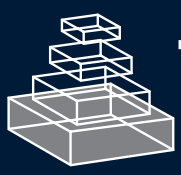
frontiers

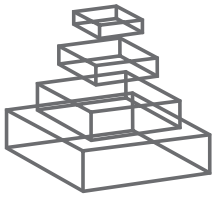
RESEARCH TOPICS

MICROBIAL GENOMICS CHALLENGE DARWIN

Topic Editors
Didier Raoult and Eugene V. Koonin



frontiers in
CELLULAR AND INFECTION MICROBIOLOGY



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-122-2
DOI 10.3389/978-2-88919-122-2

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

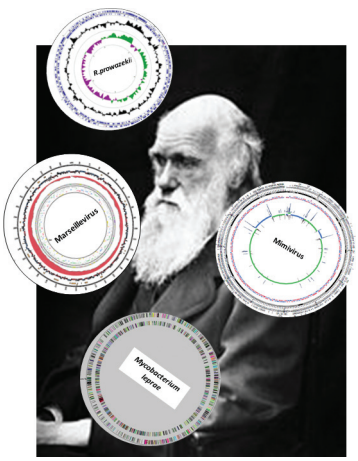
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

MICROBIAL GENOMICS CHALLENGE DARWIN

Topic Editors:
Didier Raoult, Université de la Méditerranée, France
Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, USA



The 200th anniversary of Darwin's birthday was celebrated in 2009, making the concept of Darwinism even more popular than at the time it was originally proposed, to the extent that it has acquired quasi-religious status. His theory revolves around a Tree of Life in which all living organisms are considered to have descended from a single ancestor, and each node represents a common ancestor. It comprises hierarchy and dichotomy, which are typical characteristics of the post-biblical 19th century vision. Indeed, according to post-modern philosophy (also called the French theory) the majority of theories, including scientific ones, are based only on meta-narratives expressing the influence of a culture at a given time. Buddhism or Hinduism may have generated a very different story of evolution.

Charles Darwin picture:
http://fr.wikipedia.org/wiki/Charles_Darwin

Marseille virus et Mimivirus genome:
Produced in our lab and not published

R. prowazekii genome:
Bechah Y et al., Genomic, proteomic, and transcriptomic analysis of virulent and avirulent Rickettsia prowazekii reveals its adaptive mutation capabilities. *Genome Res.* 2010 May;20(5):655–63
http://genome.cshlp.org/site/misc/terms.xhtml

Mycobacterium leprae circular genome
http://commons.wikimedia.org/wiki/File:Mycobacterium_leprae_circular_genome.png

Our way of thinking about life, and the way we describe evolution, have changed radically in the 21st century due to the genomic revolution. Comparative genome analyses have demonstrated that gene repertoires are characterized by plasticity, and there is strong evidence that nearly all genes have been exchanged at some point. Genomic data show that the genetic information of living organisms is inherited not only vertically but also laterally. Lateral gene transfers were at first observed only in bacteria, which contain genes originating from eukaryotes, Archaea and viruses. Such transfers were subsequently identified in all living organisms; giant

viruses have chimeric genomes and the human genome is a mosaic of genes with eukaryotic, bacterial, and viral origins. We cannot identify a single common ancestor for the gene repertoire of any organism. Furthermore, a very high proportion of genes have been newly created through gene fusion or degradation, and others show no homology to sequences found in other species. It is now clear that every living organism has a variety of ancestors, while exchanges between species are intense, and the creation of new genes is frequent and permanent in all living organisms. Our current genomic knowledge contradicts the tree of life theory, as established by Darwin. Recent analyses have produced bushes rather than resolved trees, with the structure of some parts remaining elusive. It becomes more and more obvious that phylogenetic relationships are better described by forests and networks and that species evolution looks more like a rhizome. The chimerism and mosaic structure of all living organisms through both non-vertical inheritance and *de novo* creation can only be assimilated and described by a post-Darwinist concept.

In this Research Topic we wish to highlight the influence of microbiology and genomics on our understanding of the complexity of gene repertoires, and also demonstrate how current knowledge does not support Darwin's theory. Microbiology has offered a great advance in the way we perceive life. Evidence obtained from studies on bacterial and viral evolution, lateral inheritance, phylogenetic trees and biodiversity continues to challenge what constituted, until recently, an unimpeded dogma in biology.

Table of Contents

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

EDITORIAL
published: 12 October 2012
doi: 10.3389/fcimb.2012.00127

Microbial genomics challenge Darwin

Didier Raoult*[1] *and* ***Eugene V. Koonin*[2]**

[1] *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Faculté de Médecine, CNRS UMR 6236, IRD 198, Aix Marseille Université, Marseille, France*
[2] *National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, USA*
**Correspondence: didier.raoult@gmail.com*

Edited by:
Yousef A. Kwaik, University of Louisville School of Medicine, USA

Reviewed by:
Yousef A. Kwaik, University of Louisville School of Medicine, USA

This collection of 14 articles in *Frontiers in Cellular and Infectious Microbiology* aims to re-assess Darwinian and neo-darwinian concepts of biological evolution in the light of the discoveries in comparative genomics of microbes in the twenty-first century. At the time of the publication of the Origins of species in 1859 (Darwin, 1859), Darwin's vision of evolution revolutionized the scientific worldview and even the human perception of the world beyond science. However, a century later, with the consolidation of the Modern Synthesis (neo-darwinism), evolutionary biology has adopted a rather rigid, somewhat dogmatic framework.

Evolutionary biologists have accepted as indisputable truths that the mutations were entirely random. They believe that:

(i) it was only selection that brought determinism and directionality into the process of evolution
(ii) all evolutionarily consequential heritable changes were extremely small in scale (the principle of gradualism that was staunchly defended by Darwin himself)
(iii) natural selection was the only important factor that shaped the evolving phenotypes and genotypes, driving in particular the emergence of biological complexity
(iv) the history of all life forms, at least in principle, could be adequately described by a single Tree of Life (following the famous sole illustration in the *Origin*) (Dobzhansky, 1937; Huxley, 1942).

At the end of the twentieth and beginning of the twenty-first centuries, genomics, especially comparative genomics of microbes, shattered each of these key tenets of (neo)Darwinism. We are now fully aware that many of the most important genomic changes are by no account miniscule; that the mutational process is far from being completely random; that evolution of complexity via routes distinct from natural selection is possible; and that pervasive horizontal gene transfer makes the original concept of the Tree of Life largely obsolete. Perhaps even more remarkably, the study of genome evolution, in particular in microbes, has brought to fore completely novel aspects of the evolutionary process of which Darwin and the architects of the Modern Synthesis were utterly unaware. Conceivably, the foremost of these phenomena is the unending arms race between cellular life forms and genomic parasites such as viruses and mobile elements that shapes the genomes of both the hosts and the parasites (Raoult, 2010a; Koonin, 2011).

In the nineteenth century, Darwin's concept of evolution was most brutally assaulted by Nietzsche who had a very low opinion of Darwin's work whose rational side, he found, was totally incompatible with life (Nietzsche, 1995). Essentially, Nietzsche believed that Darwin's theories were too simple to be true! Nietzsche's philosophical successors, especially the French post-modern philosophers, extended this radical attack on the dominant concepts of evolution. Among these, Deleuze and Guattari's reflections in "*Le rhizome*" (Deleuze and Guattari, 1976) are inspired by Jacob's work on bacteriophages. Deleuze and Guattari presciently declare that heredity is a mixture of horizontal and vertical descent, and that the origin of all living things was more like a rhizome than a tree. Thus, Deleuze and Guattari seem to have anticipated the key importance of networks in science and life years before the advent emergence of the Internet (Raoult, 2010b). These epistemological insights prepare the ground for acceptance of the revolutionary impact of the twenty-first century discoveries in genomics and epigenomics.

Although the major importance of microbial genomics for understanding evolution is beyond doubt, researchers differ in their opinions as to how radical are the changes brought about by the new discoveries. Some hold that the new discoveries only add details to the neo-Darwinian view of evolution whereas others maintain that the basic tenets of (neo)Darwinism have been falsified in the Popperian sense (Popper, 1959); yet others strive to find a middle ground by positing that, although the bulk of today's evolutionary biology consists of data and concepts that simply did not exist even 30 years ago, at the core there remains the key Darwinian principle of descent with modification.

The present *Frontiers* collection encompasses all these views. In a series of five articles on different aspects of evolution (microbial and beyond) (Georgiades et al., 2011; Colson and Raoult, 2012; Georgiades and Raoult, 2012; Ramulu et al., 2012) including a sweeping overview of modern evolutionary biology (Merhej and Raoult, 2012), Raoult and colleagues promote radical upstaging of the Darwinian paradigm. This proposed overhaul focuses primarily on the Rhizome of Life, the network representation of evolution that under this view is to supplant the Tree of Life. Two articles, by Forterre (Forterre, 2012) and by Gupta and colleagues (Bhandari et al., 2012), present the contrasting, conservative view, that the Darwinian principles remain both necessary and sufficient to understand evolution. Forterre, however, also emphasizes the fundamental importance of cell-virus conflicts that could not have been known by Darwin and his early followers

(Forterre, 2012). The review article by Koonin and Wolf (Koonin and Wolf, 2012) strives to balance the radical and the conservative approaches to Darwinian legacy by emphasizing integration more than a paradigm shift *sensu* Kuhn (Kuhn, 1962). Koonin and Wolf outline the fundamental impact of new discoveries while acknowledging Darwinian descent with modification as the surviving core of evolutionary biology. Segerman discusses the major differences in the evolutionary modalities of the stable core and the dynamic compendium of accessory genes in bacteria (Segerman, 2012). Two articles, by Danchin and Rosso (Danchin and Rosso, 2012) and by Aravind et al. (2012) address the effect that gene transfer between eukaryotes and prokaryotes as well as (once again) conflicts between selfish genetic elements and their hosts affect the evolution of eukaryotes. Bertelli and Greub discuss a more specific model system, the phagocytic amoebas and show that these organisms are veritable melting pots of horizontal gene exchange (Bertelli and Greub, 2012).

The Darwinian revolution in the nineteenth century went far beyond the scientific domain and had the broadest philosophical and cultural implications (Raoult, 2010a). It seems appropriate therefore that two articles in the present collection venture outside evolutionary biology and into economic theory and philosophy. Salvucci emphasizes the importance of moving away from simplistic models of market economy, such as those of Smith and Malthus that inspired Darwin to more integrative approaches suitable for analysis of the diverse interactions between different life forms that are central to evolution (Salvucci, 2012). Finally, Baquero and Moya address the problem of intelligibility of complex microbial systems by turning to the ideas of Wittgenstein and advocate the development of complex models that will be commensurate with the complexity of life (Baquero and Moya, 2012).

The philosophical aspect of today's evolutionary biology is especially fascinating. Clearly, we are moving away from the rigid positivist views that dominated rational thinking in Darwin's day and inevitably influenced his thought to a much richer, dynamic philosophical framework of incessant change heavily affected by chance that reverberates with the post-modern thought of the twentieth century but also harks back to the great pre-Socratic philosophers of Greece, Democritus, Parmenides, Heraclites, and Empedocles (Darwin, 1859; Dobzhansky, 1937; Huxley, 1942).

It is our hope that in this collection of articles, the interested reader finds a rich rhizome of ideas that not only summarize the key developments of evolutionary biology in the first decade of the twenty-first century but also might presage some of the directions it will take in the decades to come.

REFERENCES

Aravind, L., Anantharaman, V., Zhang, D., de Souza, R. F., and Iyer, L. M. (2012). Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front. Cell. Infect. Microbiol.* 2:89. doi: 10.3389/fcimb.2012.00089

Baquero, F., and Moya, A. (2012). Intelligibility in microbial complex systems: Wittgenstein and the score of life. *Front. Cell. Infect. Microbiol.* 2:88. doi: 10.3389/fcimb.2012.00088

Bertelli, C., and Greub, G. (2012). Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front. Cell. Infect. Microbiol.* 2:110. doi: 10.3389/fcimb.2012.00110

Bhandari, V., Naushad, H. S., and Gupta, R. S. (2012). Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution. *Front. Cell. Infect. Microbiol.* 2:98. doi: 10.3389/fcimb.2012.00098

Colson, P., and Raoult, D. (2012). Lamarckian evolution of the giant Mimivirus in allopatric laboratory culture on amoebae. *Front. Cell. Infect. Microbiol.* 2:91. doi: 10.3389/fcimb.2012.00091

Danchin, E. G., and Rosso, M. N. (2012). Lateral gene transfers have polished animal genomes: lessons from nematodes. *Front. Cell. Infect. Microbiol.* 2:27. doi: 10.3389/fcimb.2012.00027

Darwin, C. (1859). *On the Origin of Species*. London: Murray.

Deleuze, G., and Guattari, F. (1976). *Rhizome: Introduction*. Paris: Éd. de Minuit.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. New York, NY: Columbia University Press.

Forterre, P. (2012). Darwin's goldmine is still open: variation and selection run the world. *Front. Cell. Infect. Microbiol.* 2:106. doi: 10.3389/fcimb.2012.00106

Georgiades, K., Merhej, V., and Raoult, D. (2011). The influence of Rickettsiologists on post-modern microbiology. *Front. Cell. Infect. Microbiol.* 1:8. doi: 10.3389/fcimb.2011.00008

Georgiades, K., and Raoult, D. (2012). How microbiology helps define the rhizome of life. *Front. Cell. Infect. Microbiol.* 2:60. doi: 10.3389/fcimb.2012.00060

Huxley, J. S. (1942). *Evolution: The Modern Synthesis*. London: Allen and Unwin.

Koonin, E. (2011). *The Logic of Chance: The Nature and Origin of Biological Evolution*. 1st Edn. New Jersey, NJ: FT Press Science.

Koonin, E. V., and Wolf, Y. I. (2012). Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front. Cell. Infect. Microbiol.* 2:119. doi: 10.3389/fcimb.2012.00119

Kuhn, T. S. (1962). *The Structure of Scientific Revolution*. 1st Edn. Chicago, IL: University of Chicago Press.

Merhej, V., and Raoult, D. (2012). Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin. *Front. Cell. Infect. Microbiol.* 2:113. doi: 10.3389/fcimb.2012.00113

Nietzsche, F. W. (1995). *La volonté de Puissance*. 1st Edn. Paris: Gallimard.

Popper, K. (1959). *The Logic of Scientific Discovery*. New Edn. London: Taylor and Francis Group.

Ramulu, H. G., Raoult, D., and Pontarotti, P. (2012). The rhizome of life: what about metazoa? *Front. Cell. Infect. Microbiol.* 2:50. doi: 10.3389/fcimb.2012.00050

Raoult, D. (2010a). *Dépasser Darwin*. Saint-Amand-Montrond (Cher): Plon.

Raoult, D. (2010b). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Salvucci, E. (2012). Selfishness, warfare, and economics; or integration, cooperation, and biology. *Front. Cell. Infect. Microbiol.* 2:54. doi: 10.3389/fcimb.2012.00054

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* 2:116. doi: 10.3389/fcimb.2012.00116

Received: 26 September 2012; accepted: 26 September 2012; published online: 12 October 2012.

Citation: Raoult D and Koonin EV (2012) Microbial genomics challenge Darwin. Front. Cell. Inf. Microbio. 2:127. doi: 10.3389/fcimb.2012.00127

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 13 September 2012
doi: 10.3389/fcimb.2012.00119

Evolution of microbes and viruses: a paradigm shift in evolutionary biology?

Eugene V. Koonin* and Yuri I. Wolf

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Kalliopi Georgiades, Unité de Recherche en maladies infectieuses tropicales emergentes, France
L. Aravind, National Institutes of Health, USA

****Correspondence:***
Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, NIH Bldg. 38a, Bethesda 20894, MD, USA.
e-mail: koonin@ncbi.nlm.nih.gov

When Charles Darwin formulated the central principles of evolutionary biology in the *Origin of Species* in 1859 and the architects of the Modern Synthesis integrated these principles with population genetics almost a century later, the principal if not the sole objects of evolutionary biology were multicellular eukaryotes, primarily animals and plants. Before the advent of efficient gene sequencing, all attempts to extend evolutionary studies to bacteria have been futile. Sequencing of the rRNA genes in thousands of microbes allowed the construction of the three- domain "ribosomal Tree of Life" that was widely thought to have resolved the evolutionary relationships between the cellular life forms. However, subsequent massive sequencing of numerous, complete microbial genomes revealed novel evolutionary phenomena, the most fundamental of these being: (1) pervasive horizontal gene transfer (HGT), in large part mediated by viruses and plasmids, that shapes the genomes of archaea and bacteria and call for a radical revision (if not abandonment) of the Tree of Life concept, (2) Lamarckian-type inheritance that appears to be critical for antivirus defense and other forms of adaptation in prokaryotes, and (3) evolution of evolvability, i.e., dedicated mechanisms for evolution such as vehicles for HGT and stress-induced mutagenesis systems. In the non-cellular part of the microbial world, phylogenomics and metagenomics of viruses and related selfish genetic elements revealed enormous genetic and molecular diversity and extremely high abundance of viruses that come across as the dominant biological entities on earth. Furthermore, the perennial arms race between viruses and their hosts is one of the defining factors of evolution. Thus, microbial phylogenomics adds new dimensions to the fundamental picture of evolution even as the principle of descent with modification discovered by Darwin and the laws of population genetics remain at the core of evolutionary biology.

Keywords: Darwin, modern synthesis, comparative genomics, tree of life, horizontal gene transfer

INTRODUCTION

Charles Darwin's *On the Origin of Species* that appeared in London in 1859 (Darwin, 1859) was the first plausible, detailed account of biological evolution, after the simultaneous and independent brief outlines by Darwin and Alfred Russell Wallace that were published the previous year (Darwin, 1858; Wallace, 1858). Darwin did not discover evolution and did not even offer the first coherent description of evolution: exactly 50 years before the appearance of the *Origin*, the French botanist and zoologist Jean-Baptiste Lamarck published his magnum opus *Philosophie Zoologique* (Lamarck, 1809) in which he outlined his vision of the history of life in considerable detail. However, the cornerstone of Lamarck's worldview was the purported intrinsic drive of evolving organisms toward "perfection," a patently non-scientific, irrational idea. Moreover, Lamarck's view of the role of evolution in the history of life was severely limited: he did not postulate deep common ancestry of life forms but rather believed in multiple acts of creation, perhaps a separate act for each species. Prescient ideas on evolutionary changes of organisms actually have been developed centuries before Lamarck and Darwin, most notably by the great Roman thinker Titus Lucretius Carus (2011).

However, the fact remains that it was Darwin's first evolutionary synthesis that had launched the field of evolutionary biology in a sense close to the modern one and had remained central to biological thinking over the last 150 years inasmuch as "nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973). Darwin's concept lacked the essential foundation in genetics for the obvious reason that mechanisms of heredity were unknown in his day. Hence Darwin's deep concern over the so-called Jenkin nightmare, the objection to Darwin's concept according to which beneficial changes would be "diluted" after several generations in the progeny of organisms in which they occurred. The genetic basis of evolution was established after the rediscovery of Mendel's laws, with the development of population genetics in the first third of the twentieth century, primarily, through the pioneering work of Fisher, Wright, and Haldane (Fisher, 1930; Haldane, 1932). The new, advanced understanding of evolution, informed by theoretical and experimental work in genetics, was consolidated in the Modern Synthesis of evolutionary biology, usually, associated with the names of Dobzhansky, Julius Huxley, Mayr, and Simpson (Dobzhansky, 1937; Simpson, 1944). Apparently, the Modern Synthesis reached its mature form

during the 1959 centennial celebration for the *Origin* in Chicago (Tax and Callender, 1960; Browne, 2008).

Now, 50 years after the consolidation of the Modern Synthesis, evolutionary biology undoubtedly faces a new major challenge and, at the same time, the prospect of a new conceptual breakthrough (Rose and Oakley, 2007). If the Modern Synthesis can be succinctly described as Darwinism in the Light of Genetics (often referred to as neodarwinism), then, the new stage is Evolutionary Biology in the Light of Genomics and Microbiology. The combination of genomics and microbiology is indeed critical in the advent of this new age of evolutionary biology (Koonin and Wolf, 2008; Koonin, 2009a; Woese and Goldenfeld, 2009). Lamarck and Darwin (let alone Lucretius) were plainly unaware of the existence of genomes and microbes. The architects of the Modern Synthesis certainly knew about genomes and microbes "in principle" but, in the former case, did not know enough to incorporate information on genomes beyond the (important but limited) level of formal genetics, and in the latter case, did not realize the importance of microbes for understanding evolution at all.

In this article, we attempt to outline the key changes to the basic tenets of evolutionary biology brought about primarily by comparative and functional microbial genomics and argue that, in many respects, the genomic stage could be a more radical departure from the Modern Synthesis than the latter was from classic Darwinian concepts.

FROM THE TREE OF LIFE TO THE WEB OF GENE TREES

The famous sole illustration of the Origin of Species shows a Tree of Life (or more precisely, a series of trees presumably depicting the evolution of different divisions of organisms). Obviously, Darwin was not the first to use a tree to depict history. Before him, trees had been employed for many centuries to capture human genealogy, e.g., that of the Old Testament patriarchs as well as later monarchs. Darwin, however, was the first to make the crucial conceptual step by boldly proposing that the entire history of life could (at least in principle) be accurately represented by a tree growing from a single root. Darwin's tree was a sheer scheme, without any attempt to assign real life forms to the branches but in just a few years Ernst Haeckel populated the tree by a huge variety of organisms, almost exclusively animals (Haeckel, 1997). Haeckel inferred the relationships between organisms reflected in the topology of his tree primarily on the data of comparative anatomy that was already advanced in his day. Over the next century, there was considerable progress in this field leading to improved resolution of the tree but qualitatively the situation has not changed. Phylogeny largely served as a tool for systematics, and the architects of the Modern Synthesis were much more interested in mechanisms of microevolution and speciation than in the course of macroevolution that is supposedly reflected in the Tree of Life. Although by mid-twentieth century microbiologists had realized full well that microbes possess genomes and can mutate, and accordingly, should evolve, in principle, similarly to animals and plants, all attempts to infer microbial evolution from morphological and physiological characters had been unqualified failures (Stanier and Van Niel, 1962).

The fortunes of phylogeny and microbial evolution changed abruptly in the late 1970s when Carl Woese and colleagues realized that the nucleotide sequence of a universally conserved molecule, 16S rRNA, could be used to infer a universal phylogenetic tree (rather incredibly, from today's vantage point, Woese's original seminal work employed oligonucleotide maps of 16S RNA rather than sequences; however, the actual sequences became readily available shortly, and the main conclusions of the early studies stood) (Woese, 1987). Comparison of 16S RNA sequences had swiftly led to the discovery of a distinct domain of life, the Archaea, and its distinct phylogenetic affinity with the eukaryotes (Woese and Fox, 1977; Woese et al., 1990; Woese, 2004). Over the following few years, major phyla of Bacteria, Archaea and unicellular eukaryotes have been established (Woese, 1987), and the famous tripartite tree (**Figure 1**) emerged as the paradigm of the history of cellular life on earth which it more or less remains to this day (Woese et al., 1990; Pace, 1997, 2006, 2009). This was a veritable triumph of molecular phylogenetics and a dramatic departure from Haeckel's Tree of Life. In Haeckel's tree, Protista (unicellular eukaryotes) and Monera (bacteria) occupied unspecified positions near the root. For all purposes, these measly, tiny creatures were not considered important in the big picture of evolution. The tripartite tree of Woese and colleagues was a complete change of perspective. Now, two of the three domains of life were represented by prokaryotes (former Monera), and within the eukaryote domain, the majority of the phyla were represented by unicellular organisms (former Protista). The life forms formerly considered "important," i.e., the complex multicellular organisms (animals and plants), represent only two among the numerous branches of eukaryotes. There is no denying the fact that the true biodiversity on this planet is the diversity of unicellular microbes.

In the 1980s, when the paradigmatic status of the three-domain Tree of Life was established, there was little concern over the fact that technically this tree represented the history of only one gene, even if a universally present and highly conserved one. The 16S RNA was unanimously considered a suitable reference gene to represent the evolution of the respective organisms. Other universal genes, such as 18S RNA ribosomal proteins or RNA polymerase subunits, were thought to be important only to the extent their inclusion could improve the resolution of phylogenetic trees.

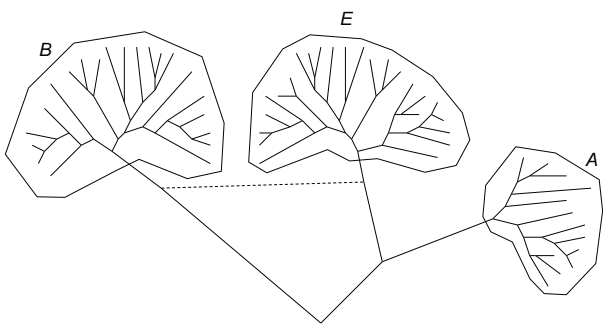


FIGURE 1 | The three-domain tree of life: a generalized schematic, A, Archaea, B, bacteria, E, Eukaryota.

Even long before the advent of the genomic era, microbiologists realized that bacteria had the capacity to exchange genetic information via horizontal gene transfer (HGT), in some cases, producing outcomes of major importance, such as antibiotic resistance (Syvanen and Kado, 2002). Multiple molecular mechanisms of HGT have been described including plasmid exchange, transduction (HGT mediated by bacteriophages), and transformation (Bushman, 2001) [indeed, the phenomenon of transformation was employed by Avery and colleagues to demonstrate the genetic function of DNA in 1944 (Avery et al., 1944a)]. However, despite these discoveries, HGT was generally viewed as a minor phenomenon that was important only under special circumstances and, in any case, did not in any manner jeopardize the Tree of Life that could be reconstructed by phylogenetic analysis of rRNA and other conserved genes.

This comfortable belief was abruptly shattered when the early findings of comparative genomics of bacteria and archaea in the late 1990s have indicated that, at least in some prokaryotic genomes, a substantial fraction of genes were acquired via demonstrable HGT, sometimes across log evolutionary distances. The pathogenicity islands and similar symbiosis islands that comprise over 30% of the genome in many pathogenic and symbiotic bacteria and obviously travel between bacteria via HGT are the prime case in point (Hacker and Kaper, 2000; Perna et al., 2001). Perhaps, more strikingly, comparative analysis of the genomes of hyperthermophilic bacteria and archaea has suggested that in shared habitats even HGT between the two domains of prokaryotes, Archaea and bacteria, can be extensive, with up to 20% of the genes of bacterial hyperthermophiles showing archaeal affinity (Aravind et al., 1998; Nelson et al., 1999; Koonin et al., 2001). Subsequent phylogenomic studies (that is analysis of phylogenies of multiple genes from numerous genomes) have led to a shocking realization: in prokaryotes at least, there seem not to exist two genes with the exact same evolutionary history (Koonin et al., 2001; Gogarten and Townsend, 2005; Gribaldo and Brochier, 2009; Zhaxybayeva, 2009; Boto, 2010; Andam and Gogarten, 2011; Zhaxybayeva and Doolittle, 2011). Apparently, this is so because all genes have experienced HGT at some stage (s) of their evolution. Although some genes, in particular those that encode components of the translation system, show substantial congruency (but not actual identity) between each other and with the standard rRNA tree, the number of such congruent trees is small. In a memorable phrase of Bill Martin and Tal Dagan, the ribosomal tree of a life is at best "a tree of one percent" (of all genes in microbial genomes) (Dagan and Martin, 2006).

Thus, "evolution of prokaryotes and the Tree of Life are two different things" (Bapteste et al., 2009; Martin, 2011). Then, the question arises: is there any substantial tree component in evolution at all and accordingly does it make any sense to speak of HGT? Indeed, horizontal transfer can be defined as such only against some standard of vertical evolution (Bapteste et al., 2005; Doolittle and Bapteste, 2007; Bapteste and Boucher, 2009). As Martin and Dagan wryly notice, if a model (in this case, the Tree of Life model) adequately describes 1% of the data, it might be advisable to abandon it and search for a better one (Dagan and Martin, 2006). Such an alternative indeed has been proposed in the form of a dynamic network of microbial evolution in which the nodes are bacterial and archaeal genomes, and the edges are the fluxes of genetic information between the genomes (Kunin et al., 2005; Dagan and Martin, 2009; Dagan, 2011; Kloesges et al., 2011). In the extreme, such a network has no vertical, tree-like component whereas the weights of the edges differ depending on the intensity of the gene exchange (**Figure 2**). Moreover, it has been persuasively argued that "tree thinking in biology" might be a sheer myth, however deeply entrenched in the textbooks and the minds of biologists (Bapteste et al., 2005; Doolittle and Bapteste, 2007; Bapteste and Boucher, 2009). Indeed, there is potential for tree-like patterns to emerge from relationships that have nothing to do with common descent as exemplified by Doolittle and Bapteste by the distribution of human names across the departments of France (Doolittle and Bapteste, 2007).

One could argue, however, that the tree pattern is not at all illusory but, on the contrary, is intrinsic and central to the entire process of biological evolution. The relevance and generality of this pattern plainly follows from the fundamental character of the replication process that underlies the evolution of life (Koonin and Wolf, 2009b). Successive generations of replicating genomes (and accordingly, dividing cells) follows an inherently binary branching pattern that, over generation naturally yields a tree. The tree pattern is predicated on a low rate of intragenic recombination which is indeed the case for all evolutionary distances large enough to prevent homologous recombination. Accordingly, evolutionary history of individual genes can be adequately represented by trees (the practical problems of accurate phylogeny reconstruction notwithstanding).

A natural, key question to ask then is: are the topologies of the trees for individual genes substantially congruent? In other words, is it possible to identify a statistically significant central trend in the vast "forest" of gene trees? Statistical analysis of thousands of phylogenetic trees for diverse genes of prokaryotes (in fact, all genes with sufficient degree of conservation to obtain a reliable tree topology) has shown that a highly significant central trend is indeed detectable in the phylogenetic forest (Puigbo et al., 2009, 2012; Koonin et al., 2011). Moreover, the consensus topology of the supertree of the (nearly) universal genes (the notorious

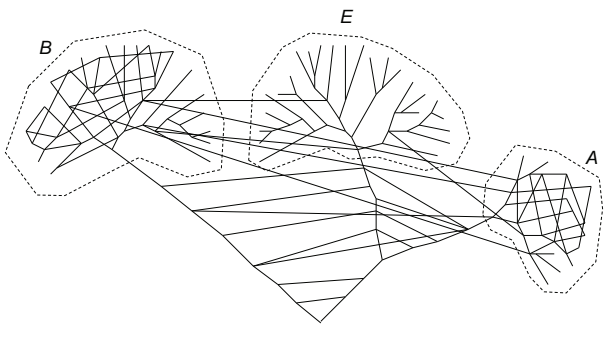


FIGURE 2 | A network representation of the evolutionary process. The network still includes some tree components such that the three domains of cellular life remains distinct but there is also an extensive horizontal component of genetic information flow that in particular dominates the earliest stages of evolution (Koonin and Wolf, 2008).

1%) turned out to be the best approximation of that central trend. Thus, although any phylogenetic tree of a central, conserved component of the cellular information-processing machinery (such as rRNA or the set of universal ribosomal proteins) represents only a minority of the phylogenetic signal across the phylogenetic forest (see details below) and so by no account can be considered an all-encompassing "Tree of Life," neither is such a phylogeny an arbitrary and irrelevant "tree of 1%." On the contrary, these trees represent a central evolutionary trend and reflect a "statistical tree of life" (O'Malley and Koonin, 2011).

THE DYNAMIC GENE UNIVERSE

For decades microbiologists knew that bacteria sometimes exchange genes (Low and Porter, 1978; Arber, 1979; Campbell, 1981; Syvanen, 1985, 1994). Moreover, the phenomena of transformation, acquisition of new traits via import of DNA from the environment and integration of the imported molecules into the bacterial genome, and transduction, transfer of genetic markers by bacteriophages, have been studied in considerable detail. In fact, transformation was the basis of the seminal 1944 experiments of Avery and colleagues which demonstrated that the genetic material of bacteria consisted of DNA (Avery et al., 1944b). In addition, microbiologists realized that such HGT could exert well-defined, major biological effects such as conferring pathogenicity (as in Avery's experiments) or antibiotic resistance on the recipients of horizontally transferred genes. However, all this knowledge notwithstanding, in the pregenomic era, HGT was considered a highly specialized genetic pathway rather than the mainstream of microbial evolution.

Comparative genomics brought the shocking realization that bacterial and archaeal genomes were literally shaped by HGT. This was clearly demonstrated by early analyses of the genomes of bacterial hyperthermophiles that were shown contain about 20% of genes of obvious archaeal origin (Aravind et al., 1998; Nelson et al., 1999; Koonin et al., 2001); conversely, genomes of mesophilic Archaea, such as *Methanosarcina*, encompass roughly the same proportion of genes clearly derived from bacteria (Deppenmeier et al., 2002; Galagan et al., 2002). These are striking examples of extensive gene exchange between the most distant prokaryotes that is stimulated by cohabitation. Not unexpectedly, the extent of gene exchange is far greater between more closely related organisms, even if often more difficult to detect (Abby et al., 2012). Nevertheless, phylogenomic analysis of a variety of bacteria and archaea clearly reveals their mosaic origins: different genes affiliate with homologs from different organisms (Koonin et al., 2001; Sicheritz-Ponten and Andersson, 2001; Koonin, 2003; Esser et al., 2007; Koonin and Wolf, 2008; Kloesges et al., 2011). These findings have been encapsulated in the concept of the Rhizome of Life under which the history of any given genome can be represented as a rhizome, with diverse sources and evolutionary histories for different genes (Raoult, 2010; Merhej et al., 2011). Recent, detailed studies indicate that at least in tight microbial communities, such as for instance the human gut microbiota, gene exchange is constant and rampant (Smillie et al., 2011).

In the face of the increasingly apparent genomic promiscuity, one cannot help asking whether "horizontal gene transfer" is a viable concept at all: indeed, for any extended span of evolution, HGT will be identifiable if and only if there is some objectively definable "vertical" standard to compare against. Otherwise, all genetic exchanges would be equal, and the only adequate depiction of evolution would be an undirected network graph. Thus, the validity of the tree representation of evolution and the very existence of HGT are inextricably linked. The results of exhaustive comparison of the individual gene trees in the "phylogenetic forest" discussed in the preceding section reveal the existence of substantial coherence of phylogenetic tree topologies, especially among highly conserved, (nearly) ubiquitous genes that encode components of the translation system (Puigbo et al., 2009). There are many exceptions to this generalization including extensive HGT of genes coding for aminoacyl-tRNA synthetases (Wolf et al., 1999; Woese et al., 2000) and even multiple cases of HGT of genes encoding ribosomal proteins (Brochier et al., 2000; Makarova et al., 2001; Yutin et al., 2012). Nevertheless, these genes appear to comprise a single, co-evolving ensemble, in at least general agreement with the so-called complexity hypothesis (Jain et al., 1999; Wellner et al., 2007; Abby et al., 2012). Under the complexity hypothesis, HGT of genes encoding subunits of macromolecular complexes is largely suppressed because of the deleterious effect caused by disruption of interactions refined by a long time of co-evolution. Indeed, a recent analysis has shown that it is the involvement in complex formation that shows a strong negative correlation with the rate of HGT, rather than any specific biological function (Cohen et al., 2011). Thus, genes encoding many translation system components probably coevolve and accordingly are rarely horizontally transferred because they are preferentially involved in large complexes (above all, the ribosome itself) rather than owing to their special biological importance or any other peculiarities of their biological function. Other genes show a much weaker but also significant phylogenetic coherence with the nearly universal genes for translation system components, perhaps also reflecting the involvement in complex formation.

The same series of phylogenomic studies that demonstrated the validity of the statistical tree of life quantified the contributions of tree-like (vertical) and web-like (horizontal) gene transmission to the relationships between bacterial and archaeal genomes (Puigbo et al., 2010, 2012). The results came out remarkably different for the ~100 nearly universal trees and the rest of the trees in the phylogenetic forest. The evolution of the nearly universal trees is dominated by the tree-like trend which contributes approximately 2/3 of the evolutionary information whereas in the rest of the forest, the ratio is the opposite, with about 2/3 of the signal coming from horizontal gene exchange (**Figure 3**).

The extensive HGT that permeates the prokaryote world is the source of gene gain by bacterial and archaeal genomes. Perhaps, the best characterized case of massive gene gain is the emergence of pathogenic bacterial strains that often evolve by acquiring the so-called pathogenicity islands that sometimes comprise over 30% of the pathogen's genome as first revealed by the comparison of the genomes of laboratory and wild strains of *E. coli* (Perna et al., 2001; Zhang et al., 2007; Eppinger et al., 2011). The opposite trend, gene loss, is at least as prominent as gene gain via HGT (Snel et al., 2002; Mirkin et al., 2003). A prime example

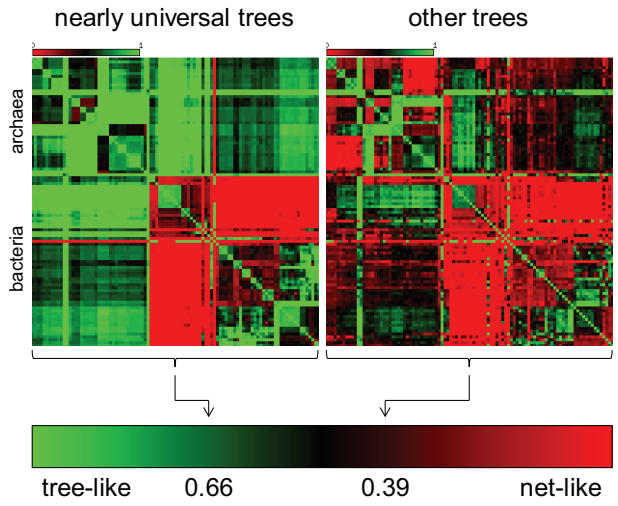


FIGURE 3 | Tree-like (vertical) and web-like (horizontal) contributions in the evolution of nearly universal genes and the entire phylogenetic forest. The two heat maps schematically depict comparison of bacterial and archaeal genomes as described previously (Puigbo et al., 2010).

is evolution of intracellular parasites and symbionts, for example, *Buchnera*, a close relative of *E. coli* that lost about 90% of the ancestral genes (Perez-Brocal et al., 2006); several other intracellular bacterial parasites and symbionts show even more drastic genome reduction (Klasson and Andersson, 2004; Perez-Brocal et al., 2006; McCutcheon and Moran, 2012).

The balance between gene gain and gene loss translates into a distinct shape of the distribution of gene occurrence in prokaryote pangenomes at all levels, from closely related bacteria (e.g., those of Enterobacteria) to the entirety of sequenced bacterial and archaeal genomes (Koonin and Wolf, 2008; O'Malley and Koonin, 2011). This universal distribution has an asymmetric U-shape and can be approximated by three exponential functions (**Figure 4**). The first of these corresponds to a small, highly conserved core (the nearly universal genes discussed above); the second exponent describes the much larger "shell" of genes with limited conservation; and the third one delineates the vast "cloud" of rare, poorly conserved genes. Thus, the gene universe is dominated by rare, sparsely distributed genes most of which are not covered by the limited available sampling of genomes and still remain to be discovered although in each particular genome the moderately conserved "shell" genes comprise the majority (**Figure 5**). The dynamic, fluid character of the prokaryote genomes yields a distinct, fractal-like structure of the gene universe (O'Malley and Koonin, 2011).

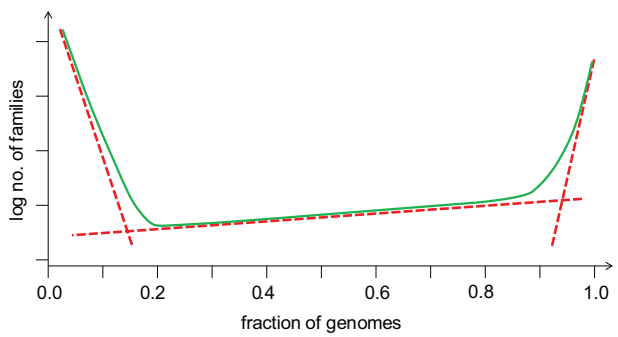


FIGURE 4 | The universal distribution of gene commonality in the microbial genomic universe: a generalized schematic. The three broken lines represent three exponential functions that fit the core (on the right), the shell (in the middle) and the cloud (on the left) of prokaryotic genes (O'Malley and Koonin, 2011).

ARE THERE SPECIES IN PROKARYOTES?

The title of Darwin's seminal book "The Origin of Species" is deeply steeped in traditions of eighteenth and nineteenth century biology that tended to view animal and plant species as key units of biological organization. Darwin himself actually saw species more as an arbitrary category in the continuum of varying life forms than a fundamental unit of life. In the twentieth century the species concept received its biological interpretation, primarily in the work of Ernst Mayr who famously defined a species as a system of panmictic populations that are genetically isolated from other such systems (Mayr, 1944). This concept indeed captures a key feature of the biology of organisms with regular, obligatory sexual reproduction such as, above all, animals and to a lesser extent plants.

Most of the prokaryotes do not engage in regular sex but instead exchange genes via HGT with diverse other microbes that they happen to cohabitate with. In general, in the prokaryote world, there are indeed no discrete, genetically isolated systems of panmictic populations but rather complex webs of gene exchange (Dagan et al., 2008; Koonin and Wolf, 2008). Thus, the very notion of species as a distinct biological category does not apply even though traditionally bacteria and archaea are still denoted by Linnaean species names (e.g., *Escherichia coli* or *Haloferax volcanii*) (Konstantinidis et al., 2006; Cohan and Perry, 2007; Doolittle and Zhaxybayeva, 2009; Fraser et al.,

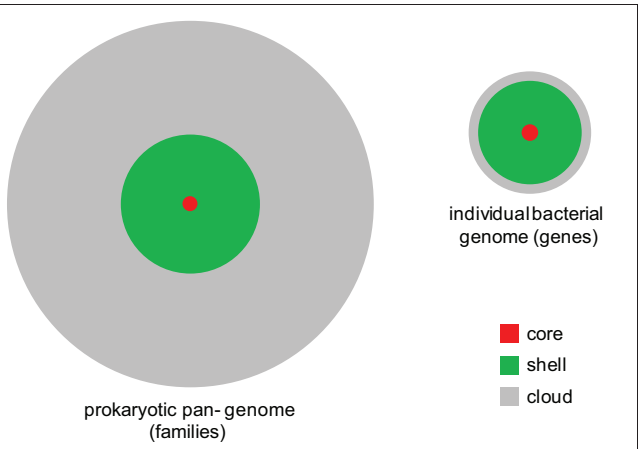


FIGURE 5 | The core, shell, and cloud of microbial genes. A generalized schematic showing the approximate contributions of the core, shell, and cloud to the pangenomes of prokaryotes and individual genomes.

2009). However, the modes of evolution substantially differ across the diversity of prokaryotes, spanning the entire continuum from fully sexual to fully clonal populations (Smith et al., 1993; Doolittle and Zhaxybayeva, 2009). Some bacteria, especially parasites such as for example *Neisseria gonorrhoeae*, have been shown to form largely isolated communities that engage in regular conjugation, the bacterial equivalent of sex, resulting in extensive homologous recombination. For these distinct organisms but not for the majority of bacteria and archaea, Mayr's biological definition of species might be a relevant concept.

The irrelevance of the (traditional) species concept for most prokaryotes by no means implies non-existence of structure in the genome space. Indeed, bacteria and archaea that share common origin in phylogenetic trees of marker genes, such as rRNA, typically also possess similar gene content. The "genome-trees" constructed on the basis of the (dis)similarity of gene content are generally congruent with phylogenetic trees of highly conserved marker genes although interesting deviations that reflect similarities in life style and/or extensive gene exchange have been detected as well (Snel et al., 1999, 2005; Wolf et al., 2002).

Thus, although the bacterial and archaeal "species" are not species in the regular sense, they are "galaxies" in the gene universe that form distinct, hierarchical clusters. Interestingly, it has been shown that, among the processes that lead to the divergence of gene content between evolving lineages of prokaryotes, gene loss appears to occur stochastically and generally follows the divergence of marker genes whereas gene gain (primarily, via HGT) is more episodic (Snel et al., 2002; Novichkov et al., 2004).

DOES EVOLUTION ADVANCE COMPLEXITY?

The idea of a general evolutionary trend toward increasing complexity is extremely popular among both lay public and scientists and certainly was shared by Darwin who wrote, for example, in famous quote: "as natural selection works solely by and for the good of each being, all corporeal and mental endowments will tend to progress toward perfection" (Darwin, 1859). This view does not imply any mysterious strive for perfection as imagined by some pre-Darwinian biologists including Lamarck (1809) or teleology of any kind. Nevertheless, Darwin's position does suggest a trend of evolution from simple to complex forms which is indeed a highly intuitive notion that has some obvious support in well known facts of the history of life on earth. For example, the most organizationally complex organisms with the largest genomes, animals, and plants, appear only at relatively late stages of evolution. Even more generally, at the earliest stages in the evolution of life, origin of complex structures, such as the cell itself, "from so simple a beginning" (Darwin, 1859) appears inevitable. Thus, notwithstanding the numerous cases of reductive evolution, in particular among parasites and symbionts, the belief in a general complexification trend in the evolution of life appears to be common.

However, is complexification the prevailing modality of evolution? Phylogenomic reconstruction, at least for bacteria and Archaea, suggests otherwise. It is not surprising that differential gene loss dominates the evolution of commensal bacteria, such as Lactobacilli, from a complex free-living ancestor (Makarova et al., 2006). However, a qualitatively similar pattern was detected in evolutionary reconstructions for all bacteria and archaea (Snel et al., 2002; Mirkin et al., 2003; Makarova et al., 2007). Strikingly, more recent reconstructions that were performed using larger genome sets and more sophisticated computational methods confidently indicate that the genome of the last common ancestor of all extant archaea apparently was at least as large and complex as that of typical modern organisms in this domain of cellular life (Csuros and Miklos, 2009). Fully compatible reconstruction results have been reported for the expanded set of cyanobacterial genomes (Larsson et al., 2011). Thus, counter-intuitively, at least in prokaryotes, genome shrinkage that is sometimes called streamlining (Lynch, 2006) and is attributed to increasing selective pressure in successful, large populations (Lynch, 2006; Koonin, 2009b), appears to be is no less and probably more common than genome growth and complexification.

THE WRIGHTEAN-DARWINIAN-LAMARCKIAN CONTINUUM OF EVOLUTIONARY PROCESSES

The Modern Synthesis of evolutionary biology emphasizes the randomness of mutations that provide the starting material for selection which engenders survival of the fittest under the given conditions and hence constitutes the adaptive, deterministic component of evolution. The insistence on such strict separation between the stochastic and deterministic aspects of evolution departs from Darwin's view that included the Lamarckian inheritance, with adaptive mutations directly caused by environmental cues, as an important, even if ancillary mechanism of evolution (Darwin, 1872).

Recently, several genetic phenomena with a distinct Lamarckian flavor have been discovered (Koonin and Wolf, 2009a; O'Malley and Koonin, 2011). Probably, the most striking case is the system of adaptive antivirus immunity, known as CRISPR-Cas (Clustered Regularly Interspaced Palindromic Repeats and CRISPR-associated proteins), that is present in most archaea and many bacteria (Koonin and Makarova, 2009; van der Oost et al., 2009; Marraffini and Sontheimer, 2010; Makarova et al., 2011). The CRISPR-Cas system integrates fragments of virus or plasmid DNA into a distinct, repetitive locus in the archaeal or bacterial genome. The transcript of this unique spacer functions as a guide RNA that is incorporated into a specific complex of Cas proteins possessing DNAse activity and directs this complex to the cognate alien DNA (or RNA) molecules that are cleaved and accordingly inactivated. The CRISPR-Cas system is amazingly efficient, with only about 10^{-5} failure rate (Deveau et al., 2008). This mechanism qualifies CRISPR-Cas as an adaptive immunity system, i.e., immunity system that adapts to a specific infectious agent, a novelty in prokaryotes (Koonin and Makarova, 2009; Bikard and Marraffini, 2012). Furthermore, the Lamarckian principle of inheritance and evolution is apparent in the mechanism of CRISPR-Cas function. Indeed, this system directly responds to an environmental cue (in this case, foreign DNA) by introducing a genetic change into the genome that is immediately adaptive with respect to that particular cue.

The discovery of the CRISPR-Cas immune system that functions on the Lamarckian principle drew attention to other phenomena that also seem to contain a Lamarckian component

(Koonin and Wolf, 2009a; O'Malley and Koonin, 2011). Some of the common, central evolutionary processes such as HGT and stress-induced mutagenesis show a "quasi-Lamarckian" character. Indeed, even if HGT cannot be viewed as being directly caused by a specific environmental factor, it certainly is the case that the repertoire of the acquired genes depends on the environment. Genes common in a given environment will be acquired often and are likely to possess adaptive value. Stress-induced mutagenesis is triggered directly by environmental stress factors, e.g., desiccation or radiation, and produces variation that is required to develop resistant phenotype (Rosenberg and Hastings, 2003; Ponder et al., 2005; Galhardo et al., 2007; Galhardo and Rosenberg, 2009). The mutations are not specific to the biologically relevant loci but the activity of the molecular machineries of stress-induced mutagenesis [the best characterized of which is the SOS repair-mutagenesis system in bacteria (Sutton et al., 2000)] generates clusters of mutations, thus locally amplifying variability and so increasing the chance of adaptation once a single mutation appears in a relevant gene (Galhardo et al., 2007).

More generally, recent empirical and theoretical studies of diverse processes of stochastic and deterministic change in genomes make it clear that evolution is not limited to the basic Darwinian scheme of random variation that is subject to selection. Evolution can be more adequately depicted as a continuum of processes from completely random ones, under the Wrightean modality defined by random variation and random fixation of changes via genetic drift; to the Darwinian modality with random changes fixed by the deterministic process of selection; to the Lamarckian mode in which both variation and fixation are deterministic (**Figure 6**) (Koonin and Wolf, 2009a; O'Malley and Koonin, 2011).

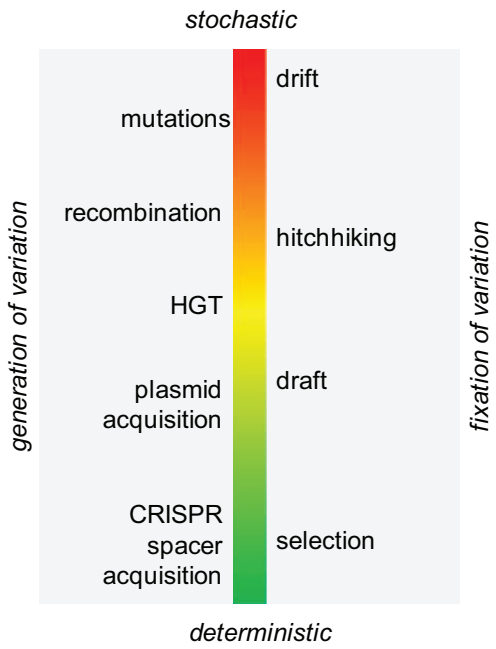


FIGURE 6 | The continuum of evolutionary processes, from stochasticity to determinism.

EVOLUTION OF EVOLVABILITY: DEDICATED MECHANISMS FOR EVOLUTION

All organisms possess a certain degree of evolvability, i.e., the ability to evolve. At the most basic level, evolvability stems from the theoretical impossibility of error-free replication. Genomic variation in evolving organisms is created by a combination of intrinsic replication errors, recombination and mutations induced external agents (mutagens). An intriguing, fundamental question in evolutionary biology is whether or not evolvability itself can evolve under selection, or put another way, whether there are dedicated mechanisms of evolution (Kirschner and Gerhart, 1998; Poole et al., 2003; Pigliucci, 2008; Brookfield, 2009). The prevailing wisdom among biologists seems to be that evolvability is not selectable but is simply maintained at a sufficient level by inevitable errors at all levels of biological information processing. Under this view, selection is always directed at minimization of the error rate but the ability to attain perfection is limited by genetic drift resulting in sufficient evolvability (Lynch, 2011). Evolutionary biologists are usually suspicious of the evolution of evolvability, generally under the old adage, "evolution has no forecast."

Nevertheless, evidence in support of "evolvability of evolvability" is mounting. The very existence of complex molecular systems for stress-induced mutagenesis (error-prone repair) the activity of which is exquisitely regulated in response to stress implies that mechanisms enhancing variation when variation is needed for survival have evolved (Galhardo et al., 2007). Another remarkable mechanism that appears to have specifically evolved to generate variation involves the Diversity Generating Retroelements (DGR) (Medhekar and Miller, 2007). Strikingly, the DGR are found both in bacteriophages where they generate diversity in cell attachment surface proteins via reverse transcription-mediated mutagenesis, resulting in host tropism switching (Doulatov et al., 2004; Guo et al., 2008), and in bacteria themselves where they produce receptor variation leading to bacteriophage resistance (Bikard and Marraffini, 2012). The analogy between the activity of DGR and hypermutagenesis in animal immune systems is obvious except that the variation generated by the DGR is inherited.

Many bacteria and some archaea possess the natural transformation ability (that was used in the Avery experiment) that requires specialized, complex pumps (recently denoted transformosomes) that internalize DNA from the environment (Claverys et al., 2009; Johnsborg and Havarstein, 2009; Kruger and Stingl, 2011). The transformation machinery potentially could be viewed as a device that evolved under selective pressure to enhance HGT (Johnsborg and Havarstein, 2009). However, one could argue that the enhancement of HGT is only a side effect of the evolution of the transformation system, its actual raison d'etre being the utilization of DNA as a rich source of replication substrates (or simply food). This argument hardly can hold with regard to the type 4 secretion systems (T4SS) that specialize in secretion of DNA from bacterial cells (Hamilton et al., 2005; Hamilton and Dillard, 2006). The recently discovered Gene Transfer Agents (GTAs) are even more striking devices for DNA donation (Paul, 2008; McDaniel et al., 2010; Lang et al., 2012). The GTAs are a distinct type of defective bacteriophages that

package in the capsid not the phage genome (which remains integrated in the host chromosome) but rather apparently random pieces of the host chromosome. The GTAs have been discovered in diverse bacteria and archaea and have been shown to infect and transfer their genetic content to a broad range of cohabitating prokaryotes (McDaniel et al., 2010). It does not seem conceivable that GTAs are anything but dedicated HGT vehicles. An additional notable aspect of T4SS and GTAs is that these devices mediate donation rather than consumption of DNA, i.e., apparently can directly benefit other microbes (recipients) rather than the donor. This seemingly altruistic behavior can be explained in terms of group selection whereby the object of selection is an ensemble of organisms that jointly benefit from adaptive mutations rather than a single organism. Group selection is a controversial subject in evolutionary biology (Maynard Smith, 1998; Borrello, 2005; Leigh, 2010) but the existence of dedicated devices for DNA donation appears to be a strong argument in its favor.

The discovery of T4SS and GTAs may be the most clear-cut pieces of evidence supporting evolution of evolvability just as the CRISPR-Cas system is the showcase for Lamarckian evolution. However, the case for the evolution of mechanisms for evolution seems to be much more general (O'Malley and Koonin, 2011). Population genetic theory holds that under a broad range of conditions a clonal population is generally doomed to collapse through the action of Muller's ratchet, the irreversible accumulation of deleterious mutations leading to gradual decline in fitness (Leigh, 2010; Bachtrog and Gordo, 2004). The effect of Muller's ratchet that has been directly demonstrated in controlled evolutionary experiments on RNA viruses (Chao, 1990; Duarte et al., 1992) and on bacteria (Andersson and Hughes, 1996). The principal way to escape Muller's ratchet is to enhance recombination via sex (in the form of meiotic crossing over in eukaryotes and in the form of conjugation in prokaryotes) or HGT. Just as sex is generally viewed as a mechanism that evolved to counteract the ratchet, HGT may be best understood as a more general variation-generating process that is supported by various evolved mechanisms. At the risk of being provocative, sex indeed can be legitimately regarded as a specialized form of HGT. Clearly, evolution maintains HGT within the optimal range rather than at the maximum possible level because the latter would eliminate genome stability and wreak havoc into selected high-fitness ensembles of genes (O'Malley and Koonin, 2011). Mechanisms that counter HGT also have evolved: these are the same that provide resistance against virus infection including CRISPR-Cas and restriction-modification (Marraffini and Sontheimer, 2008; Gardner and Olson, 2012).

At a different level, an apparent mechanism of evolution involves unusual, stable phenotype modifications that are widespread in bacteria and lead to coexistence of two distinct phenotypes in a clonal population, the so-called bistability regimes (Dubnau and Losick, 2006; Veening et al., 2008a; Piggot, 2010). For instance, under limited nutrient supply, *Bacillus subtilis* will form two subpopulations of which only the smaller one has the capacity to sporulate and thus yields the only survivors when the conditions become incompatible with cell growth and division (Veening et al., 2008a,b; Lopez et al., 2009). The coexistence is epigenetically inherited across many bacterial generations, hence this phenomenon has become known as bistability. In theoretical and experimental models bistability is rationalized as "bet hedging": for organisms that live in often and unpredictably changing environments, it is beneficial to maintain a small subpopulation of likely survivors even when their fitness is comparatively low under normal conditions (Veening et al., 2008a; de Jong et al., 2011; Libby and Rainey, 2011; Rainey et al., 2011). The cost of maintaining this subpopulation is more than compensated by the benefit of survival under adverse conditions. Thus, the evolution of the regulatory circuitry that supports bistability appears to be not just a case of evolution of an evolutionary mechanism but more specifically evolution of a kin selection mechanism or evolution of altruism in bacteria. The evolution of kin selection demonstrated by bet hedging is paralleled by the mechanism of altruistic suicide that virus-infected bacteria and archaea commit using the toxin-antitoxin or abortive infection defense systems (Makarova et al., 2009; Van Melderen and Saavedra De Bast, 2009; Hayes and Van Melderen, 2011). In this case, by killing themselves early, before the virus has a chance to replicate, the microbes save their kin from infection. The reality of kin selection, just as that of group selection, is often hotly debated by evolutionary biologists (Nowak et al., 2010; Bourke, 2011; Ferriere and Michod, 2011; Strassmann et al., 2011) but the bistability/bet-hedging phenomena and altruistic suicide in bacteria and archaea seem to plainly demonstrate not only the existence but the evolvability of this form of selection.

In parallel with experimental studies, several theoretical models have been developed that characterize evolvability as a selectable trait in fluctuating environments (Earl and Deem, 2004; Jones et al., 2007; Draghi and Wagner, 2008). Thus, on the whole, and general theoretical doubts notwithstanding, evolution of evolvability appears to be an intrinsic and fundamental, if still poorly understood, aspect of the evolutionary process.

THE VAST, ANCIENT WORLD OF VIRUSES

Viruses are no part of the modern synthesis or more generally the traditional narrative of evolutionary biology. Until very recently, viruses have been viewed primarily as pathogens of animals, plants, and bacteria. Several lines of recent discovery have radically changed this view and promoted viruses to a central position on the stage of evolution. This change in the evolutionary status of viruses and related selfish genetic elements has been discussed in detail elsewhere (Claverie, 2006; Koonin et al., 2006, 2011; Raoult and Forterre, 2008). Here we quickly recapitulate several key points, with a focus on the importance of viruses for evolutionary biology in general. Metagenomic and ecological genomics studies have shown that, astonishingly, viruses are the most common biological entities on earth (Edwards and Rohwer, 2005; Suttle, 2005, 2007). Viruses and/or virus-like mobile elements are present in all cellular life forms. Strikingly, in mammals sequences derived from mobile elements and endogenous viruses account for at least 50% of the genome whereas in plants this fraction can reach 90% (Feschotte et al., 2002; Kazazian, 2004; Devos et al., 2005; Hedges and Batzer, 2005). Even the genomes of some unicellular eukaryotes, such as *Trichomonas vaginalis*, consist mostly

of inactivated transposons (Carlton et al., 2007; Pritham et al., 2007). Recruitment of mobile element sequences for transcription regulation and other cellular functions such as microRNA formation is a common phenomenon the full extent of which is not yet fully appreciated (Jordan et al., 2003; Piriyapongsa et al., 2007; Lisch and Bennetzen, 2011). Although genomes of prokaryotes are not so overwhelmed by mobile elements, due to the intense purifying selection, nearly all of them encompass multiple prophages and mobile elements. Notably, deletion of all prophages leads to a substantial drop of fitness in *E. coli* (Wang et al., 2010).

In at least some common environments such as ocean water and soil, the number of virus particles exceeds the number of cells by factors of 10–100 (Edwards and Rohwer, 2005; Suttle, 2007; Srinivasiah et al., 2008; Breitbart, 2012). Similarly, the genetic diversity of viruses, measured as the number of distinct genes, substantially exceeds the genetic diversity of cellular life forms. Furthermore, viruses, in particular bacteriophages, are major biogeochemical agents. Periodical killing of microbes, in particular cyanobacteria, has been identified as a major contributor to sediment formation and major contributors to the nutrient cycles in the biosphere (Suttle, 2007; Rohwer and Thurber, 2009). The same process obviously is a key determinant of the population dynamics of the hosts that shapes the selection-drift balance throughout the course of evolution (Weinbauer and Rassoulzadegan, 2004).

The very fact that viruses greatly outnumber bacteria in the environment implies that antivirus defense systems are central to the evolution of bacteria and archaea. This is indeed the case as made evident by the remarkable proliferation of diverse antivirus systems including CRISPR-Cas discussed above as well as multiple restriction-modification, abortive infection, toxin-antitoxin and other, still poorly characterized defense systems that in different combinations and with different abundances are present in most prokaryotes (Juhas et al., 2009; Labrie et al., 2010; Makarova et al., 2011; Martinez-Borra et al., 2012). Taken together, these findings and theoretical considerations strongly support the view that the virus-host arms race is one of the principal processes in all evolution (Forterre and Prangishvili, 2009; Stern and Sorek, 2011).

With regard to the classification of life forms, the only defensible position appears to be that viruses (and related mobile elements) and cells are the two principal categories of biological organization (**Figure 7**) (Raoult and Forterre, 2008; Koonin, 2010; O'Malley and Koonin, 2011); this view is independent of the semantic issue of viruses being "alive" or not (Koonin et al., 2009; Moreira and Lopez-Garcia, 2009; Raoult, 2009). These two categories of biological entities can be characterized as informational (genetic) parasites, i.e., viruses and other selfish elements, and genetically self-sustained organisms, i.e., cellular life forms. Mathematical modeling indicates that genetic parasites inevitably emerge in any replicator system (Szathmary and Maynard Smith, 1997; Takeuchi and Hogeweg, 2012). This conclusion is certainly intuitively plausible: one expects that cheaters will appear in any system with limited resources—in particular, in any system of replicators, such parasites will attempt to utilize the replication machinery without making it (Koonin and Martin, 2005). Also, the notion that virus-like selfish elements are an intrinsic part of life since its inception [which can be reasonably considered to coincide with the origin of replication (O'Malley and Koonin, 2011)] is compatible with the ubiquity of these elements in nature. In mathematical modeling, the outcome of the virus-host interaction depends on the specific parameters of the adapted model. In homogeneous models, virus-like parasites tend to cause collapse of the entire systems but in models with compartmentalization, which are most relevant for the actual evolution of life, stable host-parasite coexistence is possible (Takeuchi and Hogeweg, 2009). Moreover, the destructive effect of genetic parasites on the host is mitigated when a dedicated genetic information storage medium evolves, which could be one of the driving forces behind the evolution of DNA in the primordial RNA world (Takeuchi et al., 2011).

Further support for the classification of viruses as one of the two "empires" of life is the diversity of the replication-expression cycles that is found among viruses and related elements. Indeed, while cellular life forms all use a uniform replication-expression strategy based on double-stranded (ds)DNA replication, transcription of genes into mRNA or non-coding RNA, and translation of mRNA into protein, viral genome can be represented by all known forms of nucleic acids, and alternative replication processes such as RNA replication and reverse transcription are widely used (**Figure 7**) (Koonin et al., 2006). Finally, although viral genomes are generally small compared to the genomes of cellular life forms (viruses being the ultimate genetic parasites), the range of genomic complexity is remarkable, from only about 300 nucleotides and no genes in the simplest virus-like parasites, the viroids, to over a megabase and more than 1000 genes (genomes that are more complex than those of many bacterial parasites and symbionts) in the giant mimiviruses (Raoult et al., 2004; Colson et al., 2012). Overall, the conclusion is inescapable that the entire history of life is a story of perennial interplay between genetic parasites and their hosts that is a major driver of evolution for both biological empires.

EVOLUTION OF MICROBES AND VIRUSES: A NEW EVOLUTIONARY PARADIGM?

Prokaryotes (bacteria and archaea) and viruses entered the realm of evolution with the advent of genomics. Has the comparative study of these relatively simple (compared to eukaryotes) organisms radically changed the core tenets of evolutionary biology that were first envisaged by Darwin and were augmented with the genetic foundation in the Modern Synthesis? In terms of Kuhn's concept of the development of science (Kuhn, 1962), did the study of microbial evolution engender a paradigm shift?

It is not easy to answer this question definitively, possibly because the paradigm shift model does not adequately describe the evolution of biology (regardless of whether or not it fits the evolution of physics). Probably, a more appropriate epistemological framework is that of integration, i.e., a relatively smooth incorporation of the classic concepts into the new, more general and versatile theoretical constructs. This model of the evolution of science was recognized by Kuhn himself in his later work (Kuhn, 2002) and was recently examined by O'Malley in the context of biology (O'Malley, 2012; O'Malley and Soyer, 2012).

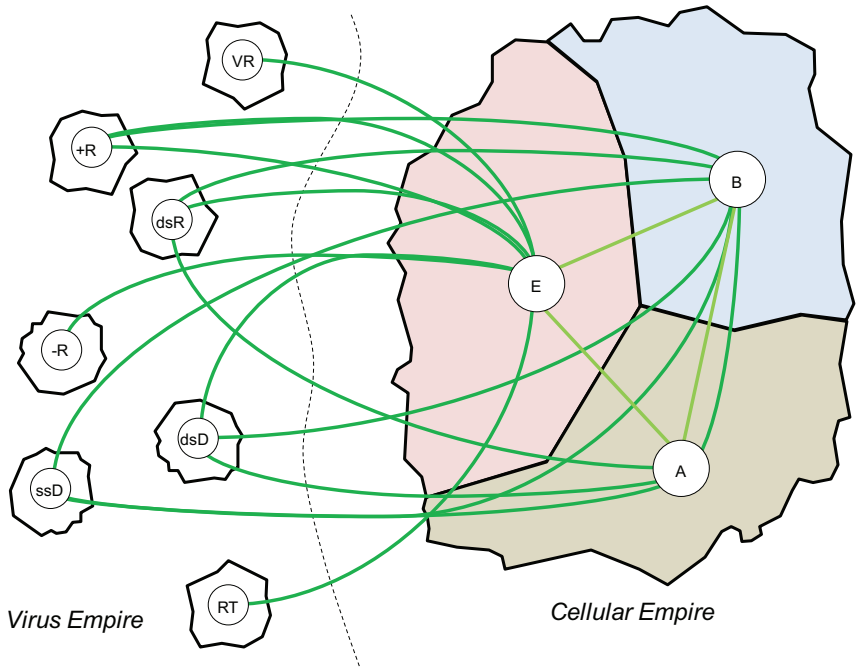


FIGURE 7 | The viral and cellular "empires" of life forms and domains within them. The cellular empire domains: A, Archaea; B, Bacteria; E, Eukaryota. The Virus empire domains: +R, positive-strand RNA viruses; −R, negative-strand RNA viruses; dsR, double-stranded RNA viruses; dsD, double-stranded DNA viruses; ssD, single-stranded DNA viruses; RT, retro-transcribing elements/viruses; VR, viroids.

The phylogenomic study of microbes and viruses uncovered new biological realms which Darwin and even the authors of the Modern Synthesis could not possibly fathom. The modes of evolution of these relatively simple organisms that, as we now realize, have dominated the biosphere since its beginning about 4 billion years ago to this day (and into any conceivable future) are different from the evolutionary regimes of animals and plants, the traditional objects of (evolutionary) biology. The study of microbial evolution has shattered the classic idea of a single, all-encompassing tree of life by demonstrating that the evolutionary histories of individual genes are generally different. Remarkably, however, these developments have not rendered trees irrelevant as a key metaphor of evolution (O'Malley and Koonin, 2011). Rather, they have shown that the bona fide unit of tree-like evolution is an individual gene not a genome, and a "tree of life" can only be conceived as a statistical trend in the "forest" of gene trees (Koonin and Wolf, 2009b). Tree-like evolution is a fundamental implication of the binary replication of the genetic material, so it served Darwin well to use a tree as the single illustration of his book. Without, obviously, knowing anything of DNA replication, Darwin grasped the central principle of the evolution of life, descent with modification, and the tree pattern followed naturally.

Microbiology yielded the first clear-cut case of Lamarckian evolution, the CRISPR-Cas system, and subsequent re-examination of other evolutionary phenomena (in both prokaryotes and eukaryotes) has strongly suggested that the (quasi)Lamarckian modality is common and important in all evolving organisms, completing the range of evolutionary phenomena from purely stochastic (drift, Wrightean evolution) to deterministic (Lamarckian evolution). Again, these findings not so much overturned but rather expanded the vision of Darwin who seriously considered Lamarckian mechanisms as being ancillary to natural selection (only the Modern synthesis banished Lamarck).

Crucially, the study of microbial evolution presented apparently undeniable cases of evolution of evolvability such as the GTAs and the DGRs. Moreover, the discovery of bet-hedging strategies and altruistic suicide in bacteria shows that kin selection (a subject of considerable controversy in evolutionary biology) is evolvable as well. Again, as in the case of Lamarckian mechanisms, these discoveries force one to re-examine many more phenomena and realize that evolution is not limited to fixation of random variation and survival of the fittest but rather is an active process with multiple feedback loops, and that dedicated mechanisms of evolution exist and themselves evolve. This is a major generalization that substantially adds to the overall structure of evolutionary biology but one has to realize that the principle of descent with modification remains at the core of all these complex evolutionary phenomena.

We now realize that evolution of life is to a large extent shaped by the interaction (arms race but also cooperation) between genetic parasites (viruses and other selfish elements) and their cellular hosts. Viruses and related elements, with their distinctive life strategy, informational parasitism, actually dominate the biosphere both physically and genetically, and represent one of the two principal forms of life that as intrinsic to the history of the biosphere as cells are. This new dimension of evolution simply

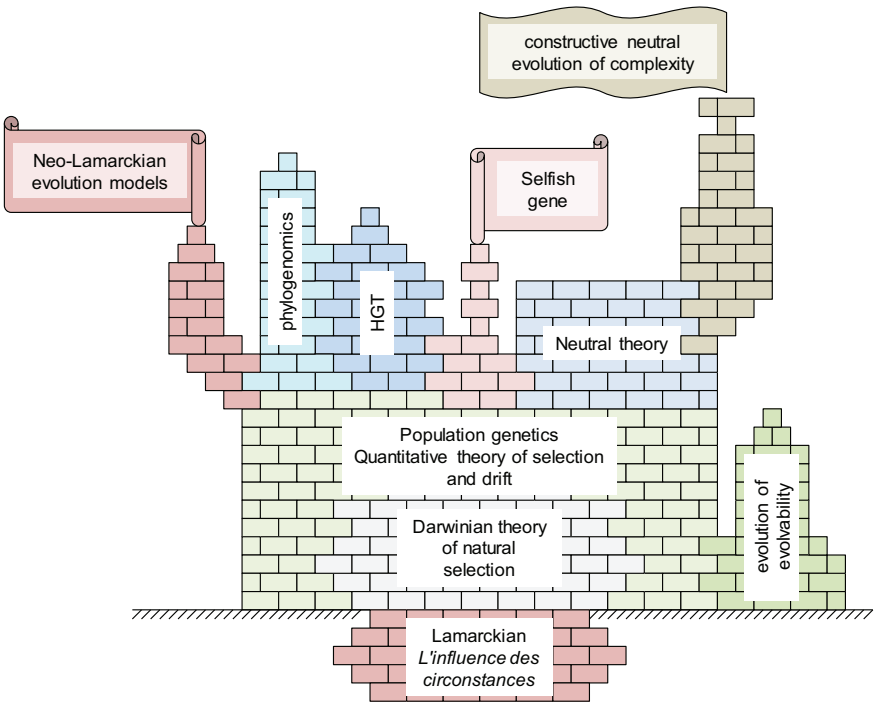


FIGURE 8 | The conceptual structure of evolutionary biology: the Darwinian core and the new levels of complexity.

could not be perceived by Darwin or even the creators of the Modern Synthesis due to the lack of relevant data.

Thus, we are inclined to view the change in evolutionary biology brought about by phylogenomics of microbes and viruses as a case of integration rather than an abrupt departure from the paradigm of the Modern Synthesis (**Figure 8**). Darwin realized the importance of descent with modification and the tree pattern of evolution it implies whereas Fisher, Wright, and Haldane derived the laws of population genetics that still constitute the core of our understanding of evolution. However, recent advances, in particular those of microbial phylogenomics, added multiple, new and interconnected layers of complexity (**Figure 8**) such that the conceptual core is but a small part of the current big picture of evolutionary biology.

AUTHOR CONTRIBUTIONS

Eugene V. Koonin and Yuri I. Wolf wrote the manuscript.

ACKNOWLEDGMENTS

The authors' research is supported by intramural funds of the US Department of Health and Human Services (to National Library of Medicine, NIH).

REFERENCES

Abby, S. S., Tannier, E., Gouy, M., and Daubin, V. (2012). Lateral gene transfer as a support for the tree of life. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4962–4967.

Andam, C. P., and Gogarten, J. P. (2011). Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* 9, 543–555.

Andersson, D. I., and Hughes, D. (1996). Muller's ratchet decreases fitness of a DNA-based microbe. *Proc. Natl. Acad. Sci. U.S.A.* 93, 906–907.

Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R., and Koonin, E. V. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444.

Arber, W. (1979). Promotion and limitation of genetic exchange. *Experientia* 35, 287–293.

Avery, O. T., MacLeod, C. M., and McCarty, M. (1944a). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 79, 137–158.

Avery, O. T., MacLeod, C. M., and McCarty, M. (1944b). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 149, 297–326.

Bachtrog, D., and Gordo, I. (2004). Adaptive evolution of asexual populations under Muller's ratchet. *Evolution* 58, 1403–1413.

Bapteste, E., and Boucher, Y. (2009). Epistemological impacts of horizontal gene transfer on classification in microbiology. *Methods Mol. Biol.* 532, 55–72.

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupre, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4, 34.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R. L., and Doolittle, W. F. (2005). Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* 5, 33.

Bikard, D., and Marraffini, L. A. (2012). Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Curr. Opin. Immunol.* 24, 15–20.

Borrello, M. E. (2005). The rise, fall and resurrection of group selection. *Endeavour* 29, 43–47.

Boto, L. (2010). Horizontal gene transfer in evolution: facts and

challenges. *Proc. Biol. Sci.* 277, 819–827.

Bourke, A. F. (2011). The validity and value of inclusive fitness theory. *Proc. Biol. Sci.* 278, 3313–3320.

Breitbart, M. (2012). Marine viruses: truth or dare. *Ann. Rev. Mar. Sci.* 4, 425–448.

Brochier, C., Philippe, H., and Moreira, D. (2000). The evolutionary history of ribosomal protein RpS14, horizontal gene transfer at the heart of the ribosome. *Trends Genet.* 16, 529–533.

Brookfield, J. F. (2009). Evolution and evolvability: celebrating Darwin 200. *Biol. Lett.* 5, 44–46.

Browne, J. (2008). Birthdays to remember. *Nature* 456, 324–325.

Bushman, F. (2001). *Lateral DNA Transfer: Mechanisms and Consequences.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Campbell, A. (1981). Evolutionary significance of accessory DNA elements in bacteria. *Annu. Rev. Microbiol.* 35, 55–83.

Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M. Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., and Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science* 315, 207–212.

Carus, T. L. (2011). *De Rerum Natura.* New York, NY: Nabu Press.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature* 348, 454–455.

Claverie, J. M. (2006). Viruses take center stage in cellular evolution. *Genome Biol.* 7, 110.

Claverys, J. P., Martin, B., and Polard, P. (2009). The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol. Rev.* 33, 643–656.

Cohan, F. M., and Perry, E. B. (2007). A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17, R373–R386.

Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28, 1481–1489.

Colson, P., de Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* 55, 321–332.

Csuros, M., and Miklos, I. (2009). Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* 26, 2087–2095.

Dagan, T. (2011). Phylogenomic networks. *Trends Microbiol.* 19, 483–491.

Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U.S.A.*105, 10039–10044.

Dagan, T., and Martin, W. (2006). The tree of one percent. *Genome Biol.* 7, 118.

Dagan, T., and Martin, W. (2009). Getting a better picture of microbial evolution en route to a network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2187–2196.

Darwin, C. (1858). On the tendency of species to form varieties; and on the perpetuation of vareities and species by natural means of selection. I. Extract from an unpublished work on species, II. Abstract of a letter from Darwin, C., esq., to Prof. Asa Gray. *J. Proc. Linn. Soc. Lond.* 3, 45–53.

Darwin, C. (1859). *On the Origin of Species*. London: John Murray.

Darwin, C. (1872). *Origin of Species*. New York, NY: The Modern Library.

de Jong, I. G., Haccou, P., and Kuipers, O. P. (2011). Bet hedging or not? A guide to proper classification of microbial survival strategies. *Bioessays* 33, 215–223.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., Bruggemann, H., Lienard, T., Christmann, A., Bomeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H. P., Gunsalus, R. P., Fritz, H. J., and Gottschalk, G. (2002). The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* 4, 453–461.

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* 190, 1390–1400.

Devos, K. M., Ma, J., Pontaroli, A. C., Pratt, L. H., and Bennetzen, J. L. (2005). Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. *Proc. Natl. Acad. Sci. U.S.A.* 102, 19243–19248.

Dobzhansky, T. (1937). *Genetics and the Origin of Species.* New York, NY: Columbia University Press.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.*104, 2043–2049.

Doolittle, W. F., and Zhaxybayeva, O. (2009). On the origin of prokaryotic species. *Genome Res.* 19, 744–756.

Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., Simons, R. W., Zimmerly, S., and Miller, J. F. (2004). Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* 431, 476–481.

Draghi, J., and Wagner, G. P. (2008). Evolution of evolvability in a developmental model. *Evolution* 62, 301–315.

Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. U. S.A.* 89, 6015–6019.

Dubnau, D., and Losick, R. (2006). Bistability in bacteria. *Mol. Microbiol.* 61, 564–572.

Earl, D. J., and Deem, M. W. (2004). Evolvability is a selectable trait. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11531–11536.

Edwards, R. A., and Rohwer, F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3, 504–510.

Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011). Genomic anatomy of Escherichia coli O157, H7 outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20142–20147.

Esser, C., Martin, W., and Dagan, T. (2007). The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol. Lett.* 3, 180–184.

Ferriere, R., and Michod, R. E. (2011). Inclusive fitness in evolution. *Nature* 471, E6–E8. author reply E9–E10.

Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3, 329–341.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: The Clarendon Press.

Forterre, P., and Prangishvili, D. (2009). The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann. N.Y. Acad. Sci.* 1178, 65–77.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., and Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323, 741–746.

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway de Macario, E. Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532–542.

Galhardo, R. S., Hastings, P. J., and Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol.* 42, 399–435.

Galhardo, R. S., and Rosenberg, S. M. (2009). Extreme genome repair. *Cell* 136, 998–1000.

Gardner, S. P., and Olson, J. W. (2012). Barriers to horizontal gene transfer in *Campylobacter jejuni*. *Adv. Appl. Microbiol.* 79, 19–42.

Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer,

genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687.

Gordo, I., and Charlesworth, B. (2000). The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics* 154, 1379–1387.

Gribaldo, S., and Brochier, C. (2009). Phylogeny of prokaryotes: does it exist and why should we care? *Res. Microbiol.* 160, 513–521.

Guo, H., Tse, L. V., Barbalat, R., Sivaamnuaiphorn, S., Xu, M., Doulatov, S., and Miller, J. F. (2008). Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol. Cell* 31, 813–823.

Hacker, J., and Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679.

Haeckel, E. (1997). *The Wonders of Life: A Popular Study of Biological Philosophy*. London: General Books, LLC.

Haldane, J. B. S. (1932). *The Causes of Evolution*. London, New York, Toronto: Longmans, Green and Co.

Hamilton, H. L., and Dillard, J. P. (2006). Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Mol. Microbiol.* 59, 376–385.

Hamilton, H. L., Dominguez, N. M., Schwartz, K. J., Hackett, K. T., and Dillard, J. P. (2005). Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system. *Mol. Microbiol.* 55, 1704–1721.

Hayes, F., and Van Melderen, L. (2011). Toxins-antitoxins: diversity, evolution and function. *Crit. Rev. Biochem. Mol. Biol.* 46, 386–408.

Hedges, D. J., and Batzer, M. A. (2005). From the margins of the genome: mobile elements shape primate evolution. *Bioessays* 27, 785–794.

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806.

Johnsborg, O., and Havarstein, L. S. (2009). Regulation of natural genetic transformation and acquisition of transforming DNA in Streptococcus pneumoniae. *FEMS Microbiol. Rev.* 33, 627–642.

Jones, A. G., Arnold, S. J., and Burger, R. (2007). The mutation matrix and the evolution of evolvability. *Evolution* 61, 727–745.

Jordan, I. K., Rogozin, I. B., Glazko, G. V., and Koonin, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 19, 68–72.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393.

Kazazian, H. H. Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632.

Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8420–8427.

Klasson, L., and Andersson, S. G. (2004). Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* 12, 37–43.

Kloesges, T., Popa, O., Martin, W., and Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.* 28, 1057–1074.

Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1929–1940.

Koonin, E. V. (2003). Horizontal gene transfer: the path to maturity. *Mol. Microbiol.* 50, 725–727.

Koonin, E. V. (2009a). Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37, 1011–1034.

Koonin, E. V. (2009b). Evolution of genome architecture. *Int. J. Biochem. Cell Biol.* 41, 298–306.

Koonin, E. V. (2010). The two empires and three domains of life in the postgenomic age. *Nat. Educ.* 3, 27.

Koonin, E. V. (2011). *The Logic of Chance: The Nature and Origin of Biological Evolution*. Upper Saddle River, NJ: FT Press.

Koonin, E. V., Makarova, K. S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* 55, 709–742.

Koonin, E. V., and Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.* 1, 95.

Koonin, E. V., and Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends Genet.* 21, 647–654.

Koonin, E. V., Puigbo, P., and Wolf, Y. I. (2011). Comparison of phylogenetic trees and search for a central trend in the "forest of life". *J. Comput. Biol.* 18, 917–924.

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2006). The ancient Virus World and evolution of cells. *Biol. Direct* 1, 29.

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2009). Compelling reasons why viruses are relevant for the origin of cells. *Nat. Rev. Microbiol.* 7, 615.

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719.

Koonin, E. V., and Wolf, Y. I. (2009a). Is evolution Darwinian or/and Lamarckian? *Biol. Direct* 4, 42.

Koonin, E. V., and Wolf, Y. I. (2009b). The fundamental units, processes and patterns of evolution, and the Tree of Life conundrum. *Biol. Direct* 4, 33.

Kruger, N. J., and Stingl, K. (2011). Two steps away from novelty–principles of bacterial DNA uptake. *Mol. Microbiol.* 80, 860–867.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kuhn, T. S. (2002). *The Road since Structure*. Chicago, IL: University of Chicago Press.

Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C. A. (2005). The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* 15, 954–959.

Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327.

Lamarck, J.-B. (1809). *Philosophie Zoologique, ou Exposition des Considérations Relatives à l'histoire Naturelle des Animaux*. Paris: Dentu.

Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482.

Larsson, J., Nylander, J. A., and Bergman, B. (2011). Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol. Biol.* 11, 187.

Leigh, E. G. Jr. (2010). The group selection controversy. *J. Evol. Biol.* 23, 6–19.

Libby, E., and Rainey, P. B. (2011). Exclusion rules, bottlenecks and the evolution of stochastic phenotype switching. *Proc. Biol. Sci.* 278, 3574–3583.

Lisch, D., and Bennetzen, J. L. (2011). Transposable element origins of epigenetic gene regulation. *Curr. Opin. Plant Biol.* 14, 156–161.

Lopez, D., Vlamakis, H., and Kolter, R. (2009). Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiol. Rev.* 33, 152–163.

Low, K. B., and Porter, D. D. (1978). Modes of gene transfer and recombination in bacteria. *Annu. Rev. Genet.* 12, 249–287.

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349.

Lynch, M. (2011). The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* 3, 1107–1118.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J., and Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D. M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J. H., Diaz-Muniz, I., Dosti, B., Smeianov, V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B., and Mills, D. (2006). Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15611–15616.

Makarova, K. S., Ponomarev, V. A., and Koonin, E. V. (2001). Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol.* 2, RESEARCH 0033.

Makarova, K. S., Sorokin, A. V., Novichkov, P. S., Wolf, Y. I., and Koonin, E. V. (2007). Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct* 2, 33.

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2009). Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* 4, 19.

Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel

defense systems. *J. Bacteriol.* 193, 6039–6056.

Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845.

Marraffini, L. A., and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181–190.

Martinez-Borra, J., Gonzalez, S., and Lopez-Larrea, C. (2012). The origin of the bacterial immune response. *Adv. Exp. Med. Biol.* 738, 1–13.

Martin, W. F. (2011). Early evolution without a tree of life. *Biol. Direct* 6, 36.

Maynard Smith, J. (1998). The units of selection. *Novartis Found. Symp.* 213, 203–11. discussion: 211–217.

Mayr, E. (1944). *Systematics and the Origin of Species*. New York, NY: Columibia University Press.

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26.

McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., and Paul, J. H. (2010). High frequency of horizontal gene transfer in the oceans. *Science* 330, 50.

Medhekar, B., and Miller, J. F. (2007). Diversity-generating retroelements. *Curr. Opin. Microbiol.* 10, 388–395.

Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences. *Mol. Biol. Evol.* 28, 3213–3223.

Mirkin, B. G., Fenner, T. I., Galperin, M. Y., and Koonin, E. V. (2003). Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* 3, 2.

Moreira, D., and Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7, 306–311.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399, 323–329.

Novichkov, P. S., Omelchenko, M. V., Gelfand, M. S., Mironov, A. A., Wolf, Y. I., and Koonin, E. V. (2004). Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. *J. Bacteriol.* 186, 6575–6585.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature* 466, 1057–1062.

O'Malley, M. A. (2012). Evolutionary systems biology: historical and philosophical perspectives on an emerging synthesis. *Adv. Exp. Med. Biol.* 751, 1–28.

O'Malley, M. A., and Koonin, E. V. (2011). How stands the Tree of Life a century and a half after The Origin? *Biol. Direct* 6, 32.

O'Malley, M. A., and Soyer, O. S. (2012). The roles of integration in molecular systems biology. *Stud. Hist. Philos. Biol. Biomed. Sci.* 43, 58–68.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740.

Pace, N. R. (2006). Time for a change. *Nature* 441, 289.

Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* 73, 565–576.

Paul, J. H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* 2, 579–589.

Perez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., Silva, F. J., Moya, A., and Latorre, A. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science* 314, 312–313.

Perna, N. T., Plunkett, G. 3rd., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157, H7. *Nature* 409, 529–533.

Piggot, P. (2010). Epigenetic switching: bacteria hedge bets about staying or moving. *Curr. Biol.* 20, R480–R482.

Pigliucci, M. (2008). Is evolvability evolvable? *Nat. Rev. Genet.* 9, 75–82.

Piriyapongsa, J., Marino-Ramirez, L., and Jordan, I. K. (2007). Origin and evolution of human microRNAs from transposable elements. *Genetics* 176, 1323–1337.

Ponder, R. G., Fonville, N. C., and Rosenberg, S. M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol. Cell* 19, 791–804.

Poole, A. M., Phillips, M. J., and Penny, D. (2003). Prokaryote and eukaryote evolvability. *Biosystems* 69, 163–185.

Pritham, E. J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390, 3–17.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a Tree of Life in the thicket of the phylogenetic forest. *J. Biol.* 8, 59.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* 2, 745–756.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2012). Genome-wide comparative analysis of phylogenetic trees: the prokaryotic forest of life. *Methods Mol. Biol.* 856, 53–79.

Rainey, P. B., Beaumont, H. J., Ferguson, G. C., Gallie, J., Kost, C., Libby, E., and Zhang, X. X. (2011). The evolutionary emergence of stochastic phenotype switching in bacteria. *Microb. Cell Fact.* 10(Suppl. 1), S14.

Raoult, D. (2009). There is no such thing as a tree of life (and of course viruses are out!). *Nat. Rev. Microbiol.* 7, 615.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* 6, 315–319.

Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* 459, 207–212.

Rose, M. R., and Oakley, T. H. (2007). The new biology: beyond the Modern Synthesis. *Biol. Direct* 2, 30.

Rosenberg, S. M., and Hastings, P. J. (2003). Microbiology and evolution. Modulating mutation rates in the wild. *Science* 300, 1382–1383.

Sicheritz-Ponten, T., and Andersson, S. G. (2001). A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* 29, 545–552.

Simpson, G. G. (1944). *Tempo and Mode in Evolution*. New York, NY: Columbia University Press.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244.

Smith, J. M., Smith, N. H., O'Rourke, M., and Spratt, B. G. (1993). How clonal are bacteria? *Proc. Natl. Acad. Sci. U.S.A.* 90, 4384–4388.

Snel, B., Bork, P., and Huynen, M. A. (1999). Genome phylogeny based on gene content. *Nat. Genet.* 21, 108–110.

Snel, B., Bork, P., and Huynen, M. A. (2002). Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 12, 17–25.

Snel, B., Huynen, M. A., and Dutilh, B. E. (2005). Genome trees and the nature of genome evolution. *Annu. Rev. Microbiol.* 59, 191–209.

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K. E. (2008). Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res. Microbiol.* 159, 349–357.

Stanier, R. Y., and Van Niel, C. B. (1962). The concept of a bacterium. *Arch. Mikrobiol.* 42, 17–35.

Stern, A., and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays* 33, 43–51.

Strassmann, J. E., Page, R. E. Jr., Robinson, G. E., and Seeley, T. D. (2011). Kin selection and eusociality. *Nature* 471, E5–E6. author reply E9–E10.

Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361.

Suttle, C. A. (2007). Marine viruses–major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812.

Sutton, M. D., Smith, B. T., Godoy, V. G., and Walker, G. C. (2000). The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu. Rev. Genet.* 34, 479–497.

Syvanen, M. (1985). Cross-species gene transfer; implications for a new theory of evolution. *J. Theor. Biol.* 112, 333–343.

Syvanen, M. (1994). Horizontal gene transfer: evidence and possible consequences. *Annu. Rev. Genet.* 28, 237–261.

Syvanen, M., and Kado, C. I. (2002). *Horizontal Gene Transfer*. San Diego, CA: Academic Press.

Szathmary, E., and Maynard Smith, J. (1997). From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* 187, 555–571.

Takeuchi, N., and Hogeweg, P. (2009). Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. *PLoS Comput. Biol.* 5:e1000542. doi: 10.1371/journal.pcbi.1000542

Takeuchi, N., and Hogeweg, P. (2012). Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. *Phys. Life Rev*. doi: 10.1016/j.plrev.2012.06.001. [Epub ahead of print].

Takeuchi, N., Hogeweg, P., and Koonin, E. V. (2011). On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput. Biol.* 7:e1002024. doi: 10.1371/journal.pcbi.1002024

Tax, S., and Callender, C. (1960). *Evolution after Darwin; the University of Chicago Centennial.* Chicago, IL: University of Chicago Press.

van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., and Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34, 401–407.

Van Melderen, L., and Saavedra De Bast, M. (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet.* 5:e1000437. doi: 10.1371/journal.pgen.1000437

Veening, J. W., Smits, W. K., and Kuipers, O. P. (2008a). Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* 62, 193–210.

Veening, J. W., Stewart, E. J., Berngruber, T. W., Taddei, F., Kuipers, O. P., and Hamoen, L. W. (2008b). Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4393–4398.

Wallace, A. R. (1858). On the tendency of species to form varieties; an don the perpetuation of varieties and species by natural mean of selection. III. On the tendency of varieties to depart indefinitely from the original type. *J. Proc. Linn. Soc. London* 3, 53–62.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., and Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* 1, 147.

Weinbauer, M. G., and Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* 6, 1–11.

Wellner, A., Lurie, M. N., and Gophna, U. (2007). Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.* 8, R156.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271.

Woese, C. R. (2004). The Archaeal Concept and the World it Lives in: A Retrospective. *Photosynth. Res.* 80, 361–372.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090.

Woese, C. R., and Goldenfeld, N. (2009). How the microbial world saved evolution from the scylla of molecular biology and the charybdis of the modern synthesis. *Microbiol. Mol. Biol. Rev.* 73, 14–21.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579.

Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. (2000). Aminoacyl-tsynthetases, RNA, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* 64, 202–236.

Wolf, Y. I., Aravind, L., Grishin, N. V., and Koonin, E. V. (1999). Evolution of aminoacyl-tRNA synthetases–analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome. Res.* 9, 689–710.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. (2002). Genome trees and the tree of life. *Trends Genet.* 18, 472–479.

Yutin, N., Puigbo, P., Koonin, E. V., and Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* 7:e36972. doi: 10.1371/journal.pone.0036972

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A. K., Taboada, E., and Gannon, V. P. (2007). Genome evolution in major *Escherichia coli* O157, H7 lineages. *BMC Genomics* 8, 121.

Zhaxybayeva, O. (2009). Detection and quantitative assessment of horizontal gene transfer. *Methods Mol. Biol.* 532, 195–213.

Zhaxybayeva, O., and Doolittle, W. F. (2011). Lateral gene transfer. *Curr. Biol.* 21, R242–R246.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 August 2012; accepted: 27 August 2012; published online: 13 September 2012.

*Citation: Koonin EV and Wolf YI (2012) Evolution of microbes and viruses: a paradigm shift in evolutionary biology? Front. Cell. Inf. Microbio. **2**:119. doi: 10.3389/fcimb.2012.00119*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 28 August 2012
doi: 10.3389/fcimb.2012.00113

Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin

Vicky Merhej and Didier Raoult*

URMITE, UM63, CNRS 7278, IRD 198, INSERM U1095, Aix Marseille Université, Marseille, France

Edited by:
Eugene V. Koonin, National Institutes of Health, USA

Reviewed by:
James E. Graham, University of Louisville School of Medicine, USA
Jose Vazquez-Boland, University of Edinburgh, UK

***Correspondence:**
Didier Raoult, Faculté de Médecine, URMITE, UMR CNRS 7278, IRD 198, INSERM U1095, 27 Bd Jean Moulin, 13385 Marseille cedex 5, France.
e-mail: didier.raoult@gmail.com

Darwin's theory about the evolution of species has been the object of considerable dispute. In this review, we have described seven key principles in Darwin's book *The Origin of Species* and tried to present how genomics challenge each of these concepts and improve our knowledge about evolution. Darwin believed that species evolution consists on a positive directional selection ensuring the "survival of the fittest." The most developed state of the species is characterized by increasing complexity. Darwin proposed the theory of "descent with modification" according to which all species evolve from a single common ancestor through a gradual process of small modification of their vertical inheritance. Finally, the process of evolution can be depicted in the form of a tree. However, microbial genomics showed that evolution is better described as the "biological changes over time." The mode of change is not unidirectional and does not necessarily favors advantageous mutations to increase fitness it is rather subject to random selection as a result of catastrophic stochastic processes. Complexity is not necessarily the completion of development: several complex organisms have gone extinct and many microbes including bacteria with intracellular lifestyle have streamlined highly effective genomes. Genomes evolve through large events of gene deletions, duplications, insertions, and genomes rearrangements rather than a gradual adaptative process. Genomes are dynamic and chimeric entities with gene repertoires that result from vertical and horizontal acquisitions as well as *de novo* gene creation. The chimeric character of microbial genomes excludes the possibility of finding a single common ancestor for all the genes recorded currently. Genomes are collections of genes with different evolutionary histories that cannot be represented by a single tree of life (TOL). A forest, a network or a rhizome of life may be more accurate to represent evolutionary relationships among species.

Keywords: catastrophes, Darwin, gene creation, giant viruses, micorbial genomics, rhizome of life, sequence exchange

INTRODUCTION

The theory of evolution became a subject of deep reflection toward the end of the twentieth century. The development of the theory of evolution has benefited from the contributions of several authors, including Lamarck and Darwin (Koonin and Wolf, 2009). Their findings have been subjected to intense criticism. Indeed, their claim that all living species were transformed over time to give rise to new species was much to the dismay of the creationists (the equivalent of the "fixistes" in France) who believed that each species was created once and for all and that no species had disappeared since the creation. This latter perception of the worlds is a synthesis between the Socratic Greek philosophy, the harmonious cosmos and the essentialism of Plato (427–327 BCE) and Aristotle (384–322 BCE) on one hand and the Christians' view of the world's creation as described in the bible on the other hand. In contrast, the monistic view of Heraclitus (535–475 BCE), the constant motion of Democritus (460–370 BCE) and the dynamic theory of atomic motion described by Lucretius (94?–55 BCE) considered life to be an interplay of physical-chemical forces immanent to matter and in which living things live in perpetual motion. In this context, Lucretius' Epicurean poem, *De rerum natura*, postulated the extinction of species that are not well suited to surviving and reproducing successfully (Lucretius, 1995).

Darwin developed a highly disputed theory that was largely influenced by the works of Buffon on transformism (de Buffon, 1753), the concept of the differential fertility of Malthus (Malthus, 1798; Barlow, 1958) and the gradualism of Leibniz (Leibniz, 1996). Darwin proposed a straightforward mechanism of evolution that involves an interplay between heritable variation and natural selection, collectively described as the survival of the fittest. Under Darwin's concept, the material for evolution is provided by heritable random variation; natural selection is the main driving force of evolution, which introduces order and produces increasingly complex adaptive features of organisms. Darwin thought of natural selection in terms of the fixation of

beneficial changes, i.e., evolutionarily relevant mutations. These beneficial changes have infinitesimally small effects on fitness, and, as a result, evolution occurs via numerous, successive and slight modifications according to the theory of *strict gradualism*. Finally, Darwin suggested that all life forms evolved from a single common ancestor (Darwin, 1859). Indeed, based on his observations on the evolution of animals, Darwin attempted to issue a general theory about the evolution of life. He proposed that the relationships among all species resemble a tree, the Tree of Life (TOL), in which all living organisms are considered to have descended from a single ancestor (Darwin, 1859).

Darwin's theory was later the object of considerable dispute, particularly because Darwin was unaware of Mendel's work and of the importance of genetics for understanding evolution (Charlesworth and Charlesworth, 2009). Fisher, Haldane, Dobzhansky, Wright and Mayr, among many others, integrated genetics, paleontology, systematics, and cytology within a newly expanded structure of biological thought that is often referred to as "the modern Synthesis" (Huxley, 1942; Koonin, 2009d). The modern synthesis provided useful foundations for biological thought, including the idea that changes in genotype, the genetic material, precede changes in the phenotype, which determines the appearance of an individual. The modern synthesis framework provided many fundamental insights into evolutionary biology, especially with regards to the main topic of Darwin's famous book, *The Origin of Species* (Darwin, 1859). Darwin thought that species were the result of the human predilection to perceive discontinuity among continuously varying individuals. Mayr's extensive knowledge about variation in morphology, overlain with an understanding of the biogeographic distributions of bird species, led him to develop the biological species concept. Mayr explained the geographic mechanisms of speciation and insisted that the geographic separation of populations that prohibits a homogenizing gene flow between them leads to the divergence of such populations and to reproductive isolation. Based on these concepts, Mayr defined allopatric speciation as, the process of the evolution of geographically isolated populations into distinct species, and sympatric speciation as, the evolution of new species inhabiting the same geographic region (Mayr, 1944, 1963).

In this paper we outline the changes brought about by comparative genomics and phylogenetic studies as determined by Koonin to the concepts of evolution proposed by Darwin (Koonin, 2009a) (**Table 1**). We have identified seven key principles in *The Origin of Species* (Darwin, 1859): (1) the concept of evolution according to a positive directional selection that favors advantageous mutations to increase fitness, (2) the struggle for existence, (3) the complexity associated with development, (4) gradualism and progressive evolution, (5) strict vertical inheritance, (6) a single common ancestor, and (7) the TOL. We discuss these points to identify the hypotheses that survive critical analysis and respect current knowledge. We attempt to highlight the influence of microbial genomics on our understanding of the evolution of genetic repertoires.

NATURAL SELECTION

The question of our origins has always fascinated humans. From the earliest times, the existence of life has typically been attributed to supernatural intervention. Naturalistic models of origins based on logic and philosophy can be traced to approximately the fifth century BC in Greece at the time of the pre-Socratic philosophers and scientists (Anaximander, Heraclitus, Empedocles, Parmenides, Zeno, Democritus…). Anaximander argued that life originated in the sea and deduced living beings gradually developed, from moisture and warmth. He further proposed that the first human, in the form known today, originated from animals of another sort (Barnes, 1983). Empedocles claimed that living creatures might have originated by chance (Barnes, 1983). In contrast, the process of development was denied by the philosophers Plato and Aristotle. These philosophers denied any continuous change of ideas or forms, i.e., the forms, or archetypal ideas, remain eternally what they are. Thus, evolution was considered by Plato and Aristotle to be a general trend in which

Table 1 | Darwin's propositions in the face of evolutionary genomics.

Darwin's proposition	Genomic challenge
The general trend of evolution is the fixation of beneficial changes	Natural selection is one of the evolutionary forces. However, random selection is largely produced by catastrophic stochastic processes
According to the principle of the "survival of the fittest," organisms evolve toward the most well-adapted state	Genomes contain many genes that do not increase the fitness and genes that are not required for the survival in current ecosystems
The general trend of evolution leads to complex adaptive organisms	Complex organisms represent very small part of living species. Several complex organisms have completely disappeared
Organisms evolve through the gradual fixation of infinitesimally small variations by natural selection	Genomes evolve through large events of gene deletions and duplications and insertions and genomes rearrangements. Evolution rarely follows a gradual adaptative process
Organisms evolve through vertical inheritance of ancestral characters	The gene repertoire results from vertical and horizontal acquisitions as well as *de novo* gene creation
All cellular life forms have one common ancestor	The chimeric character of the genomes excludes the possibility of finding a single common ancestor for all the genes recorded currently
The evolution of life can be depicted as a single tree (TOL)	Genomes are collections of genes with different evolutionary histories that cannot be represented by a single tree of life

everything in nature has a certain order or purpose. The physical world is wholly dominated by purpose (Aristotle, 2008). Aristotle developed a "*scala* of *naturae*," a great chain of being, in which he arranged all beings on a ladder beginning with inanimate matter and climbing to plants, invertebrates, and vertebrates. Among the vertebrates, Aristotle placed the fish at the lowest rung of the ladder and humans on the highest rung. This scale of nature is a graded scale of perfection that represented a continual progression from simple and undeveloped organisms to the complex and more perfect organisms (Singer, 1931; Mayr, 1982).

Darwin was not the first to describe the origin of species as one from another as a formal doctrine. In addition to the Greeks mentioned above, Lamarck denied the immutability of species and forms and claimed to have demonstrated by observation the gradual development of the animal kingdom (Lamarck, 1809). What is new in Charles Darwin's work is not his theory of descent but its confirmation by the theory of natural selection and the survival of the fittest in the struggle for existence. The major contribution of Darwin to the idea of selection can be summarized with the words "chance and necessity." According to Darwin, changes in the genome occur by chance and are maintained if necessary (natural selection). The resulting genomic repertoire corresponds to a rational end in a purely mechanical process without any cooperation of teleological principles and without any innate tendency of the organisms to proceed to a higher stage. This theory postulates that the later organisms deviated from the earlier ones and that these deviations, in so far as they are improvements, perpetuate themselves, and become generic marks of differentiation. Interestingly, the words "chance and necessity" were used for the first time by Democritus, who ascribed the causes of things either to necessity or to chance and the absence of purpose (Barnes, 1983). Democritus showed that apparently orderly effects can be produced without goal-oriented forces or purpose. Nietzsche prone the realm of chance "Those iron hands of necessity which shake the dice-box of chance, lay their game for an infinite length of time …" (Nietzsche, 2006). Similarly, Jacques Monod [winner of the Nobel Prize in Physiology or Medicine (1965)], in his famous book "*Chance and Necessity*" (Monod, 1972), described the structural teleonomy of living organisms with apparent intended goals and refuted the idea of purpose in nature.

Charles Lyell, a famed geologist and paleontologist, befriended Charles Darwin and strongly influenced his thought. Lyell's interpretation of geologic change prompted Darwin to think of evolution as a slow process in which small changes gradually accumulate over immense spans of time. Lyell had shown how gigantic valleys had been formed by gradual erosion; similarly, Darwin believed that natural selection occurred through the preservation and accumulation of a great number of infinitesimally small inherited modifications (Lyell, 1830). This theory intentionally ignored the catastrophic (chaotic) events (such as earthquakes) that the creationists used to explain evolution and the presence of fossils by defining fossils as living beings that coexisted with current living beings but that had disappeared under the impact of a disaster. These events drastically reduced population size and resulted in genetic drift. Indeed, more than 99% of a population might be killed by disasters, allowing only a few genetic features to be selected. This random selection occurs frequently in microbiology especially in the digestive microbiota. The invasion of a new bacterial or a viral species can cause diarrhea, which can lead to the extermination of up to 10^{13} bacteria, archaea and bacteriophages; the use of a specific antibiotic treatment, such as metromidazole, can eradicate 90% of the population in a few days. Interestingly, these ecosystems are repopulated at considerable speed and contain new species particularly in the presence of antibiotics that prevent the revival of the original flora. Thus, we can observe the effect of disasters on microbial populations, and there is no reason the same types of disasters, less common but just as critical for evolution, have not affected all living things.

When considering the important role of catastrophic events in the selection of living beings, evolution more closely resembles a random process than a mechanism driving positive selection. Recent work particularly that of Abi Rached et al., has shown that humans are a mosaic of three currently known hominids: Cro Magnon, Denisovan and Neanderthal. It is likely that following a series of catastrophic accidents, some mixed populations survived in different parts of the world (Abi-Rached et al., 2011). Horses evolved between 54 million years ago to about 10,000 years ago, spreading throughout North America. Then, suddenly, without apparent reason, between 10,000 and 8000 years ago, *Equus* disappeared from North America. Various theories have been advanced including destruction by drought, disease, or extinction as a result of hunting by growing human populations. At any rate, the horse was gone, and the horse was not seen again on its native continent until the Spanish explorers brought horses by ship in the sixteenth century. In total, the elements that create a visible disaster bottleneck are likely key to the selection of species (Remington, 1889).

THE STRUGGLE FOR EXISTENCE

Darwin believed that each organism must fight for its existence to survive, as derived from Malthus' theory on the population (Malthus, 1798). Darwin noted that each generation tends to increase, a process that should produce an overabundance of beings in nature. However, space and food are limited. The permanent destruction of some living beings is therefore necessary because not every individual can survive; this results in fierce competition among individuals of the same species and among individuals of different species. Thus, the struggle for existence arises from the inherent limitations of an ecological environment and the increasing number of species. Natural selection is the result of a struggle for existence, what Darwin called "the survival of the fittest" (Darwin, 1859). Natural selection eliminates some lineages and supports the species best adapted to their environments. Favorable variations in terms of survival and reproduction tend to be preserved, and unfavorable ones are destroyed. In contrast to the well-accepted statement of Lucretius about the necessity of reproduction for a species to endure (Lucretius, 1995), the notion of a struggle for existence remains a debatable issue. Indeed, some mutations may alter the fitness without decreasing the ability to multiply and perpetuate.

According to Darwin, organisms always evolve toward the most well-adapted state; thus, nearly all components of the genome should have a beneficial function. However, in microbial

organisms, the gene content raises questions about the principle of the survival of the fittest. Bacterial species have a number of genes in common that compose the core genome, and partially shared and strain-specific genes; the total of these genes constitutes the pan-genome (Tettelin et al., 2005; Schoen et al., 2008). Some bacteria that live in an ecosystem with highly variable conditions and with many other bacteria (sympatric bacteria), such as *Escherichia coli* or *Pseudomonas aeruginosa*, have very broad pan-genomes. Other bacteria that live in ecosystems with very restricted physicochemical conditions and limited partners (living in allopatry) have much smaller pan-genomes (Moliner et al., 2010). Interestingly, sympatric species retain some unused genes that are expressed at low rates. This pool of genes is not required for survival in the current ecosystem but may become necessary after future changes in the ecosystem. Indeed, bacteria contain laterally transferred sequences of DNA that are generally nearly neutral to the recipient and exert no effect on its fitness (Gogarten and Townsend, 2005). Much of the bacterial genomes consists of selfish elements with no appreciable phenotypic effect and that function only to ensure their own self-preservation within the genome (Orgel et al., 1980; Ogata et al., 2000).

Evolution as described by Darwin is a process of unidirectional positive selection that favors advantageous variations and results in increased fitness. In microbiology, mutations in DNA gyrase and in RNA polymerase that confer resistance to antibiotics such as quinolones or rifampicin may allow a bacterium to persist in its environment and, thus, seems to illustrate the Darwinian adaptive evolution. However, in most cases, these changes are not accompanied by an increase in fitness, and the mutants are rapidly eliminated when the antibiotics are removed. Hence, the change is purely opportunistic and does not play a role in the long term. Indeed, as for antibiotic selection, the antibiotic resistance of a microorganism may be associated with a short term advantage and with loss of fitness at long term when the ecosystem changes (with no more antiobiotics). Microbial genomics shows that evolution is subject to random changes rather than governed by natural selection with the goal of increasing fitness. Indeed, stochastic and catastrophic elements can substantially reduce a population and leave only a few survivors. The proportion of those survivors can be so low that it is difficult to imagine that their survival is due to anything other than chance. For example, during a plane crash, the chances of a passenger surviving are not improved by any particular inherent genetic advantage. Population bottlenecks are an indiscriminate sampling process, and genetic drift is independent of positive selection (**Figure 1**). In the same way that the sampling of colored balls from an urn is not influenced by the color difference among the balls, the effect of the gene is irrelevant to evolution. In summary, even beneficial adaptations may be permanently eliminated by bottlenecks. The immediate effect of a population bottleneck is decreased genetic diversity. In the long-term, repeated population bottlenecks and the accumulation of deleterious alleles through random

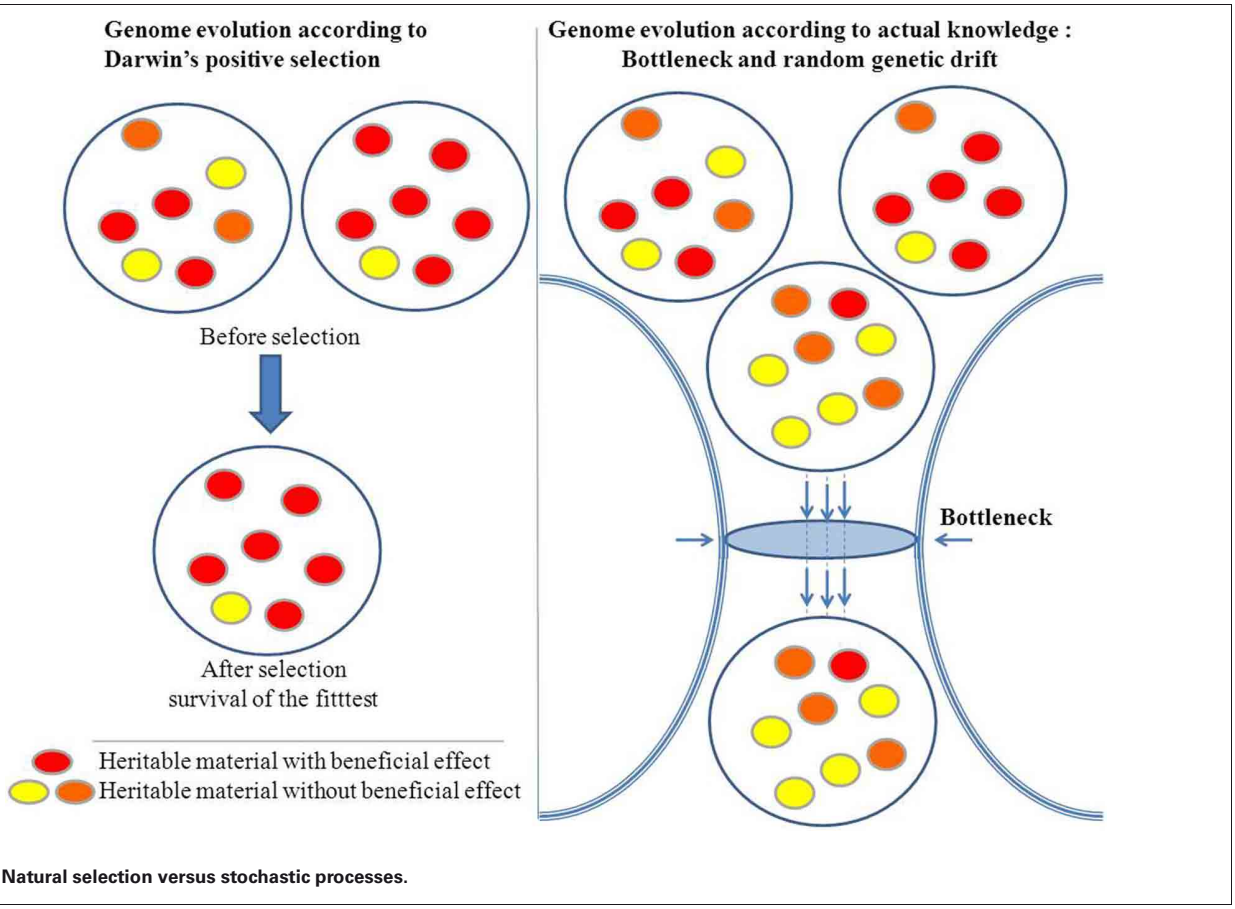


FIGURE 1 | Natural selection versus stochastic processes.

genetic drift in small populations can negatively affect their fitness (Ohta, 1973).

IMPROVEMENT AND COMPLEXITY

Darwin proposed a mechanism for the transformation of random variation into adapted, elaborate and complex devices that perform highly specific functions and increase the fitness of their carriers. Accordingly, the complex organisms, especially the multicellular eukaryotes (exhibiting large families of paralogous genes, the complicated regulation of gene expression, alternative splicing, and other genomic attributes) were considered more advanced and more successful than the simpler microbial organisms. However, genomics shows that the history of life is not a uniform trend for increasing complexity (Lynch, 2006). While, most eukaryotic taxa seem to have followed the route of junk recruitment, leading to complex organisms (Koonin, 2011c), different lineages, particularly in bacteria with an obligatory intracellular lifestyle, followed the route of genomic streamlining (Andersson and Kurland, 1998). Moreover, it has been assumed that organized multicellular life appeared with the "Cambrian explosion" some 600 million years ago. Interestingly, a group of 2.1-billion-year-old fossilized organisms (up to 12 cm) was recently found in Gabon (El Albani et al., 2010). This new discovery indicates that some large living things disappeared despite their size and complexity (**Figure 2**). Ultimately, genome size of present species has been revealed to be especially diverse across the different domains of life, ranging 1000-fold in viruses, bacteria and *Archaea* and 1,000,000-fold in eukarya as for the protists (**Figure 2**). This large diversity shows that evolution does not follow a unidirectional route towards increasing complexity.

One surprising outcome of analyzing the genome size is the lack of apparent correlation between the genome sizes and genetic and/or morphological complexity. This is the "C-value paradox," the C-value being a measure of genome size, typically expressed in base pairs of DNA per haploid genome (Thomas, 1971; Gregory, 2005). The C-value paradox implies that organisms with similar complexity may have very different genome sizes and conversely organisms with similar C-values may not be equally complex. Thus, the organism with the largest genome [and the largest number of open reading frames (ORFs)] is not necessarily the most complex. For example, the flagellated protist *Trichomonas vaginalis* has a genome of 160 Mb with ~60,000 protein-coding genes and many repetitive regions (up to 65% of the genome) (Carlton et al., 2007). Humans and mice have a genome size of around 3 billion base pairs whereas the unicellular protozoan *Amoeba dubia* has a genome size of around 700 billion base pairs, about 200 times as big (**Figure 2**). Indeed, genome size and number of genes cannot be used as a predictor of genetic

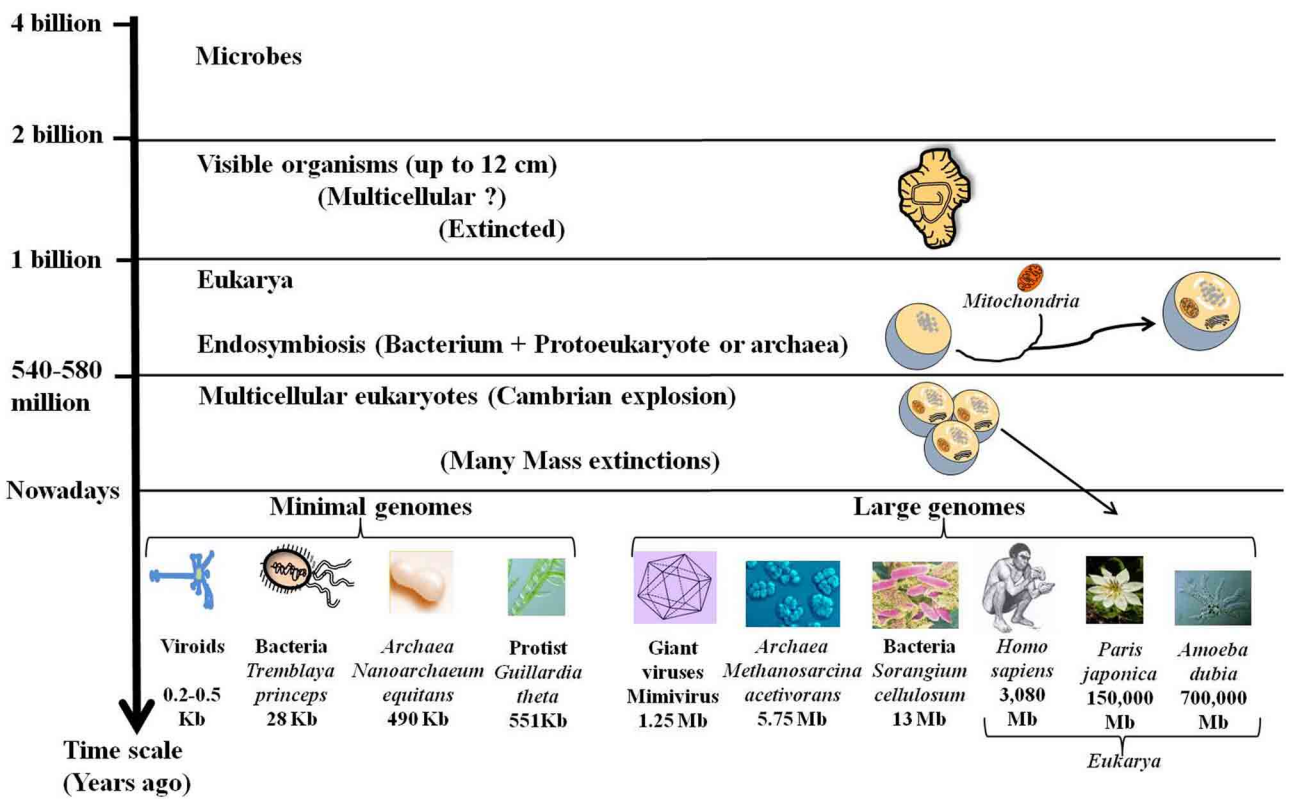


FIGURE 2 | Fossils, genome size, and complexity. Stratified evolution as deduced from fossils discoveries is characterized by the apparition and extinction of many organisms. Nowadays some complex organisms have disappeared while others are still present. These complex organisms with large genomes including giant viruses and large eukarya coexist with simple organisms like bacteria and viroids that are able to live and multiply despite their tiny genomes. The lack of correlation of genome size with organismal complexity confirms the C-value paradox.

or morphological complexity. While, the trend towards genome streamlining led to miniaturized genomes with a high-density of protein-coding DNA, most of DNA of large genomes is non-coding. The increase in gene number in multicellular species is accompanied by an expansion in the size and number of intragenic spacers and a proliferation of mobile genetic elements. The persistence of all of these sequences in the complex genomes may be due to an inefficient purifying selection in relation to the population size (Lynch, 2007). Genome complexity seems to be an indirect consequence of reduced effective population sizes accompanied by an increase in organism size. Therefore, the evolution of genome complexity may be a non-adaptive process that occurs in response to a reduction in the population size (Lynch and Conery, 2003).

Genomics is in total disagreement with the idea that species progress toward greater complexity and increasing fitness. Indeed, the notion of success and advances can be revisited in light of microbial genomic data. The compactness of microbial genomes and their widespread abundance in the biosphere highlight the power and competitiveness of simple and streamlined genomes (Koonin, 2011c). The most common bacteria on earth *Pelagibacter ubique* has a small genome (1.31 Mb). Likewise, the delta agent virus has a single genome of approximately 1700 bp and is capable of multiplying (Hughes et al., 2011) and viroids with genetic sequences of 150–500 bases of RNA represent the simplest known elements of life and are also able to spread, multiply, and cause diseases. Moreover, the most specialized bacterial species, those with an obligate intracellular lifestyle, are the most effective at a given time and in a given ecosystem, yet, they have lost their ability to adapt outside the host cell. This loss of adaptability is likely the cost of conserving a gene pool, especially in terms of translation, that slows multiplication. Therefore, by specializing in response to particular conditions, these microorganisms lose their ability to adapt to ecosystem changes. Bacteria in an optimized system are no longer able to adapt to other systems. This principle has been described by Pasteur for immunization. Indeed, the adaptation of *Pasteurella multocida* (the agent of fowl cholera) to a new ecosystem (axenic medium) resulted in the loss of their ability to multiply in their former ecosystem (chickens) (Pasteur, 1880).

The modern study of biology has shown that random processes result in perpetual change and ongoing evolution. As Heraclitus stated, nothing in the world, even for a moment, remain identical to itself: everything passes, everything changes, and everything dies every moment (Barnes, 1983). The term "evolution" implies progress; however, the observed changes do not necessarily correspond to an optimization. Recently, along with many other critics, Cathy Cox has proposed replacing the word evolution with the term "biological changes over time," which is a much more precise definition of the reality (http://www.georgiaencyclopedia.org/nge/Article.jsp?id=h-2622). This redefinition has resulted in major conflicts. Many biologists believe that these alternative terms exclude the idea of positive change and progress. In contrast, according to Gould, this progress is an illusion and is only a subjective interpretation of the statistics (Gould, 1984).

In practice, there is no goal of "progress" or "evolution" behind these biological changes. Some organisms are moving toward greater simplicity, whereas others become more complex without a general direction of evolution. Finally, the increasing morphological complexity does not go hand in hand with the gene repertoire complexity or with the increasing ability of adaption.

GRADUALISM

Darwin thought of evolution as a gradual accumulation of small changes. This proposal is a major component and one of the most controversial of Darwin's theories. He repeated several times the Latin phrase "natura non-falcit saltum" (nature does not make laps). The punctuationists believe that all are species have their own history, appearing and then disappearing, whereas gradualists consider species with much less interest, as a concept of convenience. Like Lamarck, Darwin believed that species changed gradually by undergoing changes and modifications over time without sharp changes. Because the evolutionarily relevant mutations are supposed to have infinitesimally small fitness effects, the Darwinian model of evolution inevitably leads to the concept of gradual progressive improvement (Darwin, 1859). This vision comes from his early training as a geologist who intentionally ignored disasters and catastrophic events in evolution. We know that this view is false in both geology and biology. Rather than small, gradual changes, massive events occur that affect living beings. Thus, because a gene must be present or absent to produce an inherited effect, Mendel assumed that the appearance of a new function would occur at once rather than gradually, as Darwin imagined. Later the zoologist Ernst Mayr showed that new species generally appears in geographic isolation and undergo a true "revolution" that rapidly transforms their gene pool. Studies on the frequency and geographical distribution of chronological horse fossils show that species evolution is not linear but consist of periods of stasis (gradual changes) interspersed with "crises," which lead to sudden extinctions and the appearance of new species. Indeed, different species could coexist with their original species while that ancestor remained unchanged, and there have even been reversals in evolutionary characteristics. These are all different evolutionary phenomena that explain the diversity of fossils and constitute a direct rebuttal to the principle of gradualism.

Moreover, Darwin's principle has been challenged by the Birth, Death, and Innovation Model of gene family evolution (Karev et al., 2002). In this model, duplication and lateral gene transfer give "birth" to new paralogous genes, "death" refers to gene elimination, and innovation corresponds to the acquisition of a new gene family via duplication and rapid evolution or via *de novo* creation. These events induce large and profound variations in genome size and gene repertoire (**Figure 3**). Thus, bacterial lineages that are specialized, including those with an obligatory intracellular lifestyle, show a repeated pattern of reduction in genome size through gene loss (Andersson and Kurland, 1998; Merhej et al., 2009). Bacterial genomes expand through lateral gene transfer and duplication. As a result, a considerable proportion (up to 14% of the ORFs) of most bacterial genomes consists of horizontally acquired genes (Nakamura et al., 2004).

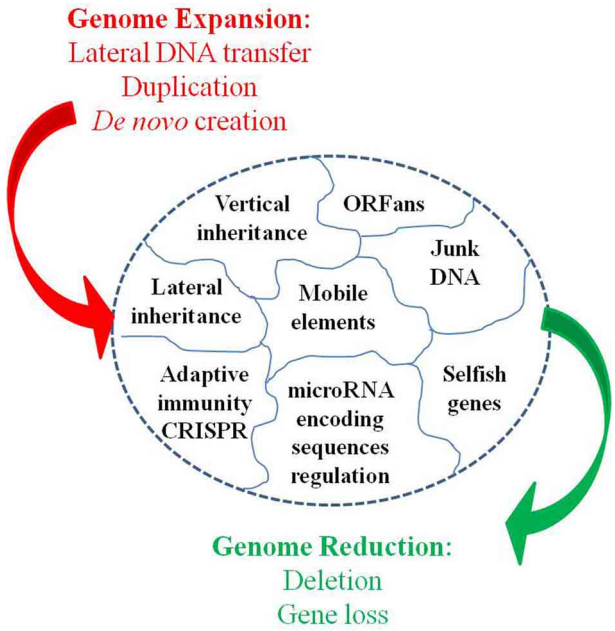


FIGURE 3 | Dynamic entity of the bacterial genomes.

Lateral transfer allows for the acquisition of xenobiotic functions (Treangen and Rocha, 2011). Lederberg's work in microbiology showed that these alterations can be transmitted in a heritable manner (Lederberg, 1949). Plasmids of several hundred kilobases can be transferred, as can bacteriophages, in bacteria. This phenomenon also occurs in eukaryotes. The virus HHV6 can integrate into the genome of humans and be transferred to their children (Arbuckle et al., 2010; Raoult, 2011). Additionally, the entire genome of the intracellular bacterium, *Wolbachia* was found to be integrated into the genome of its host (Dunning Hotopp et al., 2007; McNulty et al., 2010). Some of these inserted sequences are transcribed within eukaryotic cells, indicating that they may be functionally relevant to the evolution of the microbe's host. Finally, bacterial genomes exhibit a significant number of paralogous genes due to duplication (Fitch, 1970), ranging from 7% in *Rickettsia conorii* to 41% in *Streptomyces coelicolor* A3 (Gevers et al., 2004). Gene duplication represents an important path to the evolution of new biological functions via neofunctionalization (Ohno, 1970; Innan and Kondrashov, 2010). Clearly, loss, the lateral acquisition of genes, and the emergence of a new gene as a result of duplication or *de novo* creation are far from being "infinitesimal" changes, and if such large events occur, they are too abrupt so that the gradualist paradigm is not valid.

VERTICAL INHERITANCE

Darwin believed inheritance to be strictly vertical. In contrast Lamarck believed that the adaptation of a species to an ecosystem results in the acquisition of transferable characters. Like the "infective heredity" described by Lederberg, Lamarck insisted on the "inheritance of acquired characteristics"(Koonin and Wolf, 2009). Indeed, an organic modification acquired by an individual is genetically transmitted to offspring. In contrast, Darwin thought that traits were acquired by chance and not influenced by the environment. Natural selection retains the favorable changes *a posteriori*. This view has been challenged in a number of cases. First, the theory of "use it or lose it" holds that when genes are not used in a given ecosystem, they disappear (Moran, 2002). In this case, the phenotype precedes the genotype. Second, genomics have revealed the lateral acquisition of immunity in relation to the environment rather than by chance or vertically. Indeed, clustered regularly interspaced short palindromic repeats (CRISPRs) are found in the genomes of bacteria and archaea (Grissa et al., 2007; Horvath and Barrangou, 2010). These short sequences are stored in-between repeated elements; they function as acquired immunity genes against viruses (Weinberger et al., 2012). They can be transmitted to offspring allowing them to fight against the infection of viruses that have infected their ancestor in the past. Third, the high level of the transmission of sequences between organisms is particularly remarkable for the transmission of antibiotic resistance sequences between microorganisms. During the administration of a certain antibiotic, the sequence encoding for antibiotic resistance genes amplify by recombination or by duplication or by activating the expression and spread among different microorganisms. Moreover, some antibiotics may induce generalized transduction and help to propagate resistance genes (Rolain et al., 2011). The fourth challenge to vertical inheritance is the chimerical aspect of genomes, which will be developed in the next chapter. Finally, the vertical descent theory ignores the phenomenon of increase copy number and spread of repetitive DNA elements, like the selfish genes in a dynamic that usually has little or no benefit to the fitness of the organism.

Microbial genomes are not simply bags of faithfully inherited genes from an ancestor; rather, they are varied in their organization (Huynen and Bork, 1998). Bacterial genomes often exhibit high levels of plasticity and high levels of gene gain and loss during the evolution of species and strains. The genomes of closely related bacteria with different lifestyles showed remarkable variability with respect to gene content and gene order (Perna et al., 2001; Edwards et al., 2002). The microbial genomic architecture, or the arrangement of genes in a genome, exhibits great evolutionary instability (Koonin, 2009b). With the exception of the organization of small groups of functionally linked genes in operons, there is relatively little conservation of gene order, even among closely related organisms (Koonin, 2011a) (**Figure 4**). Various elementary mechanisms underlie the substantially dynamic character of genome evolution. Indeed, genome rearrangements such as inversions and translocations profoundly destroy the conservation of gene order. Moreover, recombination frequently occurs and generates sequence diversity by incorporating short DNA fragments (Feil et al., 2000; Hanage et al., 2005).

Comparative genomics shows large diversity in the gene repertoires among and within species. The genomes of obligate intracellular bacteria contain a subset of the genes present in their ancestors' larger genomes as the result of reductive evolution and gene loss (Merhej et al., 2009). The degree of genome flexibility correlates with the genomic content of repeated and

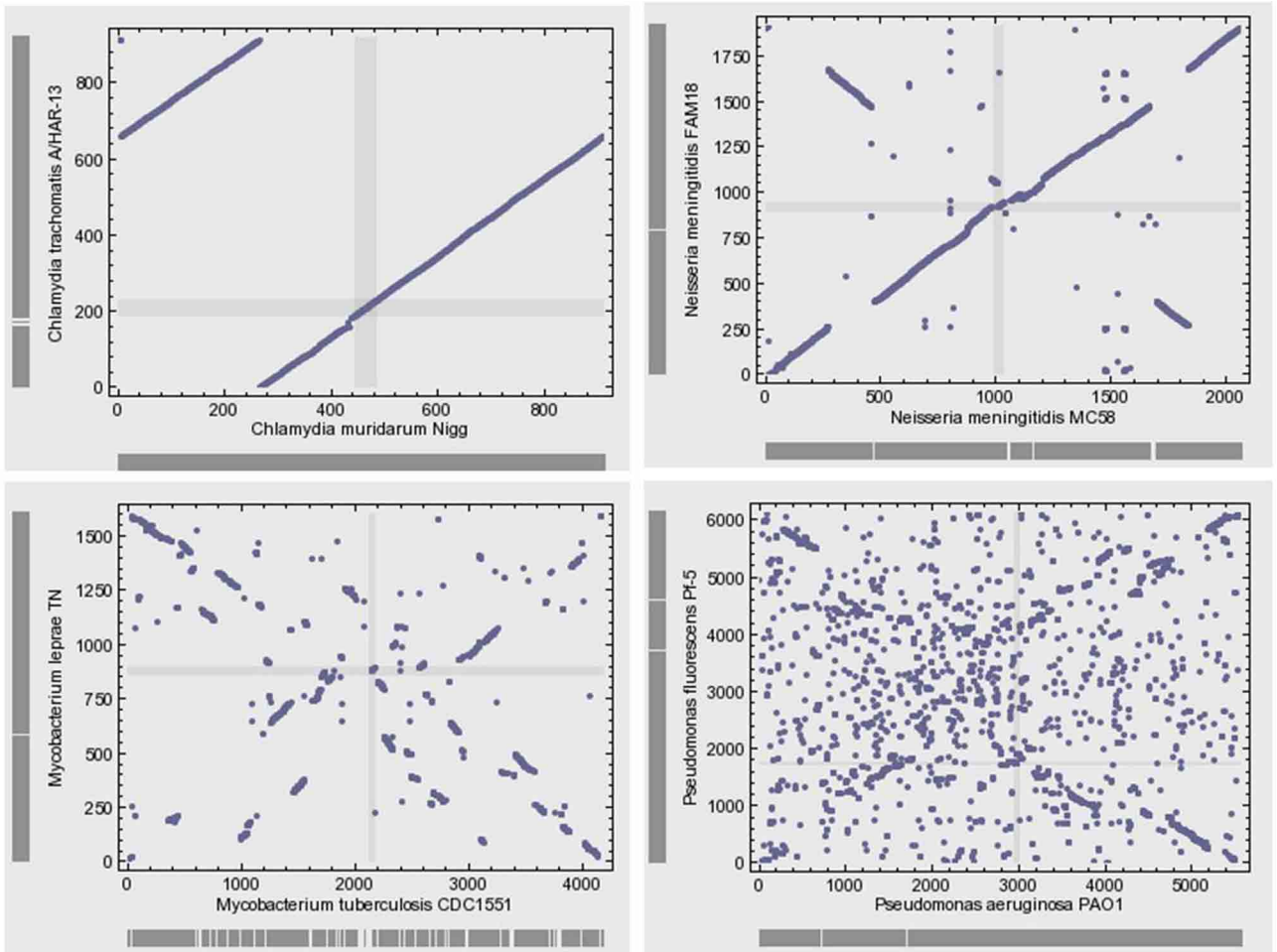


FIGURE 4 | Gene position plots of pairs of 4 selected genome pairs of bacteria show large variation between related species. This dynamic view of the genome rejects the evolution by infinitesimal variation of vertical inheritance. Each dot represents a pair of orthologous genes identified using the bidirectional best hit approach.

mobile sequences such as insertion sequence elements, plasmids and phages (Mira et al., 2002). The differential gene repertoires among closely related species is most likely due to gene transfer (Perna et al., 2001; Zhang et al., 2007) and recombination of repeated sequences (Tamas et al., 2002) rather than strict vertical inheritance and random variation as Darwin suggested. A substantial fraction of the differences in gene content is due to gene duplication (Zhang et al., 2007) and to the presence of ORFans, ORFs with no detectable sequence similarity to any other sequence in the databases (Fischer and Eisenberg, 1999); they correspond consequently to hypothetical or putative proteins. It has been proposed that the majority of ORFans are derived from bacteriophages (Daubin and Ochman, 2004). The variation in gene content often yields large pan-genomes (Tettelin et al., 2005; Schoen et al., 2008). Thus, the pan-genome of the genus *Streptococcus* likely exceeds by at least three times the average genome size of a typical *Streptococcus* species (Lefebure and Stanhope, 2007). The relationship between the pan-genome of a taxon and the gene content of a specific genome in the same taxon is far from being simple (**Figure 4**).

Lateral transfer has been viewed as a marginal phenomenon except for the transfer of pathogenicity islands (Perna et al., 2001; Juhas et al., 2009) and antibiotic resistance (Brisson-Noel et al., 1988; Shoemaker et al., 2001; Barlow, 2009). However, the analysis of multiple complete genome sequences became feasible, lateral transfer was revealed to play a major role in the evolution of microbial organisms (Lawrence and Retchless, 2009) especially by contributing to the metabolic innovation (Ochman et al., 2000). Oxygenic photosynthesis seems to have spread by lateral transfer (Mulkidjanian et al., 2006) via bacteriophages (Lindell et al., 2005). Thus, lateral transfer plays a major role in the biochemical diversity of microbial organisms and allows them to make up the vast majority of living cells on the planet and to be the principal agents in the biosphere. In contrast to the complexity hypothesis, genomics has shown that no gene is completely refractory to lateral transfer (Jain et al., 1999; Wellner et al., 2007). It was previously thought that informational genes were less prone to lateral transfer, but genomic analysis showed that genes essential for transcription and translation had also experienced multiple lateral transfers (Brochier et al., 2000; Merhej

et al., 2011). Lateral transfer affects the functions of genes to different extents, and all genes are susceptible to lateral transfer because the mechanism of transfer is random (Hao and Golding, 2008; Merhej et al., 2011). Moreover, lateral inheritance does not cleanly move a gene that is defined by start and end codons but rather involves DNA sequences that can be non-coding or include a single gene or a block of genes. It would thus be more accurate to talk of lateral sequence transfer (LST) than lateral gene transfer. Finally, the length of the lateral transfer can vary from a few bases (recombination) to multiple kilobases (Chan et al., 2009; Merhej et al., 2011). The analysis of the genomic sequence of *Wolbachia* demonstrates that LST occurs independently of the length of the sequence. Indeed, a small proportion and nearly the entire genome of *Wolbachia* in some cases, was found to be integrated into the genome of its host (Dunning Hotopp et al., 2007; McNulty et al., 2010).

Comparative genomics have shown that bacterial genomes are extremely heterogeneous and dynamic entities. The striking diversity in gene content and the flexibility of the genomic architecture challenge the theory of strict vertical acquisition and, to a greater extent the concept of species.

COMMON ANCESTOR

Darwin theorized that all extant life forms originated from a unique ancestor, which is now commonly referred to as the LUCA (last universal common ancestor). Koonin's seminal book vindicated Darwin's conjecture on the common origin of life and discussed the reconstruction of the gene repertoire of the LUCA (Koonin, 2011b). Indeed, comparative genomics revealed the universal conservation of hundreds of genes that are involved in gene expression and are thereby evidence in support of a common ancestral heritage (Koonin, 2003; Mirkin et al., 2003). The universally conserved features include the genetic code, i.e., the 64 codons that encode 20 amino acids and the stop signals; the three core subunits of the RNA polymerase; and the translation machinery composed of approximately 30 tRNAs, several translation factors, 18 amino-acyl-tRNA synthetases, and tRNA modification enzymes. Thus, by comparing the genes that present-day organisms have in common, evolutionary genomics indicate that the LUCA was a cellular organism with complete translation machinery, a core transcription system, and several metabolic pathways that included the genes required for purine and pyrimidine nucleotide biosynthesis.

The reconstruction of this ancestral cell is not plausible, because although the ancestor is primitive, its gene repertoire lacks key components that are essential for life (Mirkin et al., 2003). In particular, it is missing the genes necessary for DNA replication. Moreover, the idea of a common origin for all living beings faces substantial difficulties, including the lack of homology in the core DNA replication system components and the distinct enzymes required for lipid membrane biosynthesis in archaea and bacteria (Leipe et al., 1999; Pereto et al., 2004). As for the replication system, it has been hypothesized that the LUCA contained an RNA genome. The replacement of the RNA genome with a DNA genome and the appearance of the corresponding molecule systems would have occurred independently in the three domains of life—archaea, bacteria, and eukarya—after their divergence. Thus, the replication system was thought to have evolved in three distinct DNA viruses (prior to the existence of the DNA cell) and then transferred to the three life domains (Forterre, 1999, 2006). Another scenario is that a LUCA with a DNA genome underwent a subsequent replacement of its DNA-replication systems by non-homologous counterparts in the bacterial, archaeal, and eukaryotic lineages (Forterre, 2002). Finally, it has been suggested that a non-cellular LUCAS (last universal common ancestral state) existed as a pool of virus-like genetic elements in which the cellular key components originated. Archaea and bacteria might have independently emerged from the LUCAS, likely with numerous life forms now extinct (Koonin, 2009c). An alternative scenario postulates that the LUCA was a complex, protoeukaryotic lineage with an RNA genome present in a metabolically and morphologically heterogeneous community that gave rise to bacteria and archaea through differential gene loss (Glansdorff et al., 2008).

Multiple scenarios have been proposed to explain the origin of living beings. Regardless of which scenario is the most accurate, it has become obvious that the large diversity among species cannot be logically explained only by mutations that occurred on a unique ancestral genome ("descent with modification"). Likewise, the idea of a single mating pair at the origin of all human beings present on earth today cannot be accepted (Raoult, 2011). Several geneticists agree that "Eve" was not the only woman to conceive children who are ancestors of the current human population. Human evolution appears to be much more chimerical. Add to this, the theory of endosymbiosis showed that mitochondria were of bacterial origin from a species closely related to the Rickettsiales. Darwins reluctantly allowed for the principle of endosymbiosis but limited it to a single event suggesting that the exception does not undermine the principle of a common ancestor. However, we recently demonstrated that mitochondria were not the result of a single event but rather resulted from multiple events of gene transfer from different sources, leading to variation among organisms (Georgiades and Raoult, 2011). Mitochondria seem to have different bacterial origins, which are mainly, but not exclusively, from the group of Rickettsiales. Similarly, human beings are chimeras that contain retroviral DNA and many genes of bacterial and archaeal origin (Raoult, 2011). Genes from *Trypanosoma cruzi* are likely to integrate into the genome of infected patients and to be passed on to children according to infective heredity. Finally, giant viruses were shown to be chimeras composed of the genes of viral, bacterial, archaeal, and eukaryotic origins (Boyer et al., 2009). The notion of common ancestry completely undermines the existence of chimeras. Chimerism seems to give a clearer view of genome evolution than does common ancestry.

The hypothesis of a LUCA as a living organism with a ribosome has never been demonstrated. Livings have been classified into three domains commonly known as the Bacteria, the Archae, and Eukaryotes on the basis of ribosomal RNA sequences (Woese et al., 1990; Pace, 2006); viruses were excluded from this classification because they do not seem to possess a core of genes related pathogenicity and they lack ribosomes (Moreira and Lopez-Garcia, 2009). The idea of defining the livings based on the analysis of ribosomal genes implies that all genes are derived

from a ribosome-containing organism. However, metagenomic studies that test all of the sequences in an environment show that only 15% of the sequences identified in these conditions can be linked to a cell with a ribosome. These sequences have different origins, some are viral, and others are of unknown origin. These last sequences may be either from viruses that have not yet been identified or genes that were created *de novo* (ORFans). In the other hand, the core genome of nucleocytoplasmic large DNA viruses was shown to be as ancient as the other domains of life (Boyer et al., 2010; Colson et al., 2012). Thus, asserting that life began with the existence of a ribosome and is defined by this (Moreira and Lopez-Garcia, 2009) is a form of neo-creationism. Indeed, Woese and Crick (Woese, 1967, 1970; Crick et al., 1976; Andersson and Kurland, 1990) proposed that translation started long before the ribosome creation. The initial synthesis of polypeptides did not require the elaborate machinery of ribosomes, activating factors, and enzymes, but was rather accomplished using only RNA messenger and a few primitive tRNAs. This confirms that there was life before the ribosome-containing "LUCA." Therefore, current cells with ribosomes have incorporated sequences from viruses, newly created genes and sequences predating the ribosome apparition. All these data are contradicting the LUCA theory of a single ancestor of all currently living organisms.

Given what we know about microbiology, a scenario based on the theory of punctuated equilibrium is more likely than the Darwinian phyletic gradualism. According to Gould, long periods of relative evolutionary stability, called "stasis," are interrupted by evolutionary changes that occur relatively rapidly (Eldredge and Gould, 1972; Gould, 2002). Some chaotic changes, such as, geological catastrophes, can be destructive steps that create a bottleneck with few survivors. It is likely that during the evolution of life there was a catastrophic event that created a bottleneck, and the surviving cells had a ribosome and, potentially, a repertoire of ancestral genes other than those encoding the ribosome, particularly the genes encoding for RNA polymerase. The selection process resulting from the bottleneck is completely random and is not influenced by the genes that may confer a greater likelihood of survival in the ecosystem. Survival of a disaster may not confer further evolutionary advantages and can in no way be regarded as natural selection of the fittest. Rather, this process is a non-directional selection without an adaptive goal; is merely chance. Migration from the area of a bottleneck gives rise to increased diversity and the creation of new species. Heterogeneous populations result from the accumulation of mutations and LST. From time to time, a stochastic event may create a new stage and induce the proliferation of a species in an ecosystem. Thus, the capacity of specialized bacteria to multiply is linked to a limited number of events; one event that seems to be particularly important is the limitation of translation capabilities. Indeed, in at least seven bacterial phyla, the evolutionary history of specialized bacteria seems to begin with the disappearance or the malfunction of the ribosomal operon, which forced the bacterium to specialize while limiting its production only to useful proteins. This change allows the specialized bacteria to expand more rapidly than others in their specific niche (Merhej et al., 2009). However, the gene repertoires of living beings did not completely disappear but some have been used and are present in a certain number of organisms that exist today. Thus, genes have an evolutionary history that is different from that of the whole organism, as postulated by Dawkins (2006).

Our hypothesis is that ancestral organisms were sorted by successive disasters, and some of them were able to improve their ability to live in the ecosystems in which they now live. These species represent chimeras made by combining ancestral genes with laterally acquired sequences, a mixture of genes that have been recycled from organisms that are now extinct, and genes that were newly created (**Figure 5**). The idea of a unique common

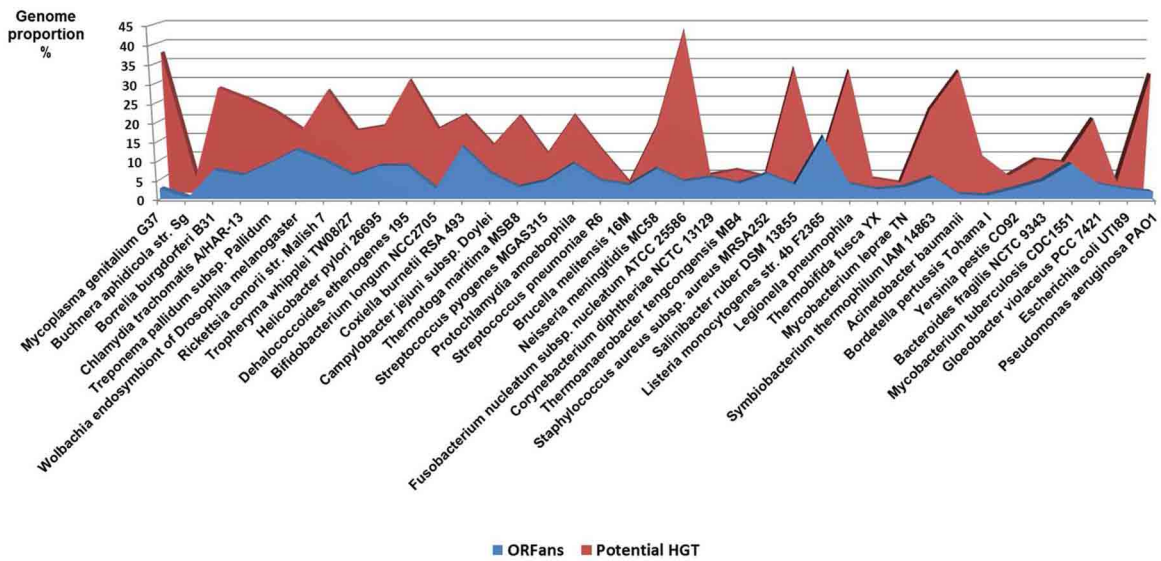


FIGURE 5 | Proportion of the ORFans and potential lateral transfer in bacterial genomes.

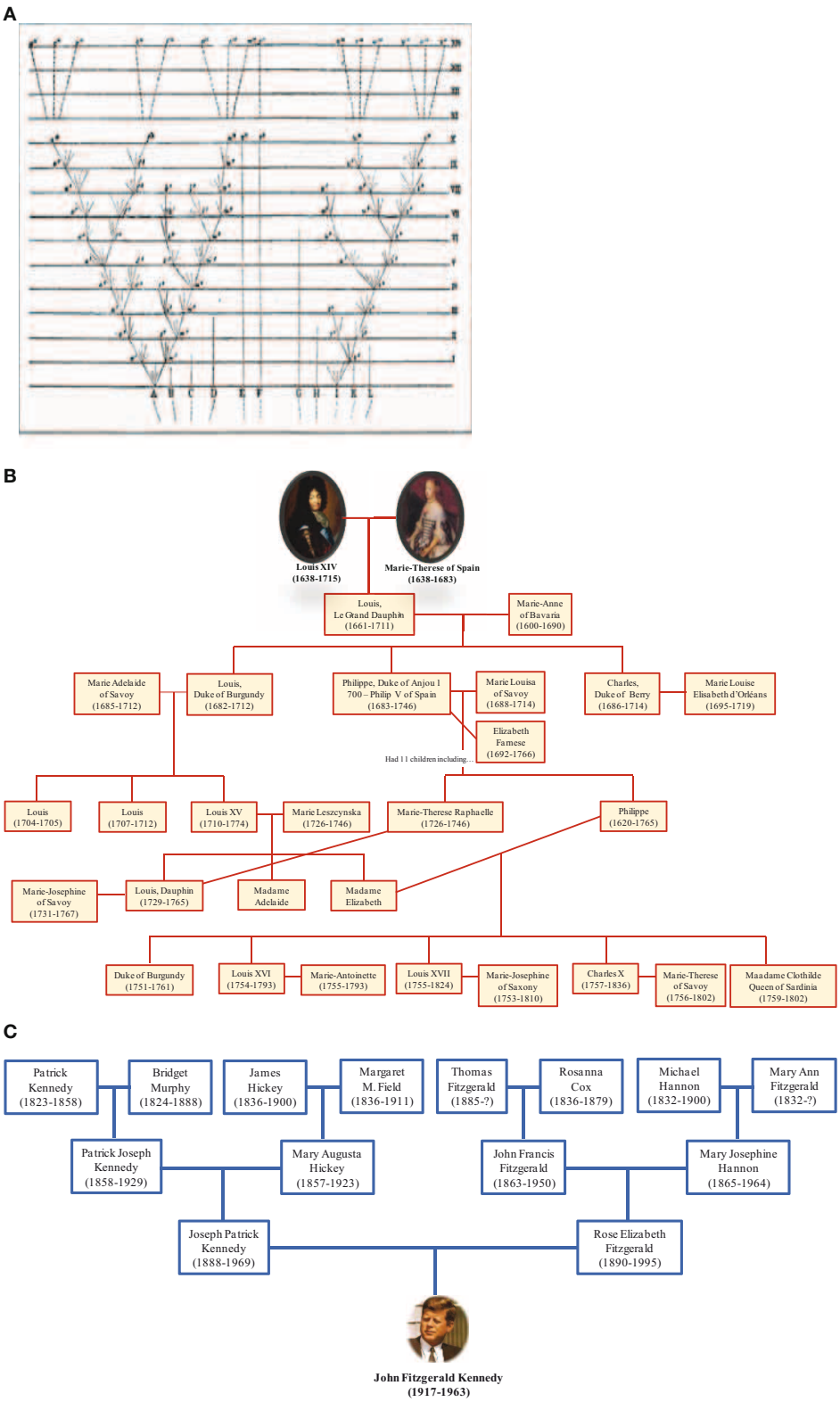


FIGURE 6 | Genealogical trees. (A) *Darwin's* illustration of the origin of species in the form of a tree with lineage splitting. The trunk growing from the root split into two branches, marking the creation of two new species. The branching continued right on up to the top of the tree representing species alive nowadays. **(B)** Darwin's tree is compatible with this tree showing descendants of Louis XIV. **(C)** In contrast, the ascendant genealogy of John F. Kennedy is showing his multiple ancestors. This representation is consistent with our current knowledge.

ancestor denies chimerism and traces the creative origin of life today to an event. Many scientists adhere to this theory and end up denying the very existence of life outside of cells with a ribosome such as, viruses that may be excluded from "life." In contrast, we believe that life cannot be considered anything other than the expression of the language contained in genetic sequences.

TREE OF LIFE

The depiction of relationships in the form of a bifurcating tree was not invented by Darwin; it had been used for many centuries to represent descendant genealogies such as those of royal families. Likewise, the term "tree of life" (TOL) was not invented by Darwin. It is a biblical metaphor that refers to the tree of knowledge that bears the fruit that gives eternal life (Bible, Genesis 2:9 and 3:22). Darwin adopted the living tree analogy to illustrate the mechanism of evolution by showing continuity between populations and species and demonstrating that certain lineages of species compete and supplanted other lineages (Penny, 2011). Thus, Darwin represented the evolution of life as a hierarchical pattern of relationships that reflects the "natural order" (Doolittle, 1999). Darwin's TOL assumes that all life forms originated from a single node corresponding to a last common ancestor through a branching evolutionary process (Doolittle and Bapteste, 2007). From this perspective, it is interesting to see the representation of trees of life in the form of family trees. Some genealogical trees begin with an ancestor and show his descendants. It should be noted that this tree does not represent reality because we know that current human beings do not descend from a single ancestor but result from many couples, forming an inverted genealogic tree (**Figure 6**). These genealogical representations ignore chimerism and LST and instead show existence of our human lineage as descending from a single ancestor. People commonly understand evolution in terms of multiple species descending from a common ancestor; the reality may more closely resemble the opposite, with multiple ancestors contributing genes to individual species.

The accuracy of the common origin and the relevance of the tree-like representation as a model of evolution have been frequently questioned (Bapteste et al., 2009; Dagan and Martin, 2009; Puigbo et al., 2010). The TOL concept presumes that all organisms are descended from a predecessor. This is true for a number of genetic sequences but not for some ORFans, including the functional ORFans. Indeed, some genes and proteins have been entirely invented in the last million years. For example, genes that are specific to the species of *Drosophila*, have been demonstrated to be essential, or at least useful, for the current life of *Drosophila* (Chen et al., 2010). These genes originated in an ancestor of *Drosophila* for which they were useful, but they were never created elsewhere. The TOL is a perception of conservative nature that lost the ability to create anymore new function since the ancestor was alive. This is contrary to our current knowledge. The analysis of bacterial genomes shows that between 10 and 15% of the genes of each species has no equivalent in other species and are likely due to "gene creativity (**Figure 5**)." Some of these genes may have been created by the reconstruction of old genes or by the genetic drift of unused genes, resulting in useful features that persevered, while other genes disappeared. This demonstrates the constant creative trial that enables the creation of new life forms.

Other evidence that deeply undermines the TOL is LST. Indeed, single gene phylogenies often yielded conflicting topologies that are distinct from the rRNA phylogenetic tree (Maddison, 1997). The causes of these discrepancies can be analytical such as limitations of the models of amino acid sequence evolution, taxon

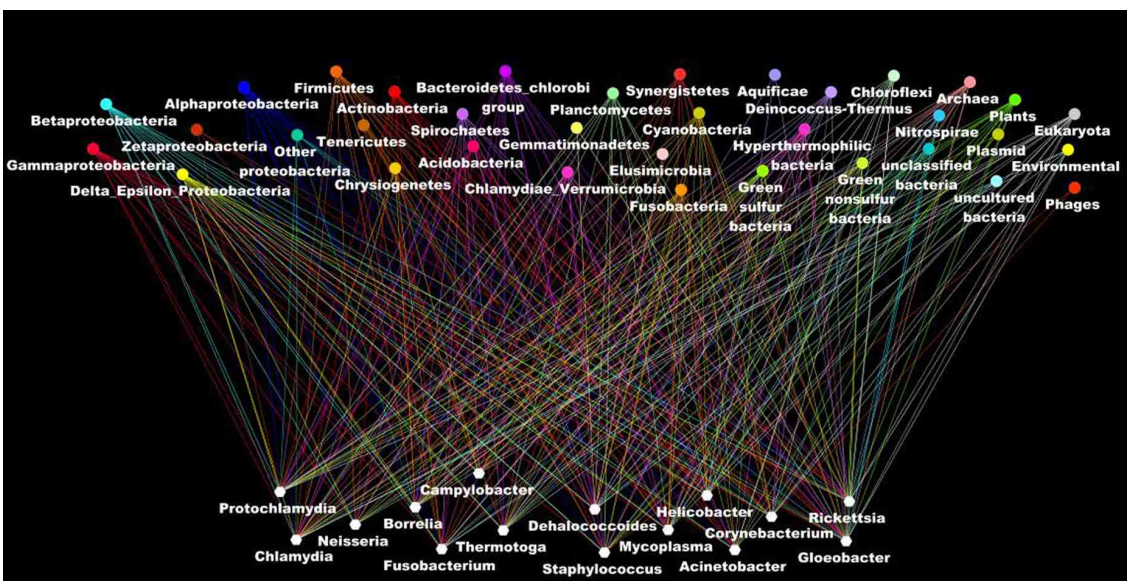


FIGURE 7 | The ascendant genealogy or the rhizome of bacteria. Bacterial genomes (at the bottom in blank) have a mosaic structure as a result of lateral inheritance from the different organisms in the different taxonomic group (at the top of the figure). Each line indicates the taxonomic origin of the putative closest phylogenetic organism as deduced from the BLastP analysis of all the genes in the genomes.

sampling, and selection bias (Rokas et al., 2003). Nevertheless, it has been stated that beyond the analytical limitations, the evolution of genes is rather reticulate due to lateral DNA transfer, and the history of life cannot be properly represented by bifurcating trees (Doolittle, 1999). Indeed, microbial genomes contain multiple selfish elements, such as bacteriophages, gene transfer agents (Paul, 2008), plasmids, and transposable elements, that are known as the mobilome. They are involved in the lateral transfer of their associated genes via different mechanisms, including conjugation, transduction, and transformation (Frost et al., 2005; Thomas and Nielsen, 2005; Asadulghani et al., 2009). Comparative genomic and phylogenetic analyses have provided evidence of extensive LST (Ochman et al., 2000; Gogarten et al., 2002; Boucher et al., 2003; Gogarten and Townsend, 2005). Thus, hyperthermophilic bacteria were found to exhibit much higher sequence similarity to the archaea that share the same habitat than to mesophilic bacteria, likely as the result of archaea-to-bacteria LST (Aravind et al., 1998; Nelson et al., 1999). Likewise, our analysis of 16 bacterial genomes found a significant proportion of genes without homologs in closely related species but with homologs in distantly related taxa (**Figure 5**). These genes were likely acquired through lateral transfer. Evidence of LST according to a sympatric model of evolution is present in obligate intracellular bacteria that share the same host (Moliner et al., 2010; Coscolla et al., 2011; Georgiades et al., 2011; Merhej et al., 2011). The high prevalence of LST raised the notion of a connected microbial "gene pool" with no barrier (Beiko et al., 2005; Koonin, 2011d) while questioning the concept of bacterial species (Bapteste et al., 2009). Moreover, the dynamic nature of evolution, in which the genetic information of living organisms is inherited not only vertically but also laterally, challenges the representation of the evolution of life in the form of a Darwinian bifurcating tree (Bapteste et al., 2004, 2005; O'Malley and Boucher, 2005; Jeffroy et al., 2006; Susko et al., 2006; Marttinen et al., 2012) (**Figure 7**).

The fluidity of microbial genomes has instigated many efforts to find a better representation of the dynamic relationships that shape microbial evolution. It has been proposed that congruent topologies of trees for several highly conserved genes might better represent the history of the majority of the genes (Wolf et al., 2002; Dagan and Martin, 2006). Using a comprehensive comparison of individual gene tree topologies, the "forest of life" (FOL), a collection of phylogenetic trees for all genes, was proposed as an alternative to a single tree (Puigbo et al., 2009; Koonin et al., 2011). In this approach, the topologies of the

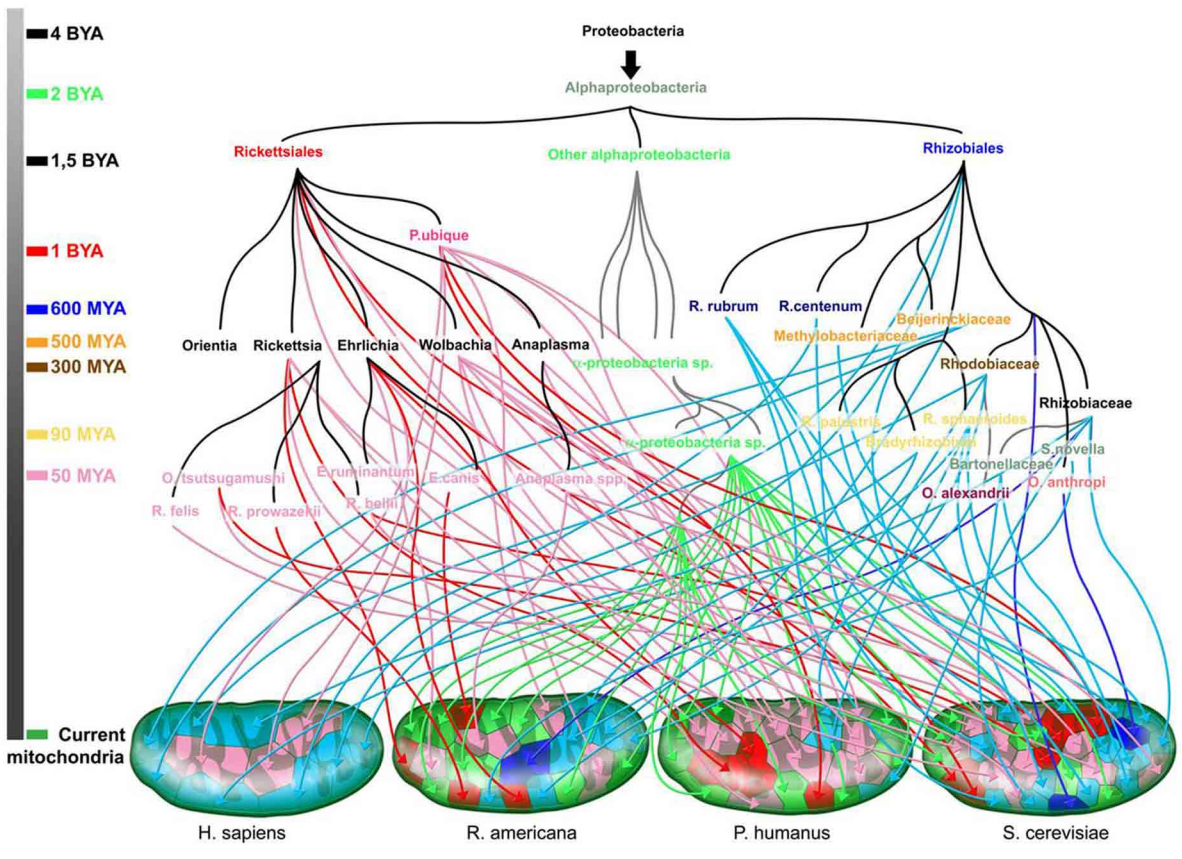


FIGURE 8 | The rhizome of mitochondria (adapted from Georgiades and Raoult (2011). The origin of each gene is represented along with the time scale of the species divergence. Dark blue and red arrows are for sister taxa with high bootstrap values, and light blue and pink arrows for sister taxa with low bootstrap values. Green arrows are for sister taxa from the Alphaproteobacteria subgroup. Colors on the time scale coincide with the emergence of the corresponding colored species.

102 nearly universal trees (NUTs) were highly consistent and seemed to represent a central evolutionary trend in the FOL. The consensus topology of the NUTs has been proposed as an accurate representation of the evolution of organisms. For other scientists, the dynamic picture of the prokaryotic world is best represented as a complex network of genetic elements that exchange genes. Considering the high level of horizontal inheritance, microbial evolution more closely resembles a rhizome than a bifurcating tree (Raoult, 2010; Merhej et al., 2011; Ramulu et al., 2012) and the tree-like representation should be completely abandoned in favor of a web-like representation of evolution (Sneath, 1975; Gogarten et al., 2002; Doolittle and Bapteste, 2007; Puigbo et al., 2010; Popa et al., 2011). Unlike the hierarchical tree-like model, the novel representations that consider the broad-spectrum of gene origins, including vertical descent, lateral inheritance, and *de novo* creation, are promising representations of microbial genome evolution.

CONCLUSIONS

None of the seven points laid out in the introduction to this manuscript can be permanently retained, as established by Darwin's theory, which was at the time a fight against the creationists. This theory cannot be upheld in its entirety. Recent advances from genomics refute the ideas of gradualism, exclusive vertical inheritance, evolution selecting the fittest, a common ancestor and the TOL. Indeed, there may not be any two genes that have the same evolutionary tree. Moreover, it is less the genes that are traded than the sequences themselves. Genes may have portions of sequences with different evolutionary origins because of recombination. An accurate representation of the genealogy of genes in a repertoire should take into account the different origins of closely and distantly related organisms as well as organisms that have gone extinct. A single tree is largely inadequate. We prefer to represent evolution as a family tree or in the form of a rhizome, which corresponds to a more authentic description of our present knowledge than the TOL (**Figure 8**).

Finally Darwin has contributed to the debate on the myth from the Bible and Aristotle and tried to return the history of life to the domain of science. At the same time, he created a cultural and religious context, a sort of scientific battle against obscuring belief. Indeed, he is considered in Britain and the United States of America as an icon of science against the obscurantist religious or the creationists (Raoult, 2008, 2010). The expression of Darwinian's idolatry peaked in the year 2009 which corresponded to the bicentenary of his birth and the 150th year of his theory, when virtually all scientific journals posted photos and texts on Darwin. Currently, even in the USA, the opinion is divided on evolution at about equal between evolutionists and creationists. This position has become ideological so that many of the major writers of the twentieth century in the field of evolution felt compelled to take a stand on the issue. Mayr, Gould, and Dawkins stated theories that are antagonist to that of Darwin. Moreover, Karl Popper claimed that the theory of evolution was not a scientific theory (Popper, 2002). From our point of view, the theory of evolution is a scientific theory however it is an outdated theory. Darwin's theory should not become a religion but remain a scientific theory from another era that can be refined based on the actual insights from microbial genomics.

REFERENCES

Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F. A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S. G., Maiers, M., Guethlein, L. A., Tavoularis, S., Little, A. M., Green, R. E., Norman, P. J., and Parham, P. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334, 89–94.

Andersson, S. G., and Kurland, C. G. (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198–210.

Andersson, S. G., and Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends Microbiol.* 6, 263–268.

Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R., and Koonin, E. V. (1998). Evidence for massive gene exchange between archaeal, and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444.

Arbuckle, J. H., Medveczky, M. M., Luka, J., Hadley, S. H., Luegmayr, A., Ablashi, D., Lund, T. C., Tolar, J., De Meirleir, K., Montoya, J. G., Komaroff, A. L., Ambros, P. F., and Medveczky, P. G. (2010). The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes *in vivo*, and *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5563–5568.

Aristotle, (2008). *Physics.* New York, NY: Oxford University Press.

Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., Nakayama, K., and Hayashi, T. (2009). The defective prophage pool of *Escherichia coli* O157, prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 5:e1000408. doi: 10.1371/journal.ppat.1000408

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W. F. (2004). Phylogenetic reconstruction, and lateral gene transfer. *Trends Microbiol.* 12, 406–411.

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupre, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution, and the tree of life are two different things. *Biol. Direct* 4, 34.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R. L., and Doolittle, W. F. (2005). Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* 5, 33.

Barlow, M. (2009). What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol.* 532, 397–411.

Barlow, N. (1958). *The Autobiography of Charles Darwin.* London: St James's place.

Barnes, J. (1983). *The Presocratic Philosophers.* New york, NY: Routledge.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14332–14337.

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E., Nesbo, C. L., Case, R. J., and Doolittle, W. F. (2003). Lateral gene transfer, and the origins of prokaryotic groups. *Annu. Rev. Genet.* 37, 283–328.

Boyer, M., Madoui, M. A., Gimenez, G., La, S. B., and Raoult, D. (2010). Phylogenetic, and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLoS ONE* 5:e15530. doi: 10.1371/journal.pone.0015530

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M. G., Suzan-Monti, M., La Scola, B., Koonin, E. V., and Raoult, D. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21848–21853.

Brisson-Noel, A., Arthur, M., and Courvalin, P. (1988). Evidence for natural gene transfer from gram-positive cocci to *Escherichia coli*. *J. Bacteriol.* 170, 1739–1745.

Brochier, C., Philippe, H., and Moreira, D. (2000). The evolutionary history of ribosomal protein RpS14, horizontal gene transfer at the heart of the ribosome. *Trends Genet.* 16, 529–533.

Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J.,

Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M. Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., and Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315, 207–212.

Chan, C. X., Beiko, R. G., Darling, A. E., and Ragan, M. A. (2009). Lateral transfer of genes, and gene fragments in prokaryotes. *Genome Biol. Evol.* 1, 429–438.

Charlesworth, B., and Charlesworth, D. (2009). Darwin, and genetics. *Genetics* 183, 757–766.

Chen, S., Zhang, Y. E., and Long, M. (2010). New genes in Drosophila quickly become essential. *Science* 330, 1682–1685.

Colson, P., de Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* 55, 321–332.

Coscolla, M., Comas, I., and Gonzales-Candelas, F. (2011). Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol. Biol. Evol.* 28, 985–1001.

Crick, F. H., Brenner, S., Klug, A., and Pieczenik, G. (1976). A speculation on the origin of protein synthesis. *Orig. Life* 7, 389–397.

Dagan, T., and Martin, W. (2006). The tree of one percent. *Genome Biol.* 7, 118.

Dagan, T., and Martin, W. (2009). Getting a better picture of microbial evolution en route to a network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2187–2196.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray.

Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 14, 1036–1042.

Dawkins, R. (2006). *The Selfish Gene*. New York, NY: Oxford university press.

de Buffon, G. L. L. (1753). *Histoire Naturelle Générale et Particulière*. Paris: Parent-Desbarres.

Doolittle, W. F. (1999). Phylogenetic classification, and the universal tree. *Science* 284, 2124–2129.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism, and the tree of life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Munoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., and Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756.

Edwards, R. A., Olsen, G. J., and Maloy, S. R. (2002). Comparative genomics of closely related salmonellae. *Trends Microbiol.* 10, 94–99.

El Albani, A., Bengtson, S., Canfield, D. E., Bekker, A., Macchiarelli, R., Mazurier, A., Hammarlund, E. U., Boulvais, P., Dupuy, J. J., Fontaine, C., Fursich, F. T., Gauthier-Lafaye, F., Janvier, P., Javaux, E., Ossa, F. O., Pierson-Wickmann, A. C., Riboulleau, A., Sardini, P., Vachard, D., Whitehouse, M., and Meunier, A. (2010). Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago. *Nature* 466, 100–104.

Eldredge, N., and Gould, S. J. (1972). *Punctuated Equilibria: An Alternative to Phyletic Gradualism*. San Francisco, CA: Freeman, Cooper, and Company.

Feil, E. J., Enright, M. C., and Spratt, B. G. (2000). Estimating the relative contributions of mutation, and recombination to clonal diversification: a comparison between *Neisseria meningitidis*, and *Streptococcus pneumoniae*. *Res. Microbiol.* 151, 465–469.

Fischer, D., and Eisenberg, D. (1999). Finding families for genomic ORFans. *Bioinformatics* 15, 759–762.

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Syst. Zool.* 19, 99–113.

Forterre, P. (1999). Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.* 33, 457–465.

Forterre, P. (2002). The origin of DNA genomes, and DNA replication proteins. *Curr. Opin. Microbiol.* 5, 525–532.

Forterre, P. (2006). Three RNA cells for ribosomal lineages, and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3669–3674.

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732.

Georgiades, K., Merhej, V., El Karkouri, K., Raoult, D., and Pontarotti, P. (2011). Gene gain, and loss events in Rickettsia, and Orientia species. *Biol. Direct* 6, 6.

Georgiades, K., and Raoult, D. (2011). The rhizome of *Reclinomonas americana*, *Homo sapiens*, *Pediculus humanus*, and *Saccharomyces cerevisiae mitochondria*. *Biol. Direct* 6, 55.

Gevers, D., Vandepoele, K., Simillon, C., and Van de Peer, Y. (2004). Gene duplication, and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol.* 12, 148–154.

Glansdorff, N., Xu, Y., and Labedan, B. (2008). The last universal common ancestor: emergence, constitution, and genetic legacy of an elusive forerunner. *Biol. Direct* 3, 29.

Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.

Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer, genome innovation, and evolution. *Nat. Rev. Microbiol.* 3, 679–687.

Gould, S. (1984). *Darwin et les Grandes Enigmes de la vie*. Paris: Seuil.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA: Belknap Press of Harvard University press.

Gregory, T. R. (2005). Synergy between sequence, and size in large-scale genomics. *Nat. Rev. Genet.* 6, 699–708.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database, and tools to display CRISPRs, and to generate dictionaries of spacers, and repeats. *BMC Bioinformatics* 8, 172.

Hanage, W. P., Fraser, C., and Spratt, B. G. (2005). Fuzzy species among recombinogenic bacteria. *BMC Biol.* 3, 6.

Hao, W., and Golding, G. B. (2008). Uncovering rate variation of lateral gene transfer during bacterial genome evolution. *BMC Genomics* 9, 235.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria, and archaea. *Science* 327, 167–170.

Hughes, S. A., Wedemeyer, H., and Harrison, P. M. (2011). Hepatitis delta virus. *Lancet* 378, 73–85.

Huxley, J. S. (1942). *Evolution: The Modern Synthesis*. London: Allen, and Unwin.

Huynen, M. A., and Bork, P. (1998). Measuring genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5849–5856.

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying, and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108.

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806.

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic isl, ands: tools of bacterial horizontal gene transfer, and evolution. *FEMS Microbiol. Rev.* 33, 376–393.

Karev, G. P., Wolf, Y. I., Rzhetsky, A. Y., Berezovskaya, F. S., and Koonin, E. V. (2002). Birth, and death of protein domains: a simple model of evolution explains power law behavior. *BMC Evol. Biol.* 2, 18.

Koonin, E. V. (2003). Comparative genomics, minimal gene-sets, and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136.

Koonin, E. V. (2009a). Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37, 1011–1034.

Koonin, E. V. (2009b). Evolution of genome architecture. *Int. J. Biochem. Cell Biol.* 41, 298–306.

Koonin, E. V. (2009c). On the origin of cells, and viruses: primordial virus world scenario. *Ann. N.Y. Acad. Sci.* 1178, 47–64.

Koonin, E. V. (2009d). Towards a postmodern synthesis of evolutionary biology. *Cell Cycle* 8, 799–800.

Koonin, E. V. (2011a). "Comparative genomics: evolving genomescapes," in *The Logic of Chance. The Nature, and Origin of Biological Evolution*, ed K. Jensen (New Jersey, NJ: FT Press Science), 49–79.

Koonin, E. V. (2011b). "The last universal common ancestor, the origin of cells, and the primordial gene pool,"

in *The Logic of Chance. The Nature, and Origin of Biological Evolution,* ed K. Jensen (New Jersey, NJ: FT Press Science), 329–350.

Koonin, E. V. (2011c). "The Non-adaptive Null Hypothesis of Genome Evolution, and Orgins of Biological Complexity," in *The Logic of Chance. The Nature, and Origin of Biological Evolution,* ed K. Jensen (New Jersey, NJ: FT Press Science), 225–255.

Koonin, E. V. (2011d). "The web genomics of the prokaryotic worl: vertical, and horizontal flows of genes, the mobilome, and the dynamic pangenomes," in *The Logic of Chance. The Nature, and Origin of Biological Evolution,* ed K. Jensen (New Jersey, NJ: FTP Press Science), 105–144.

Koonin, E. V., Puigbo, P., and Wolf, Y. I. (2011). Comparison of phylogenetic trees, and search for a central trend in the "forest of life". *J. Comput. Biol.* 18, 917–924.

Koonin, E. V., and Wolf, Y. I. (2009). Is evolution Darwinian or/, and Lamarckian? *Biol. Direct* 4, 42.

Lamarck, J. B. (1809). *Philosophie Zoologique, ou Exposition des Considérations Relatives à L'histoire Naturelle des Animaux.* Paris: Dentu.

Lawrence, J. G., and Retchless, A. C. (2009). The interplay of homologous recombination, and horizontal gene transfer in bacterial speciation. *Methods Mol. Biol.* 532, 29–53.

Lederberg, J. (1949). Bacterial variation. *Annu. Rev. Microbiol.* 3, 1–22.

Lefebure, T., and Stanhope, M. J. (2007). Evolution of the core, and pan-genome of Streptococcus: positive selection, recombination, and genome composition. *Genome Biol.* 8, R71.

Leibniz, G. W. (1996). *New Essays on Human Understanding*. Cambridge: Cambridge University Press.

Leipe, D. D., Aravind, L., and Koonin, E. V. (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res.* 27, 3389–3401.

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M., and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438, 86–89.

Lucretius, (1995). *On the nature of things: De rerum natura*. Baltimore, MD: The john Hopkins University Press. Translated by Esolen, A. M.

Lyell, C. (1830). *Principles of Geology.* London: John Murray.

Lynch, M. (2006). Streamlining, and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8597–8604.

Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. *Science* 302, 1401–1404.

Maddison, W. (1997). Gene trees in species trees. *Syst. Biol.* 46, 523–536.

Malthus, T. R. (1798). *An Essay on the Principle of Population.* London: St. Paul's Church-yard.

Marttinen, P., Hanage, W. P., Croucher, N. J., Connor, T. R., Harris, S. R., Bentley, S. D., and Corander, J. (2012). Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res.* 40, e6.

Mayr, E. (1944). *Systematics, and the Origin of Species.* New York, NY: Columbia University Press.

Mayr, E. (1963). *Animal Species, and Evolution.* Cambridge, MA: Harvard university Press.

Mayr, E. (1982). *The Growth of Biological Thought: Diversity, Evolution, and Inheritance.* Cambridge, MA: Harvard university Press.

McNulty, S. N., Foster, J. M., Mitreva, M., Dunning Hotopp, J. C., Martin, J., Fischer, K., Wu, B., Davis, P. J., Kumar, S., Brattig, N. W., Slatko, B. E., Weil, G. J., and Fischer, P. U. (2010). Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS ONE* 5:e11029. doi: 10.1371/journal.pone.0011029

Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric *Rickettsia felis* paradigm demonstrates the r, andom transfer of DNA sequences. *Mol. Biol. Evol.* 28, 3213–3223.

Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4, 13.

Mira, A., Klasson, L., and andersson, S. G. (2002). Microbial genome evolution: sources of variability. *Curr. Opin. Microbiol.* 5, 506–512.

Mirkin, B. G., Fenner, T. I., Galperin, M. Y., and Koonin, E. V. (2003). Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor, and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* 3, 2.

Moliner, C., Fournier, P. E., and Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol. Rev.* 34, 281–294.

Monod, J. (1972). *Chance, and Necessity: An Essay on the Natural Philosophy of Modern Biology*. New york, NY: Vintage Books.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586.

Moreira, D., and Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7, 306–311.

Mulkidjanian, A. Y., Koonin, E. V., Makarova, K. S., Mekhedov, S. L., Sorokin, A., Wolf, Y. I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., Haselkorn, R., and Galperin, M. Y. (2006). The cyanobacterial genome core, and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13126–13131.

Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* 36, 760–766.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea, and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.

Nietzsche, F. (2006). *Thus Spoke Zarathustra.* Cambridge: Cambridge University Press.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer, and the nature of bacterial innovation. *Nature* 405, 299–304.

Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D., and Claverie, J. M. (2000). Selfish DNA in protein-coding genes of Rickettsia. *Science* 290, 347–350.

Ohno, S. (1970). *Evolution by Gene Duplication.* Vienna: Springer.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature* 246, 96–98.

O'Malley, M. A., and Boucher, Y. (2005). Paradigm change in evolutionary microbiology. *Stud. Hist. Philos. Biol. Biomed. Sci.* 36, 183–208.

Orgel, L. E., Crick, F. H., and Sapienza, C. (1980). Selfish, D. N. A. *Nature* 288, 645–646.

Pace, N. R. (2006). Time for a change. *Nature* 441, 289.

Pasteur, L. (1880). De l'attenuation du virus du cholera des poules. *C. R. Acad. Sci. Paris* 91, 673–680.

Paul, J. H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* 2, 579–589.

Penny, D. (2011). Darwin's theory of descent with modification, versus the biblical tree of life. *PLoS Biol.* 9, e1001096. doi: 10.1371/journal.pbio.1001096

Pereto, J., Lopez-Garcia, P., and Moreira, D. (2004). Ancestral lipid biosynthesis, and early membrane evolution. *Trends Biochem. Sci.* 29, 469–477.

Perna, N. T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157, H7. *Nature* 409, 529–533.

Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers, and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609.

Popper, K. (2002). *The Logic of Scientific Discovery*. London: Routledge.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a 'tree of life' in the thicket of the phylogenetic forest. *J. Biol.* 8, 59.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree, and net components of prokaryote evolution. *Genome. Biol. Evol.* 2, 745–756.

Ramulu, H. G., Raoult, D., and Pontarotti, P. (2012). The rhizome of life: what about metazoa?. *Front. Cell. Infect. Microbiol.* 50, 1–11. doi: 10.3389/fcimb.2012.00050

Raoult, D. (2008). Creationism–remember the principle of falsifiability. *Lancet* 372, 2095–2096.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Raoult, D. (2011). A viral gr, andfather: genomics in 2010 contradicts

Darwin's vision of evolution. *Eur. J. Clin. Microbiol. Infect. Dis.* 30, 935–936.

Remington, F. (1889). Horses of the plains. *Century* 37, 332–343.

Rokas, A., King, N., Finnerty, J., and Carroll, S. B. (2003). Conflicting phylogenetic signals at the base of the metazoan tree. *Evol. Dev.* 5, 346–359.

Rolain, J. M., Fancello, L., Desnues, C., and Raoult, D. (2011). Bacteriophages as vehicles of the resistome in cystic fibrosis. *J. Antimicrob. Chemother.* 66, 2444–2447.

Schoen, C., Blom, J., Claus, H., Schramm-Gluck, A., Brandt, P., Muller, T., Goesmann, A., Joseph, B., Konietzny, S., Kurzai, O., Schmitt, C., Friedrich, T., Linke, B., Vogel, U., and Frosch, M. (2008). Whole-genome comparison of disease, and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3473–3478.

Shoemaker, N. B., Vlamakis, H., Hayes, K., and Salyers, A. A. (2001). Evidence for extensive resistance gene transfer among Bacteroides spp., and among Bacteroides, and other genera in the human colon. *Appl. Environ. Microbiol.* 67, 561–568.

Singer, C. (1931). *A Short History of Biology.* Oxford: Clarenden press.

Sneath, P. H. A. (1975). Cladistic representation of reticulate evolution. *Syst. Zool.* 24, 360–368.

Susko, E., Leigh, J., Doolittle, W. F., and Bapteste, E. (2006). Visualizing, and assessing phylogenetic congruence of core gene sets: a case study of the gamma-proteobacteria. *Mol. Biol. Evol.* 23, 1019–1030.

Tamas, I., Klasson, L., Canback, B., Naslund, A. K., Eriksson, A. S., Wernegreen, J. J., Sandstrom, J. P., Moran, N. A., and Andersson, S. G. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296, 2376–2379.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., Madupu, R., Brinkac, L. M., Dodson, R. J., Rosovitz, M. J., Sullivan, S. A., Daugherty, S. C., Haft, D. H., Selengut, J., Gwinn, M. L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K. J., Smith, S., Utterback, T. R., White, O., Rubens, C. E., Grandi, G., Madoff, L. C., Kasper, D. L., Telford, J. L., Wessels, M. R., Rappuoli, R., and Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae:* implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955.

Thomas, C. A. Jr. (1971). The genetic organization of chromosomes. *Annu. Rev. Genet.* 5, 237–256.

Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721.

Treangen, T. J., and Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284

Weinberger, A. D., Sun, C. L., Plucinski, M. M., Denef, V. J., Thomas, B. C., Horvath, P., Barrangou, R., Gilmore, M. S., Getz, W.M., and Banfield, J. F. (2012). Persisting viral sequences shape microbial CRISPR-based Immunity. *PLoS Comput. Biol.* 8:e1002475. doi: 10.1371/journal.pcbi.1002475

Wellner, A., Lurie, M. N., and Gophna, U. (2007). Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.* 8, R156.

Woese, C. (1970). Molecular mechanics of translation: a reciprocating ratchet mechanism. *Nature* 226, 817–820.

Woese, C. R. (1967). *The Genetic Code.* New York, NY: Evanston, and London.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. (2002). Genome trees, and the tree of life. *Trends Genet.* 18, 472–479.

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A. K., Taboada, E., and Gannon, V. P. (2007). Genome evolution in major *Escherichia coli* O157, H7 lineages. *BMC Genomics* 8, 121.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 May 2012; paper pending published: 07 July 2012; accepted: 06 August 2012; published online: 28 August 2012.

Citation: Merhej V and Raoult D (2012) Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin. Front. Cell. Inf. Microbio. 2:113. doi: 10.3389/fcimb.2012.00113

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 29 June 2012
doi: 10.3389/fcimb.2012.00089

Gene flow and biological conflict systems in the origin and evolution of eukaryotes

L. Aravind[1]*, Vivek Anantharaman[1], Dapeng Zhang[1], Robson F. de Souza[1,2] and Lakshminarayan M. Iyer[1]

[1] *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
[2] *Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil*

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Didier Raoult, Université de la Méditerranée, France
Chengzhi Wang, Cancer Research Center, USA

****Correspondence:***
L. Aravind, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA.
e-mail: aravind@ncbi.nlm.nih.gov

The endosymbiotic origin of eukaryotes brought together two disparate genomes in the cell. Additionally, eukaryotic natural history has included other endosymbiotic events, phagotrophic consumption of organisms, and intimate interactions with viruses and endoparasites. These phenomena facilitated large-scale lateral gene transfer and biological conflicts. We synthesize information from nearly two decades of genomics to illustrate how the interplay between lateral gene transfer and biological conflicts has impacted the emergence of new adaptations in eukaryotes. Using apicomplexans as example, we illustrate how lateral transfer from animals has contributed to unique parasite-host interfaces comprised of adhesion- and O-linked glycosylation-related domains. Adaptations, emerging due to intense selection for diversity in the molecular participants in organismal and genomic conflicts, being dispersed by lateral transfer, were subsequently exapted for eukaryote-specific innovations. We illustrate this using examples relating to eukaryotic chromatin, RNAi and RNA-processing systems, signaling pathways, apoptosis and immunity. We highlight the major contributions from catalytic domains of bacterial toxin systems to the origin of signaling enzymes (e.g., ADP-ribosylation and small molecule messenger synthesis), mutagenic enzymes for immune receptor diversification and RNA-processing. Similarly, we discuss contributions of bacterial antibiotic/siderophore synthesis systems and intra-genomic and intra-cellular selfish elements (e.g., restriction-modification, mobile elements and lysogenic phages) in the emergence of chromatin remodeling/modifying enzymes and RNA-based regulation. We develop the concept that biological conflict systems served as evolutionary "nurseries" for innovations in the protein world, which were delivered to eukaryotes via lateral gene flow to spur key evolutionary innovations all the way from nucleogenesis to lineage-specific adaptations.

Keywords: antibiotics, biological conflict, endosymbiosis, immunity proteins, restriction-modfication, RNAi, selfish elements, toxins

INTRODUCTION

Ever since the emergence of the endosymbiotic hypothesis as the primary model for the origin of eukaryotes there has been considerable interest in two major issues which it brought forth, namely large-scale lateral gene flow and genetic conflicts. While the exact details of the nature of this endosymbiotic event are still debated, by its very nature the endosymbiotic hypothesis implies gene flow between the alphaproteobacterial mitochondrial progenitor and the nucleo-cytoplasmic progenitor of archaeal ancestry (Martin and Muller, 1998; Esser et al., 2004; Rivera and Lake, 2004; Aravind et al., 2006; Gabaldon and Huynen, 2007; Pisani et al., 2007; Sapp, 2007). This phenomenon is not just relevant to the origin of eukaryotes, but also several other symbiogenic events that shaped the subsequent evolution of eukaryotes, such as the origin of the primary photosynthetic eukaryotes, including the plants, and the numerous secondary or tertiary photosynthetic eukaryotes (Delwiche, 1999; Palmer, 2003; Bhattacharya et al., 2004; Keeling, 2004; Huang and Gogarten, 2007; Oborník et al., 2009). In the former event, not just the well-known gene flow from cyanobacteria, but also complementary contributions from a chlamydia-like endosymbiont have been postulated (Huang and Gogarten, 2007). Additionally, there are other inter-organismal interactions that have occurred throughout eukaryotic evolution, which have resulted in comparable gene flow, albeit in a more episodic fashion (Anantharaman et al., 2007). Eukaryotes are characterized by a wide-range of close organismal associations. Indeed, cytoplasmic symbiotic bacteria, comparable to the progenitors of the mitochondria and chloroplasts, and infection by several types of large DNA viruses are a common feature of many eukaryotes, including representatives of the metazoan and amoebozoan lineages (Batut et al., 2004; Collingro et al., 2005; Ogata et al., 2006; Iyer et al., 2006b; Nikoh et al., 2008; Bertelli et al., 2010; Raoult and Boyer, 2010; Schmitz-Esser et al., 2010; Georgiades et al., 2011). There are also examples of some

rather dramatic inter-eukaryotic associations, like endoparasitism as exhibited by apicomplexans, karyoklepty, or "theft" of chlorophyte nuclei (along with the chloroplasts) observed among ciliates, or karyoparasitism, involving injection of parasitic nuclei into host cells, which is observed in certain rhodophytes (Fields and Rhodes, 1991; Goff and Coleman, 1995; Johnson et al., 2007). Further, it has been noted that the phagotrophic nutrition of many eukaryotes can also result in a more general form of genetic chimerism, facilitated by the constant engulfment of genetic material of particular types of bacteria and eukaryotes (Doolittle, 1998). Yet other eukaryotes, such as the rotifers, appear to even actively engage in uptake and incorporation of genetic material from their environments—in addition to the proposed role in compensating for the lack of sexual reproduction, this phenomenon also serves as a conduit for notable "alien" gene flow (Gladyshev et al., 2008). Thus, it has become increasingly clear in the past two decades that gene flow between distant lineages and the consequent genomic chimerism might have a notable role in the evolution of eukaryotes.

Inter-organismal and intra-organismal genetic conflicts are a quotidian feature across all organizational levels of life (Smith and Price, 1973; Maynard Smith and Szathmáry, 1995; Hurst et al., 1996; Burt and Trivers, 2006; Werren, 2011). In their simplest form they include various trophic interactions between organisms, such as predation. Such conflicts might also arise between different cells of the same species cooperatively aggregating to form a multicellular assembly or developing as a multicellular organism due the emergence of "cheaters," whose genetic interests do not align with the remaining cooperating cells (Dao et al., 2000). At the level of a single cell, as the interests of different genomes residing within it are not necessarily aligned with each other, there is potential for yet another level of genetic conflicts (Burt and Trivers, 2006). Such conflicts have a long evolutionary history in the prokaryotic superkingdoms in the form of the interactions between plasmids and the cellular genome. However, the endosymbiotic origin of eukaryotes made it one of their quintessential features because it brought together multiple distinct genomes (i.e., the nuclear and mitochondrial) in a single cell (Maynard Smith and Szathmáry, 1995; Werren, 2011). Such inter-genomic conflicts within the cell further expanded in course of eukaryotic evolution due to additional associations introducing interactions with genomes from plastids, nucleomorphs, and endosymbiotic/parasitic and intracellular bacterial predators of mitochondria (Sassera et al., 2006; Werren, 2011). In several cases symbiotic bacteria are involved in multi-level cooperation-conflict relationships: For instance, the bacterial symbiont *Photorhabdus* enables predatory nematodes to feed on insects by killing them with toxins (Bowen et al., 1998), whereas the endosymbiotic bacterium *Hamiltonella defensa* protects aphids against parasitoid wasps by deploying toxins against them (Degnan et al., 2009). Conflicts between the cellular genomes and viruses that exploit them for their own reproduction add yet another dimension to conflicts occurring within cells (Iyer et al., 2006b; Raoult and Boyer, 2010). Finally, there might be genetic conflicts within a single genome itself, arising from a wide variety of selfish elements trying to maximize their own fitness at the expense of the remaining genes (Burt and Trivers, 2006; Werren, 2011). These selfish elements are often characterized by a degree of intra- and/or inter-genomic mobility and assume a bewildering array of forms, including numerous distinct types of transposable elements, restriction-modification, and toxin-antitoxin systems (Kobayashi, 2001; Anantharaman and Aravind, 2003; Burt and Trivers, 2006; Ishikawa et al., 2010; Leplae et al., 2011). The former elements catalyze or facilitate their own proliferation, while the latter elements enforce cellular genomes to retain them by killing cells in which they are disrupted. Despite being primarily selfish elements, they might on occasions confer a fitness advantage to genomes, as this indirectly augments their own fitness (Burt and Trivers, 2006; Werren, 2011).

These conflicts are often directly mediated by particular molecules, either proteins or small molecules which act as "chemical armaments"; although in multicellular forms it might be reflected as morphological features that serve as weaponry (Smith and Price, 1973; Anantharaman and Aravind, 2003; Degnan et al., 2009; Ishikawa et al., 2010; Leplae et al., 2011; Werren, 2011; Zhang et al., 2011; Iyer et al., 2011b). Not surprisingly, each of the many levels of organismal conflict have sparked off intense "arms races" between the interacting organisms (Dawkins and Krebs, 1979), whose signatures are often seen in the form of extensive diversification of the proteins directly participating in, or synthesizing molecules deployed in conflict (Cascales et al., 2007; Zhang et al., 2011). Concomitantly, there is also a similar rapid diversification of proteins directly involved in defending or serving as antidotes against the chemical armaments deployed in the conflict (Anantharaman and Aravind, 2003; Leplae et al., 2011; Zhang et al., 2011; Iyer et al., 2011b). Importantly, both the offensive and defensive molecular adaptations involved in these conflicts can be transmitted between genomes by way of lateral transfer and is an important factor both in the spread of antibiotic production and resistance among prokaryotes (Walsh, 2003; Aminov and Mackie, 2007; Skippington and Ragan, 2011).

The ever-expanding genomic data from both eukaryotes and prokaryotes, along with genome-scale analysis, has considerably elucidated the major trends in the genomic chimerism arising from the bacterial and archaeal progenitors of the eukaryotes (Martin and Muller, 1998; Esser et al., 2004; Rivera and Lake, 2004; Aravind et al., 2006; Gabaldon and Huynen, 2007; Pisani et al., 2007). These analyses have particularly helped differentiate the cellular systems which have a primarily archaeal provenance (e.g., core DNA replication, core RNA metabolism, and translation) as against those with a primarily bacterial provenance (various aspects of energy, anabolic, and catabolic metabolism). However, uncovering the origins of specific systems, which appear to be eukaryotic synapomorphies (or shared derived characters), have required a somewhat distinct computational approach relying on in-depth analysis of protein sequences and structures (Aravind et al., 2006, 2011; Burroughs et al., 2011). Such analyses revealed glimpses of a collusion between gene flow through lateral transfer and the selective forces acting on molecular players in organismal and intra-genomic conflict in shaping the evolution of key components of systems such as eukaryotic chromatin, RNA-based gene regulation, and certain signaling pathways. However, this aspect of eukaryotic evolution is

considerably under-appreciated. Hence, in this article we present a synthetic overview of: (1) how large-scale lateral gene flow between interacting organisms has facilitated the emergence of new adaptations deployed in inter-organismal conflict. (2) How adaptations developed due to the intense selection for diversity in the molecular participants in organismal and genomic conflicts were dispersed by lateral transfer and subsequently exapted for various eukaryote-specific adaptations. Due to limitations of space, we do not provide a comprehensive survey of all known instances of the above processes. Instead, we attempt to highlight the importance of these processes in the emergence of key adaptations, not just in early eukaryotes, but also during their subsequent evolution, with diverse illustrations emerging from recent investigations. We must emphasize that in this article we mainly use published examples that have been reported in several individual studies on various biological systems or protein families. However, this is the first time they are being brought together to create a coherent picture. A detailed presentation of the methodological apparatus for sequence, structure and phylogenetic analysis of the presented examples is precluded due to limitations of space. However, we refer readers to the individual studies from which we draw our examples for details regarding the computational analysis of the proteins considered here. We use these to develop a conceptual framework for understanding the importance of the diversifying forces acting during biological conflicts in facilitating adaptations that played a role in the so-called "major transitions" of eukaryotic evolutions (Maynard Smith and Szathmáry, 1995).

MATERIALS AND METHODS

Sequence profile searches to establish the relationships between protein domains were performed using the PSI-BLAST (Altschul et al., 1997) and JACKHMMER (Eddy, 2009) programs that run against the non-redundant (NR) protein database of National Center for Biotechnology Information (NCBI). For most searches which were used to report the relationships presented in this work a cut-off *e*-value of 0.01 was used to assess significance. This was further confirmed with other aids such as secondary structure prediction and superposition on known structures, if available. Protein sequences were clustered using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html) to identify related sequences in gene neighborhoods. Multiple sequence alignments of all domains were built by the Kalign (Lassmann et al., 2009) and PCMA programs (Pei et al., 2003), followed by manual adjustments on the basis of profile-profile and structural alignments. Secondary structures were predicted using the JPred program (Cuff et al., 1998). A comprehensive database of profiles was then constructed using these multiple alignments and was used extensively in the annotation and analysis of protein domain architectures and gene neighborhoods. For other known domains, the Pfam database (Finn et al., 2010) was used as a guide, though the profiles were augmented in several cases by addition of newly detected divergent members that were not detected by the original Pfam models. Clustering with BLASTCLUST, followed by multiple sequence alignment, and further sequence profile searches were used to identify other domains that were not present in the Pfam database. Signal peptides and transmembrane segments were detected using the TMHMM and Phobius programs (Kall et al., 2007). The HHpred program was used for profile-profile comparisons to either unify poorly characterized families of proteins or find homologous structures in the PDB database (Soding et al., 2005). Structure similarity searches were performed using the DaliLite program (Holm et al., 2008). Preliminary phylogenetic analysis was conducted using a rapid but approximate-maximum-likelihood method implemented in the FastTree 2.1 program under default parameters (Price et al., 2010). In-house bench-marking suggested that these results are generally comparable to complete ML implemented in the Phylip (Proml) and Molphy packages (Felsenstein, 1989; Adachi and Hasegawa, 1992). Predicted lateral transfers to eukaryotes were further evaluated for false positives by ensuring they were embedded in contigs or complete chromosome sequences with other genes typical of eukaryotes, comparing exon-intron structure of the genes, studying their phyletic distribution within eukaryotes and comparing the protein distances of the predicted eukaryotic proteins (as measured by bit scores) with bacterial homologs. Structural visualization and manipulations were performed using the PyMol (http://www.pymol.org) program. Automatic aspects of large-scale analysis of sequences, structures and genome context were performed by using the in-house TASS package, which comprises a collection of Perl scripts.

RESULTS AND DISCUSSION

PARASITE-HOST CONFLICTS: EMERGENCE OF APICOMPLEXAN SURFACE PROTEINS FOR HOST INTERACTION DUE TO LATERAL TRANSFER

Apicomplexa are a remarkable clade of alveolate eukaryotes entirely comprised of highly specialized metazoan parasites (Levine, 1988; Vivier and Desportes, 1990). With other alveolates, such as ciliates, colpodellids, perkinsids and dinoflagellates, they share organelles known as extrusomes, which allow delivery of a payload of proteins into target cells, such as their prey or hosts (Leander and Keeling, 2003). While basal apicomplexans, the archigregarines, are partial endoparasites that insert only the forepart of their cell into the host cells to suck nutrients, the derived apicomplexans are obligate endoparasites that reside entirely within the cells they invade (Leander et al., 2006). Basal apicomplexans typically have a single-host, but many of the derived apicomplexans like the malarial parasite *Plasmodium* and *Theileria* have evolved lifecycles with two distinct hosts (Levine, 1988; Vivier and Desportes, 1990). Genome analysis of multiple apicomplexans ranging from the relatively basal *Cryptosporidium* to the highly derived *Plasmodium* have shown that they have evolved a remarkable set of secreted or membrane-anchored (surface) proteins that interact with host molecules as a part of the invasion process or other cytoadherance events during their lifecycle (Kaslow et al., 1988; Kappe et al., 1998, 1999; Anantharaman et al., 2007; Arredondo et al., 2012). While surface proteins in each apicomplexan lineage show a wide-range of lineage-specific domains (e.g., the Rifins and Dbl domain proteins in *P. falciparum*), they also contain a striking array of domains that are also found in surface proteins of animals (Patthy, 1999; Anantharaman et al., 2007) (**Figure 1**). Case by case phylogenetic analysis revealed that at least 18 types of non-catalytic

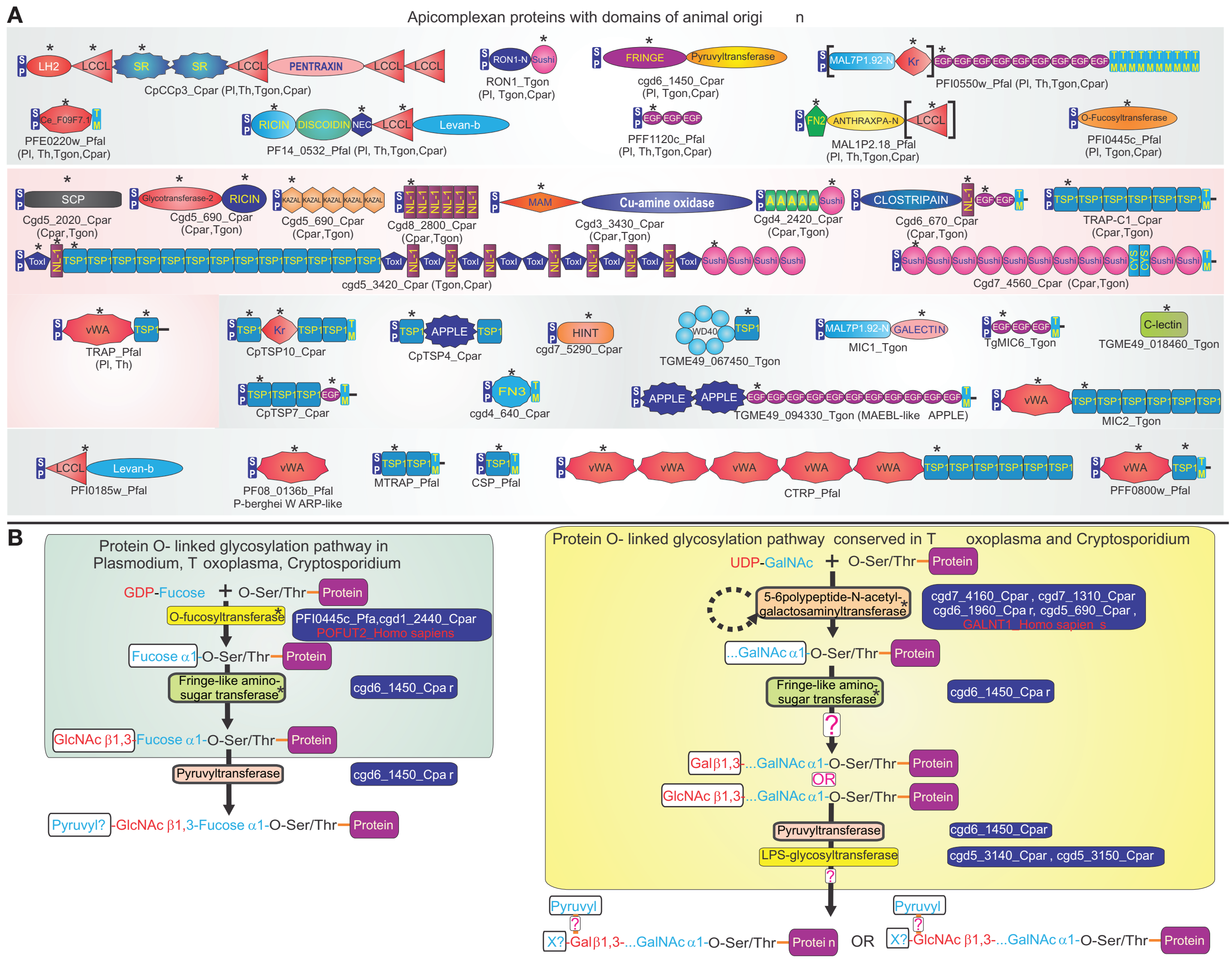


FIGURE 1 | Animal domains and animal-type O-glycosylation systems in apicomplexa. (A) Domain architectures of apicomplexan proteins containing adhesion domains of animal origin. Proteins are labeled by their gene names/common names and species abbreviation separated by an underscore, and are grouped based on their conservation in apicomplexans. If a domain architecture is present in more than one distinct apicomplexan lineage, the additional lineages are shown in brackets. Domains of animal origin are marked with an asterisk above the domain. If a domain is present in multiple copies in a protein, only one (the first) instance of it is labeled with an asterisk. Domains not present in all orthologs of a protein are enclosed in square brackets. Standard abbreviations are used for domains. Species abbreviations are as follows: Cpar: *Cryptosporidium parvum*, Pl: *Plasmodium*, Pfal: *Plasmodium falciparum*, Th: *Theileria*, Tgon: *Toxoplasma gondii*. **(B)** Protein O-linked glycosylation pathways of animal provenance in apicomplexans. Gene names of enzymes involved in these pathways are shown to the right of the enzyme, along with examples of orthologous proteins from animals. The reconstructed oligosaccharide chain is represented using abbreviations for various sugars and functional groups. Speculative parts are marked with a "?". GalNAc: N-acetylgalactosamine, GlcNAc: N-acetylglucosamine, X? indicates an uncharacterized sugar added by the LPS glycosyltransferase. Enzymes of animal origin are marked with an asterisk. Species abbreviations are as in **(A)**.

domains from apicomplexans are otherwise found only in the animal lineage, or alternatively are most closely related to versions found in the animal lineage (Anantharaman et al., 2007) (**Figure 1**). Functional studies in metazoans suggest that majority of these domains, such as the thrombospondin-1 (TSP1), sushi/CCP, MAM, fibronectin-type 2, scavenger receptor, kringle, and vWA domains are involved in adhesive interactions between proteins or proteins and carbohydrates on the cell-surface (Bork, 1993; Patthy, 1999). More recently structural analysis has revealed that the SRS and s48/45 domains, respectively, from coccidian and aconoidasidan apicomplexans, were probably derived through rapid sequence divergence from the ephrin-like domain found in metazoan signaling molecules (Arredondo et al., 2012). Genome analysis suggests that while some of these "animal-like" domains were acquired early in apicomplexan evolution, yet others were acquired only later by specific lineages (**Figure 1**) (Anantharaman et al., 2007). This suggests that the acquisition of a structurally diverse, but functionally comparable group of domains from their animal hosts has been a persistent feature of apicomplexan evolution. Although functional studies on apicomplexan surface proteins with animal domains are still in relatively early stages, two major themes are beginning to emerge: (1) Some of these

proteins appear to have a parasite-specific function in relation to their sexual development, such as in gamete fusion (Pradel et al., 2004; Arredondo et al., 2012). (2) Most others have been adapted for a diverse set of interactions pertaining to invasion of host cells or localization to particular tissues and are often secreted via specialized extrusomes of apicomplexans known as rhoptries (Bradley and Sibley, 2007; Santos and Soldati-Favre, 2011). Particularly striking is the recruitment of the TSP1-domain-containing adhesins early in apicomplexan evolution as part of the conserved invasion apparatus that depends on a cytoskeletal gliding motor unique to apicomplexans (Soldati-Favre, 2008).

Genome analysis has also revealed that apicomplexans possess an animal O-like glycosylation system with two separate arms performing the fucosylation and N-acetylgalactosaminylation of hydroxyl groups of serines or threonine on target proteins (Anantharaman et al., 2007) (**Figure 1**). The first of these has at its core two enzymes, the protein O-fucosyltransferase and a *Drosophila* fringe-like glycosyltransferase that elongates the initial fucose chain with N-acetylglucosamine (Varki et al., 1999; Luo et al., 2006). Also associated with this pathway is the fucose-GDP transporter that allows parasites to take up fucose (Luhn et al., 2001). Interestingly, this pathway modifies TSP1 and EGF domains, both of which appear to have been acquired by apicomplexans through lateral transfer from animals (**Figure 1**). The second pathway displays three distinct orthologous groups of proteins, which constitute the enzyme complex that transfers UDP-linked N-acetylgalactosamine to mucin-like target proteins typified by homopolymeric stretches of serines and threonines (Varki et al., 1999). Phyletic and phylogenetic analysis revealed that enzymes of both these arms of the O-linked glycosylation system and the fucose transporter are specifically related to their animal counterparts to the exclusion of homologs from any other lineage (Anantharaman et al., 2007). Furthermore, their phyletic patterns suggest that the glycosylation pathways were acquired in the common ancestor of endoparasitic apicomplexans, though they were either partially lost in haemosporidians or completely lost in piroplasms. Interestingly, in the more basal apicomplexans, like *Cryptosporidium* and the coccidians, there is a lineage-specific expansion of surface proteins with mucin-like S/T stretches, which are likely to be primary targets of the second arm of the glycosylation system (Stwora-Wojczyk et al., 2004; Anantharaman et al., 2007). Given the gut parasitism of these apicomplexans, it is possible that these glycosylated mucin-like proteins helped homotypic interactions with the gut mucosa, which is also enriched in surface mucins (McGuckin et al., 2011). However, emergence of vertebrate blood parasitism in haemosporidians and piroplasms probably rendered these useless, and perhaps even maladaptive due to the immune response directed against them, thereby favoring their loss.

Thus, apicomplexan genomics suggests that not just adhesion domains of surface proteins, but also entire modification pathways for them were acquired on account of lateral gene flow from their hosts. It appears likely that gene transfer from the host facilitated by the initial parasitic contact allowed the development of elaborate host interaction proteins that might have been central to the emergence of the intimate endoparasitism observed in apicomplexans.

COMMON MOLECULAR ADAPTATIONS OBSERVED IN INTER-ORGANISMAL, INTER-GENOMIC AND INTRA-GENOMIC CONFLICTS

In contrast to the above-discussed example, where a unique set of adaptations emerged due to lateral transfer in course of an evolving host-parasite conflict, several other molecular adaptations appear to be common across a wide-range of biological conflicts. These commonalities appear to be a consequence of two disparate forces: (1) Convergent evolution due to strong selection for particular types of molecular interactions in conflicts; (2) Rapid dispersion over wide phylogenetic distances of certain highly effective adaptations by lateral transfer. We briefly outline some of these adaptations below.

Deployment of proteinaceous toxins

Proteinaceous toxins are the mainstay across all major levels of biological conflict. Such toxins are seen in competition between multicellular eukaryotes (e.g., castor bean ricin, *Aspergillus* sarcin and various snake venom proteins) and between them and their pathogens (e.g., anti-microbial peptide toxins and defensive RNases such as RNase A and RNase L)(Rochat and Martin-Eauclaire, 2000; Rosenberg, 2008; Wiesner and Vilcinskas, 2010). Conversely, such toxins are also used by pathogenic and symbiotic bacteria directed against their hosts (e.g., the cholera toxin and the shiga toxin) (Aepfelbacher et al., 2000; Alouf and Popoff, 2006). Similarly, the importance of protein toxins is becoming apparent in inter-bacterial conflicts (Schwarz et al., 2010; Russell et al., 2011; Iyer et al., 2011b; Zhang et al., 2011). In this regard, an exciting recent discovery has been made of a highly prevalent system of secreted multi-domain toxins, primarily involved in intra-specific conflict between related strains of prokaryotes (Aoki et al., 2011; Iyer et al., 2011b; Zhang et al., 2011). These proteins are typified by the tendency to vary their C-terminal toxin domains through a process of recombination that replaces an existing toxin domain by a distinct one encoded by standalone cassettes, while retaining the rest of the protein's architecture (i.e., N-terminal regions related to trafficking and presentation) intact (Zhang et al., 2011). Hence, these toxins are termed polymorphic toxins. They include contact-dependent versions, which have long N-terminal stalks comprised of RHS/YD or filamentous haemagglutinin repeats that present the C-terminal toxin domain at the tip, shorter diffusible versions, and versions injected or delivered via type VI and ESX/type VII secretory systems (Aoki et al., 2011; Iyer et al., 2011b; Zhang et al., 2011). Importantly, they share these delivery/presentation mechanisms with those toxins using conflicts with hosts (Schwarz et al., 2010). However, they are distinguished from them by the presence a specific immunity protein encoded by a gene downstream of the toxin gene (Aoki et al., 2011; Zhang et al., 2011). Given their role in intra-specific conflict, they are an important determinant of kin-recognition and thereby have an effect on the included fitness in prokaryotes. Inter-genomic conflicts between cellular genomes and selfish replicons residing in the same cell (e.g., classical bacteriocins and plasmid addiction toxins) and intra-genomic conflicts between selfish elements and the host genome (restriction-modification (R-M) systems and genomic toxin-antitoxin (TA) systems) also use protein toxins with related domains (Cascales et al., 2007;

Zhang et al., 2011). The protein toxins of TA systems enable them to act as selfish elements that favor their own retention or "addiction" by killing cells where they are lost or disrupted. However, they might also enhance the fitness of their prokaryotic host. Indeed, expression of chromosomally embedded TA systems has been observed in diverse pathogens such as *Mycobacterium tuberculosis* and *Brucella abortus* when they are replicating within human cells. Here, the action of the toxin actually helps the bacteria to persist effectively in the hosts (Korch et al., 2009; Heaton et al., 2012).

There are some frequently recurrent themes in these toxins deployed across different levels of biological conflict: Most prominent are enzymatic toxins that disrupt the flow of biological information—nucleases targeting genomic DNA, tRNAs and rRNAs, nucleic acid base glycosylases, nucleic acid-modifying enzymes such as deaminases, peptidases that cleave key protein targets, and protein-modifying enzymes such as ADP-ribosyltransferases and AMP/UMPylating enzymes that alter the properties of proteins, such as components of the signaling and translation apparatus (Anantharaman and Aravind, 2003; Cascales et al., 2007; Leplae et al., 2011; Zhang et al., 2011). For example, toxins with the restriction endonuclease (REase) or HNH/ENDOVII folds are seen in intra-specific, inter-specific, inter-genomic (i.e., plasmid-encoded colicins) and intra-genomic conflicts (Stoddard, 2005; Cascales et al., 2007; Zhao et al., 2007; Zhang et al., 2011). Alternatively they disrupt cellular integrity by forming pores in cellular membranes (Gilbert, 2002). The enzymatic domains deployed in these conflicts are characterized by rapid sequence and structure divergence due to selection arising from immunity proteins and resistance against them.

Use of small molecule toxins

Deployment of small molecule toxins or antibiotics, synthesized via dedicated secondary metabolism pathways, is another common strategy, primarily observed in inter-organismal conflicts (Walsh, 2003). They are particularly common in bacteria, and in certain eukaryotic clades, such as fungi and plants. Several distinct types of such molecules are synthesized, with aminoglycosidic, fatty-acid-based (polyketide) and peptide-based skeletons being prevalent (Walsh, 2003). These basic skeletons, which are often synthesized by large multi-domain or multi-protein complexes catalyzing one or more rounds of endoergic condensations of acyl moieties or amino acids, are typically subject to a wide variety of modifications enzymes such as 2-oxoglutarate-dependent hydroxylases, methylases and oxidoreductases (Walsh, 2003; Iyer et al., 2009, 2010). Related to antibiotics are siderophores that are secreted for chelation of essential environmental metals (Barry and Challis, 2009). While not being toxic, they are the center of inter-organismal conflict because several bacteria have evolved receptors for uptake of "non-self" siderophores that allow them to benefit from siderophores produced by other organisms in the environment (Lee et al., 2012). Organisms combat such siderophore-stealing by diversifying their siderophores through modifications similar to those of antibiotics (Samel et al., 2008). Similar pressures also apply to small molecule signals that are used, especially by bacteria, to communicate with each other, as they can also be potentially exploited by non-kin organisms (Brady et al., 2004). Thus, the related secondary metabolism pathways for antibiotic, signaling molecule and siderophore biosynthesis are under pressure for rapid diversification due to pressures from resistance and stealing. In most bacteria, components of these secondary metabolism pathways are encoded by multi-gene operons, which, as indicated by the large number of dioxygenases and oxidoreductases encoded by them, appear to have radiated concomitant with the first oxygenation event in Earth's history (Iyer et al., 2010). Subsequently, they appear to have undergone diversification through recruitment of multiple non-ribosomal peptide ligases and acyl condensation enzymes, sequence divergence of individual enzymatic components, and recombination between distinct biosynthetic operons to synthesize new products (Walsh, 2003; Samel et al., 2008; Iyer et al., 2009, 2010).

Enzymes that facilitate mobility and replication of selfish elements

The fitness of intra-genomic and intra-cellular selfish elements depends on a variety of enzymes that allow their efficient propagation. One group of these enzymes is directly involved in the replication and transcription of the selfish DNA and provides autonomy from the host replication and transcription systems (Galun, 2003; Burt and Trivers, 2006). These enzymes include DNA polymerases, RNA polymerases, primases and reverse transcriptases, which in certain cases are distantly related to the cellular counterparts and in other cases, represent distinct, non-homologous enzymes with analogous activities. These enzymes often face selective pressures for diversification due to exploitation by defective or satellite element which lack their own replication or due to host defensive mechanisms (Galun, 2003; Burt and Trivers, 2006). Another widely used group of enzymes that do not directly catalyze nucleic acid synthesis are transposases/integrases, which often display nuclease domains related to the nuclease domain of toxins (see above) (Lilley and White, 2000; Galun, 2003; Stoddard, 2005; Burt and Trivers, 2006; Zhao et al., 2007; Mak et al., 2010; Zhang et al., 2011). One frequently encountered catalytic domain across a wide-range of transposons is a transposase/integrase domain of the RNAseH fold which is related to the nuclease domain found in the archaeal NurA and the argonaute nucleases (Aravind et al., 2000; Nowotny, 2009). This suggests that several of these mobile elements share an ultimate common ancestry in the form of an ancient RNAseH integrase. Additionally, these enzymes from selfish elements are characterized by a mélange of structurally distinct DNA-binding domains (DBDs), which diversify considerably due to pressures for specific recognition of sequences in the selfish elements (Babu et al., 2006).

Immunity systems

Antagonistic actions in biological conflicts are countered by a variety of dedicated immunity mechanisms, which act over and beyond the immunity gained via sequence divergence of targeted proteins. The polymorphic toxins, plasmid-borne bacteriocins, TA, and R-M systems are all characterized by the presence of an antidote or immunity protein that neutralizes the toxin produced by them (Kobayashi, 2001; Leplae et al., 2011; Russell et al.,

2011; Zhang et al., 2011). Thus, they channelize their antagonistic effects primarily against non-self replicons lacking the protective immunity proteins. Conflicts between cellular and viral genomes have selected for the emergence of multiple dedicated immunity mechanisms. Both prokaryotes and eukaryotes have evolved their own dedicated RNA-based mechanisms, respectively, the CAS/CRISPR and the RNAi systems, which utilize the complementarity of processed RNA to target invasive replicons (Allis et al., 2006; Grewal, 2010; Makarova et al., 2011). Bacteria additionally have evolved less-understood DNA-based mechanisms such as the Abi and the Pgl systems to counter bacteriophages (Sumby and Smith, 2002; Chopin et al., 2005). In eukaryotes, lineage-specific expansions and concomitant sequence diversification of particular receptor molecules, commonly those with leucine-rich repeats (LRRs) are exploited to provide receptors for recognition of viral and bacterial pathogens ("antigen receptors") (Pancer and Cooper, 2006). In some cases, LRR and other domains might be combined with the SCF-type ubiquitin E3-ligases to allow degradation of proteins encoded by invasive replicons or cells (Thomas, 2006). In the vertebrate lineage, on two independent occasions, elaborate mechanisms involving mutagenesis and recombination have evolved to enable diversification of pathogen receptors, which respectively, utilize the immunoglobulin domain and LRRs (Pancer and Cooper, 2006).

COMMONALITIES IN THE MULTIPRONGED APPROACH OF INTRA-CELLULAR BACTERIA AND VIRUSES IN MANIPULATING EUKARYOTIC HOSTS

Endosymbiotic/parasitic bacteria utilize a multipronged approach by often simultaneously deploying several toxins or effectors, each with its own mode of action to manipulate the behavior of the eukaryotic hosts in which they reside. Yet genomics of these bacteria suggests that there is a relatively small set of strategies that are exploited by intra-cellular bacteria from across the bacterial tree, including representatives of alphaproteobacteria, gammaproteobacteria, chlamydiae, and bacteroidetes (Collingro et al., 2005; Ogata et al., 2006; Penz et al., 2010; Schmitz-Esser et al., 2010; Georgiades et al., 2011). The most commonly used approach is the deployment of proteins that alter action of the ubiquitin system, including E3-ligases with RING, U-Box and F-Box domains, deubiquitinating and desumoylating peptidases, especially of the OTU and SMT4/Ulp1-like families and ubiquitin-like (Ubl) proteins (Loureiro and Ploegh, 2006; Lomma et al., 2010; Penz et al., 2010; Schmitz-Esser et al., 2010). Such effectors are seen in several bacteria such as the chlamydiae, like *Chlamydia, Protochlamydia* and *Waddlia*, proteobacteria like *Odysella*, *Wolbachia* and *Legionella*, and the bacteroidetes *Amoebophilus* (**Figure 2**). Protein modification by the action of toxins/effectors with ADP-ribosyltransferase, DOC-type AMP/UMPylase, protein methylases and protein kinase domains is another widely used strategy common to several bacteria such as *Yersinia*, *Xanthomonas*, *Legionella, Amoebophilus,* and *Waddlia* (Yarbrough et al., 2009; Aravind et al., 2011; Feng et al., 2012). These modifying enzymes target proteins from various host systems such as chromatin and signaling proteins. Recent studies have indicated that deployment of diverse nucleic-acid-targeting effectors is also a common feature of numerous endoparasites/endo-symbionts. For example, effectors/toxins with nucleic deaminase domains are seen in *Orientia*, *Wolbachia,* and *Amoebophilus* (Zhang et al., 2011; Iyer et al., 2011b). Likewise, several of these bacteria also share effectors with different nuclease domains that might target both DNA and RNA. Interestingly, studies on eukaryotic viruses suggest that several of viruses also deploy a similar class of molecules. For example, numerous ubiquitin system components, including ubiquitin, SUMO and Apg8-like proteins, E3-ligases and deubiquitinating/desumoylating peptidases are encoded by nucleo-cytoplasmic large DNA viruses, baculoviruses and herpesviruses (Iyer et al., 2006b). Several Ubl proteins are also observed in polyproteins of eukaryotic RNA viruses (Burroughs et al., 2012). Similarly, protein kinases, ADP-ribosyltransferases and some other protein-modifying enzymes are also observed in several NCDLVs and baculoviruses such as the *Agrotis segetum* granulovirus (Iyer et al., 2006b; De Souza and Aravind, 2012).

Among the endosymbiotic bacteria, *Amoebophilus* and *Protochlamydia*, which infect amoebozoan eukaryotes, are particularly striking in that a notable fraction of their proteomes is comprised of diverse effectors with different kinds of catalytic domains (Collingro et al., 2005; Schmitz-Esser et al., 2010). These include numerous ubiquitin system proteins, kinases and α/β hydrolases, which might function as lipases, RNases and REase-fold DNAses (**Figure 2**). Also notable are the *Amoebophilus* effectors with a GTPase domain related to the animal GIMAP GTPases and the AIG1-like GTPases of plants, which play a role in providing scaffolds on intra-cellular membranes (Schwefel et al., 2010). It is conceivable that bacterial effectors with these GTPase domains play a comparable role in remodeling the host membranes surrounding intra-cellular bacteria. Interestingly, such GIMAP GTPases are also encoded by certain animal RNA viruses (e.g., Duck hepatitis A virus) and herpesviruses (e.g., Anguillid herpesvirus 1). Together, the above observations suggest that there are relatively few ancient routes to achieve successful colonization of eukaryotic cells. These appear to have emerged, in part convergently, and in part via lateral transfer of certain effective catalytic toxin/effector domains between unrelated or distant intra-cellular residents of eukaryotes. Interestingly, the genomes of such endosymbiotic bacteria [e.g., *Wolbachia* (Nikoh et al., 2008)] and or DNA viruses [e.g., a Herpesvirus inserted into the genome of the amphioxus (De Souza et al., 2010)] can be integrated into host genomes. Thus, they serve as an effective conduit for transfer of symbiont/parasite adaptations to their hosts.

EVOLUTION OF MAJOR EUKARYOTIC SYSTEMS: CONTRIBUTION FROM PROTEINS DEPLOYED IN INTER-ORGANISMAL, INTER-GENOMIC AND INTRA-GENOMIC CONFLICTS

In this section of the article we discuss with examples as to how several of the above-discussed players deployed in biological conflicts have played a major role in the emergence and elaboration of various eukaryotic adaptations. In doing so we take examples both from early events close to eukaryogenesis and also systems that evolved in particular eukaryotic lineages, such as metazoans.

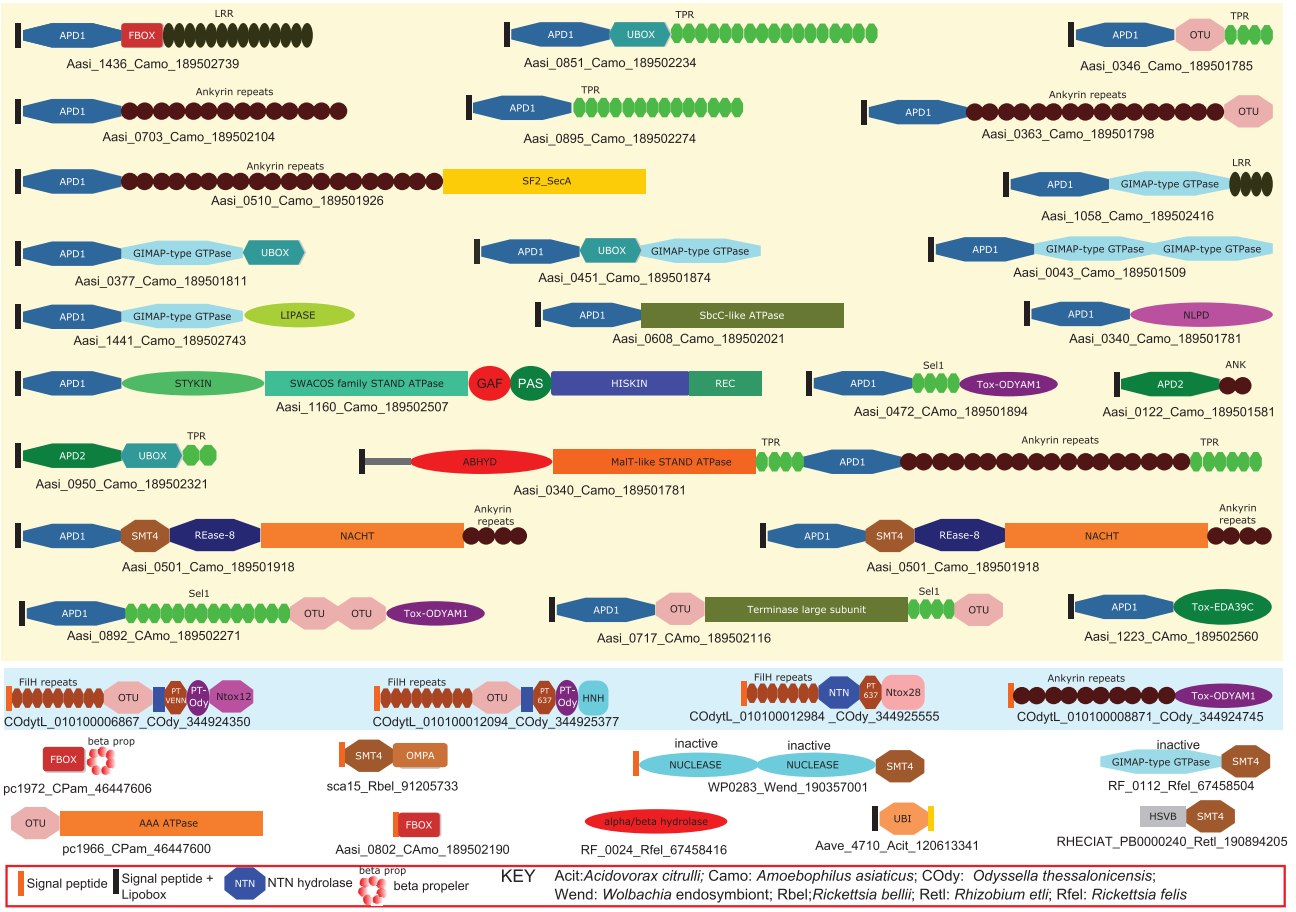


FIGURE 2 | Domain architectures of effectors deployed by endosymbiotic/parasitic bacteria illustrating certain common functional strategies. Proteins are labeled by their gene names, species abbreviations and genbank index (GI) numbers separated by underscores. Non-standard domain names and expansion of species abbreviations are given in the key below the figure. Additionally, Amoebophilus prodomain 1 (APD1) and Amoebophilus prodomain 2 (APD2) are *Amoebophilus*-specific N-terminal domains that are present immediately downstream of a signal peptide and a lipobox. These domains are likely to help in the specific localization and/or clustering of effectors from this organism.

Emergence of key players in eukaryotic chromatin protein complexes

Eukaryotes are distinguished from the two prokaryotic superkingdoms by their dynamic chromatin organized by histones with low complexity tails, which provides a veritable "ecosystem" for several protein-modifying and ATP-dependent remodelers (Allis et al., 2006; Kouzarides, 2007; Aravind et al., 2011; Iyer et al., 2011a). The mysterious origins of several of the unique components of eukaryotic chromatin have begun to considerably clear up with recent genomic data. SWI2/SNF2 ATPases, which had at least six representatives by the time of the last eukaryotic common ancestor (LECA), had already diversified to perform several distinct chromatin remodeling activities, such as sliding/ejection of nucleosomes, exchange of canonical nucleosomes with those containing alternative histones, or altering nucleosomal spacing (Iyer et al., 2008b; Hauk and Bowman, 2011; Hota and Bartholomew, 2011). Phylogenetic, domain architecture, and gene neighborhood analysis revealed that SWI2/SNF2 ATPases are superfamily II DNA helicases, which had their most extensive diversification as part of R-M systems and related systems that are likely to function as a defensive mechanism against bacteriophages (related to the phage growth limitation or Pgl system) (Iyer et al., 2008b) (**Figure 3**). In phylogenetic trees, the eukaryotic versions are nested within the radiation of SWI2/SNF2 ATPases from prokaryotic selfish elements and were transferred on at least three independent occasions to eukaryotes (**Figure 3A**). The first of these transfers occurred prior to the LECA, and by the time of the LECA had proliferated to spawn at least six distinct lineages (Iyer et al., 2008b). The remaining two transfers occurred much later in eukaryotic evolution, and gave rise to the Strawberry Notch and HARP-like SWI2/SNF2 ATPases (**Figure 3**) (Iyer et al., 2008b). Bacterial R-M systems contributed a second ATP-dependent chromatin remodeling enzyme to eukaryotes, the MORC ATPase, which contains a composite module comprised of gyrase, histidine kinase, and MutL (GHKL) and S5 domains (Iyer et al., 2008a). Analysis of R-M bacterial systems showed that they display a vast radiation of several different types of GHKL-S5 module ATPases, of which the MORCs form

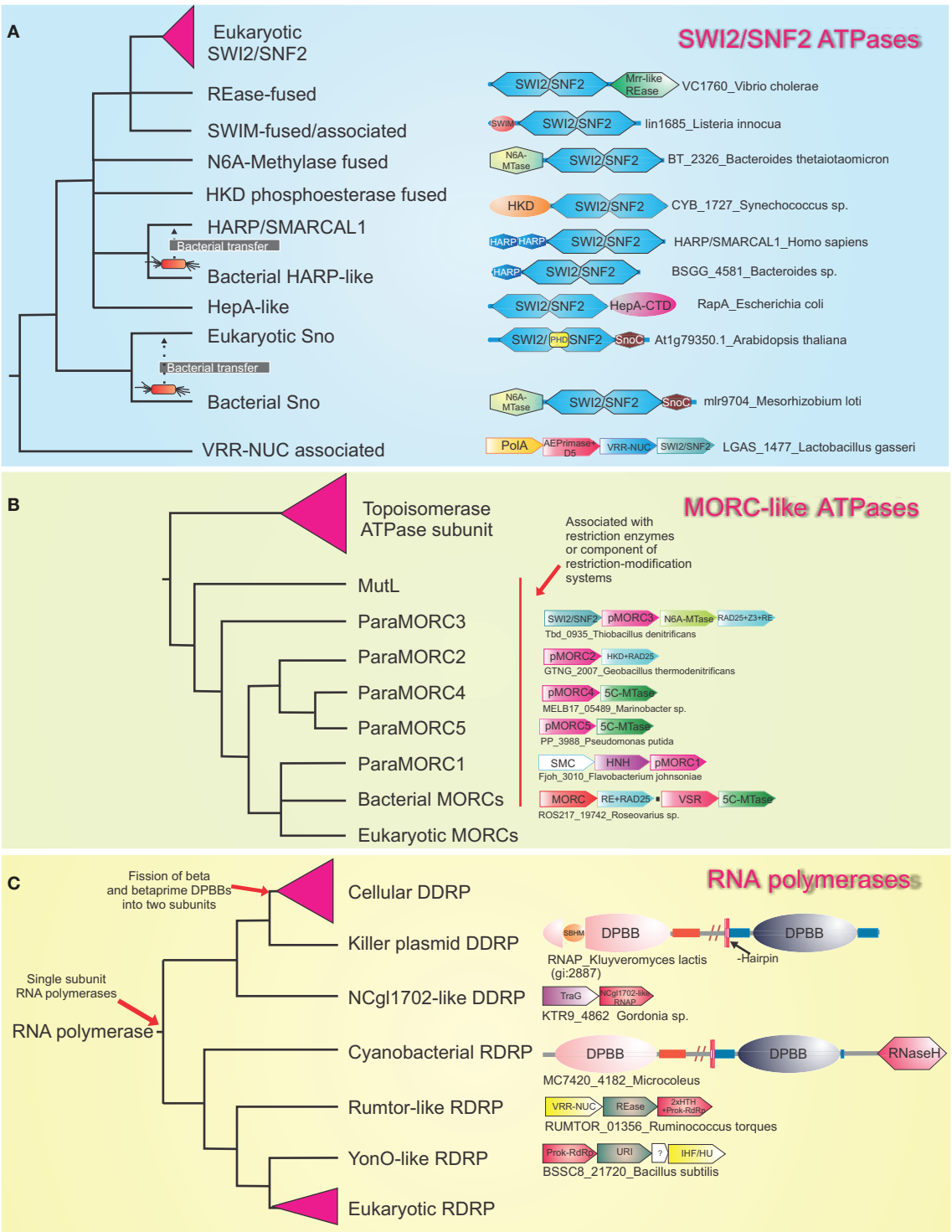


FIGURE 3 | Evolutionary relationships of various families of enzymes illustrating the origin of eukaryotic versions within radiations of systems involved in inter- and intra-genomic conflicts. Reconstructed phylogenetic trees are shown for **(A)** The bacterial radiation of the SWI2/SNF2 ATPases. **(B)** MORC-like ATPases and **(C)** The Double-psi beta barrel containing RNA polymerases. Certain clades with multiple families such as the eukaryotic SWI2/SNF2 ATPases, the Topoisomerase ATPase subunits, the cellular DDRP and eukaryotic RdRPs are collapsed into triangles for clarity. Illustrative domain architectures or gene neighborhoods are shown next to the leaf. Genes in gene neighborhoods are shown in block arrows with the arrow head pointing from the 5′ to the 3′ gene. Proteins and gene neighborhoods are labeled by the gene name and species name separated by underscores. The trees represent only the overall topology because they were obtained by a combination of conventional phylogenetic tree construction and structure-based determination of higher-order relationships.

one distinct clade (**Figure 3B**). Given that basal excavate lineages, such as parabasalids and diplomonads lack MORCs, they appear to have been acquired by eukaryotes post-LECA, prior to the radiation of the large eukaryotic clade uniting animals, fungi, amoebozoans, and plants (Iyer et al., 2008a) (**Figure 4**). Both the MORCs and the SWI2/SNF2 ATPases use ATP hydrolysis to catalyze DNA-unwinding or large-scale looping of DNA in aiding the restriction activity of the REases. This activity has been

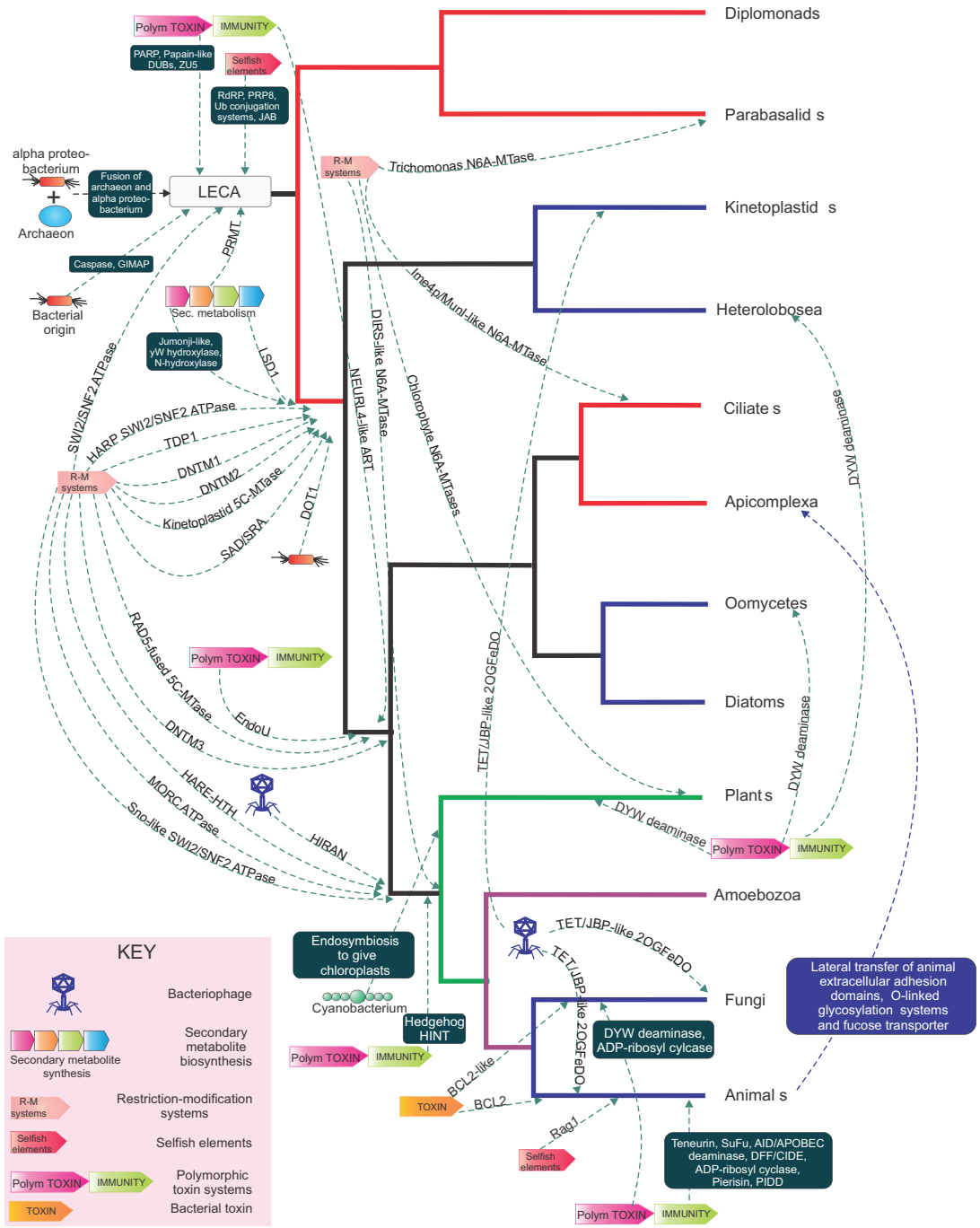


FIGURE 4 | A tree of the eukaryotic relationships illustrating the points of recruitment in eukaryotes in different functional systems of various domains from different biological conflict systems. With a eukaryotic tree as reference, the source and reconstructed point of transfer of various domains recruited from different conflict systems and symbiogenic events are shown. The transfers are shown as dashed arrows with the arrow head pointing to the ancestor in which the transfer is proposed to have taken place. The dashed lines are labeled either with a single gene or a set of genes enclosed in a box. The conflict systems are shown in the key at the bottom left.

reused in a biochemically comparable, but functionally distinct, context to remodel protein-DNA contacts or facilitate higher-order looping in eukaryotic chromatin. In a similar vein, R-M systems might also account for the origin of the eukaryotic phosphoesterase enzyme TDP1, which hydrolyzes 3'-phosphotyrosyl bonds between DNA and the active tyrosine of topoisomerase Ib to release DNA from topoisomerase adducts (Gajewski et al., 2012). Sequence relationships of TDP1 suggest that it is likely to have been derived from HKD phosphoesterase domains found fused to SWI2/SNF2 ATPases in bacterial R-M systems (Iyer et al., 2006a).

Similar studies have shown that the DNA methylases of eukaryotes, which play an important role as encoders of epigenetic information that goes over and beyond the basic genetic information, also largely owe their origin to R-M systems and related methylation systems that protect prokaryotic genomes against restriction attacks by selfish R-M systems (Bestor, 1990; Iyer et al., 2011a). Both DNA cytosine (C5) and adenine (N6) methylases of eukaryotes appear to have been derived from bacterial R-M system and dcm methylases on more than 10 independent occasions (Iyer et al., 2011a). As none of the conserved eukaryotic lineages of DNA methylases can be detected in the parabasalids and diplomonads, it appears that the classical epigenetic DNA modification of cytosine was absent in the LECA. The primary conserved cytosine DNA methylase of eukaryotes, DNMT1, appears to have emerged only just before the time the heterolobosean-kinetoplastid clade branched off from the remaining eukaryotes, and phylogenetic analysis strongly supports its origin from a bacterial R-M system methylase-related to M.NgoFVII (Iyer et al., 2011a). Most other DNA methylases of eukaryotes can be attributed to comparable later acquisitions, primarily from other types of R-M systems. Recent discoveries have indicated that the reversal of cytosine DNA methylation in several eukaryotic lineages occurs via the action of Tet-JBP family of 2-oxoglutarate and iron-dependent dioxygenases (2OGFeDOs), which remove the methyl group through oxidative conversion to hydroxymethylcytosine and further oxidized cytosine derivatives that are then cleared by base excision repair (He et al., 2011; Iyer et al., 2011a). Interestingly, related enzymes, JBP1/2, catalyze the hydroxylation of thymine in the synthesis of base J, an epigenetic modification observed in kinetoplastids (Vainio et al., 2009). Prior studies on the evolution of 2OGFeDOs revealed that the eukaryotic Tet-JBP enzymes were derived from precursors encoded by caudate bacteriophages (Iyer et al., 2011a). Bacteriophages have been known to display a rich variety of DNA modifications, including hydroxymethylated pyrimidines, which enable them to evade restriction by different R-M systems in the host genome (Gommers-Ampt and Borst, 1995). Thus, the bacteriophage Tet-JBP enzymes appear to have first emerged as part of their counter-restriction strategy, and subsequently recruited to generating and erasing epigenetic marks on DNA upon being transferred to eukaryotes. Multiple studies have also revealed that not just enzymatic domains, but also specific DBDs found in eukaryotic chromatin proteins might have been acquired from bacterial R-M systems and replication apparatus of caudate bacteriophages. The SAD/SRA domain, which is a key player in eukaryotic chromatin as an epigenetic "reader" of hemimethylated cytosine marks, has been derived from the DNA-binding domain of REases from R-M systems that discriminate between hemimethylated and fully methylated sites (Iyer et al., 2011a). Likewise, the recently described HARE-HTH domain, which might have an important role in discriminating the DNA modification generated by the cytosine methylases, and the Tet/JBP enzymes has also evolved from bacterial R-M systems, where it is combined with several distinct REase domains (Aravind and Iyer, 2012). On the other hand, another DNA-binding domain, the HIRAN domain, which among other proteins is associated with the eukaryotic chromatin remodeling RAD5-type SWI2/SNF2 ATPases appears to have emerged from the replication apparatus of caudate bacteriophages (Iyer et al., 2006a).

In stark contrast to chromatin remodeling and epigenetic DNA modifications, enzymes catalyzing epigenetic modifications of proteins in eukaryotic chromatin appear to have extensively drawn from very different types of prokaryotic systems involved in inter-organismal conflict. Two key epigenetic modifications are acetylation of lysines and methylation of both lysines and arginines in histones and other proteins in eukaryotic chromatin (Allis et al., 2006; Kouzarides, 2007). Sequence comparisons show that the eukaryotic arginine methylases (PRMT) have been derived from within a bacterial radiation of peptide methylases (Aravind et al., 2011). The closest bacterial sister groups of the eukaryotic PRMTs are encoded in antibiotic-like secondary metabolite biosynthesis operons that also contain genes for peptide dioxygenases, non-ribosomal peptide synthetases and other peptide-oxidizing enzymes such as LSD1-related amine oxidases (Aravind et al., 2011). Bacterial PRMT domains are also incorporated as domains of gigantic antibiotic biosynthesis enzymes, such as anabaenopeptilide synthetase that synthesizes a peptide toxin of the cyanobacterium *Anabaena* (Rouhiainen et al., 2000; Aravind et al., 2011). Interestingly, the LSD1-like amine oxidases observed in these and other peptide antibiotic/toxin biosynthesis operons are also the precursors of eukaryotic histone demethylases that catalyze oxidative removal of methyl groups from mono- and di-methylated histone H3K4 (Allis et al., 2006; Kouzarides, 2007). All the remaining histone demethylases in eukaryotes belong to one large superfamily of 2-oxoglutarate-dependent dioxygenases known as the Jumonji-related dioxygenases (Iyer et al., 2010). These, along with LSD1, are absent in the earliest-branching eukaryotes such as parabasalids and diplomonads, and first appear as multiple paralogous copies just prior to the divergence of the heterolobosean-kinetoplastid clade from the other eukaryotes (Iyer et al., 2010). However, each of these multiple eukaryotic paralogous lineages have their own bacterial counterparts suggesting that they had already diverged in bacteria before being acquired. In bacteria, like LSD1, they appear in one or more copies in peptide antibiotic/toxin and siderophore biosynthesis operons (Iyer et al., 2010), where they are likely to catalyze multiple oxidative modifications of peptides as previously observed in the biosynthesis of penicillin and its derivatives (Liras and Demain, 2009). Thus, it is plausible that eukaryotes acquired multiple paralogous jumonji-related dioxygenases via the transfer of a single secondary metabolism gene-cluster with multiple versions of these enzymes. In eukaryotes, other than

histone demethylation, they also radiated to give rise to enzymes catalyzing the last step in the generation of the eukaryote-specific $tRNA^{Phe}$ modification, hydroxywybutosine, and protein asparagine hydroxylation (Iyer et al., 2010). In contrast to these, the histone H3K79 methylase Dot1 appears to have emerged from a methylase effector delivered by intra-cellular symbionts and is seen in diverse bacterial endo-symbionts/pathogens of amoeboid protozoans and metazoans, like *Parachlamydia* and *Legionella* (Aravind et al., 2011).

Thus, components from R-M and virus-restriction systems, viral replication apparatus, peptide antibiotic/siderophore biosynthesis systems and effectors of intra-cellular bacteria, which are exemplars of intra-genomic, inter-genomic and inter-organismal conflict systems, have been harnessed as progenitors of distinguishing components of eukaryotic chromatin.

Conflict systems and eukaryotic RNA metabolism

Eukaryotes are characterized by the unique RNAi system, which is typified by small RNAs (usually 23–35 nt in length) that perform a number of roles ranging from post-transcription gene regulation to regulation of chromatin structure (Allis et al., 2006; Grewal, 2010). Of these small RNAs, the siRNA-type RNAs are particularly important in gene-silencing, and might be amplified by a distinctive enzyme of this system, the RNA-dependent RNA-polymerase (RdRP), which can be traced back to the LECA (Salgado et al., 2006; Ruprich-Robert and Thuriaux, 2010; Iyer and Aravind, 2011). Sequence-structure analysis of the RdRP revealed that its two catalytic double-ψ-β-barrel (DPBB) domains are related to the catalytic domain found in the two largest subunits of the cellular RNA polymerases from all life forms (Salgado et al., 2006; Ruprich-Robert and Thuriaux, 2010; Iyer and Aravind, 2011). The search for RdRP cognates outside eukaryotes showed that they are prevalent in certain bacteriophages of firmicutes and also a variety of recently identified novel selfish elements in bacterial genomes (**Figure 3C**) (Iyer and Aravind, 2011). In these potential selfish elements they are often encoded alongside genes for different DNase domains such as those belonging to the REase and URI endonuclease fold, which might aid in the mobility of the elements (**Figure 3C**). The RdRPs might also be combined with RNAse H domain in the cyanobacterial versions suggesting that might function in the context of RNA-DNA hybrids (Iyer and Aravind, 2011). Furthermore, structural analysis of the RNA-polymerases with DPBB catalytic domains showed that the RdRP-like enzymes belonged to a radiation of single-subunit RNA polymerases encoded by variety of selfish elements, from within which the cellular multi-subunit versions emerged via fission of the two catalytic domain-containing segments of the single-subunit enzyme (**Figure 3C**). It appears plausible that these RdRP-like enzymes of intra-genomic selfish elements and bacteriophages primarily arose as enzymes that aided their mobility by potentially acting as primases enabling their replication (Iyer and Aravind, 2011). Upon acquisition by the eukaryotic lineage, prior to the LECA, the enzyme appears to have been recruited as a part of the RNAi systems for amplification of small RNAs. Interestingly, the RdRP is not the only nucleic acid polymerase that has been recruited to RNA metabolism from a prokaryotic selfish element. Recent studies on the domain architectures and sequence relationships of the most conserved splicing factor of eukaryotes Prp8, which is part of the spliceosomal catalytic center, has revealed that it has been derived from the polyprotein of a retroelement replete with the reverse transcriptase, "thumb" and RNaseH domains (Dlakic and Mushegian, 2011). However, in Prp8 the active site of the reverse transcriptase domain is disrupted, suggesting that it merely functions in a nucleic acid-binding capacity rather than as an active enzyme (Dlakic and Mushegian, 2011). It is conceivable that this retroelement was associated with the ancestral group-II introns that invaded the genome in the pre-LUCA period to give rise to the spliceosomal introns of eukaryotes.

On several occasions, components of yet another prokaryotic inter-organismal conflict system, namely the recently characterized polymorphic toxin systems, appear to have contributed to eukaryotic RNA-processing and modification systems (Zhang et al., 2011). In eukaryotes, small nucleolar RNAs (snoRNAs) are required for modification and maturation of rRNA in the nucleolus. In several eukaryotes certain snoRNA, like U16 and U86, are directly released from the introns encoding them by the endonucleolytic action of the EndoU RNase (Laneve et al., 2003). Sequence and structure analysis revealed that the EndoU RNase of eukaryotes is nested within a vast radiation of RNase domains that function as toxins in bacterial polymorphic toxin and related secreted toxin systems (Zhang et al., 2011). Thus, acquisition of the EndoU domain appears to have enabled eukaryotes to bypass splicing to directly release snoRNAs from introns. RNA-editing via deamination of cytosine and adenine has considerably expanded in eukaryotes and is observed not just in tRNAs but also in mRNAs and as part of a counter-viral strategy (Iyer et al., 2011b). The origins of certain divergent metal-dependent nucleic acid deaminase domains, such as those of the AID-APOBEC clade and the DYW clade, which catalyzes massive RNA-editing in plant chloroplasts and mitochondria, were rather unclear until recently (Zehrmann et al., 2011). Analysis of the polymorphic toxins revealed that one of the widely used toxin domains was the nucleic acid deaminase that had greatly diversified in such and related secreted toxins (Iyer et al., 2011b). Importantly, the origin of the both the DYW and AID-APOBEC-like deaminases could be placed within specific prokaryotic toxin groups (see below for details).

Prokaryotic conflict systems and protein-modifying enzyme and second messenger in eukaryotic signaling systems

Recent studies on the diversity of catalytic toxin domains deployed in bacterial polymorphic and related secreted toxins systems are also throwing light on the emergence of what were previously considered uniquely eukaryotic signaling systems (**Figure 4**). One such is the polyADP-ribosylation system, which modifies aspartate, glutamate and lysine side chains in both cytoplasmic and nuclear proteins including histones, with profound effects on DNA repair, chromatin organization, telomere dynamics, centrosomal and mitotic spindle organization, and endosomal trafficking (Ame et al., 2004). The enzymes catalyzing this modification, polyADP-ribosyl polymerases (PARPs), can be traced back to the LECA, but their emergence in eukaryotes remained a mystery (Citarelli et al., 2010). The closest relatives of the PARPs are

found among toxin domains of a toxin used in inter-bacterial conflicts delivered via a distinctive phage-derived, injecting secretory system known as the *Photorhabdus* virulence cassette (Hurst et al., 2004; Zhang et al., 2012). Related PARP domains are also found as effectors of intra-cellular symbionts/parasites of amoebae and metazoa such as *Legionella drancourtii*. Recently, a novel family of ADP-ribosyltransferases (ARTs), distinct from the PARPs, was identified, and typified by the Neurl4 protein of humans (De Souza and Aravind, 2012). These ARTs might have an important role in the organization of the eukaryotic centrosome among other processes. They also seem to have been derived from effectors delivered by endoparasitic bacteria, such as *Waddlia* (Hurst et al., 2004). The use of mono-ADP-ribosyltransferases by diverse bacteria as toxins in intra- and inter-specific conflicts (i.e., polymorphic toxins) and those directed at host proteins is well-known (Koch-Nolte et al., 2008; Laing et al., 2011; De Souza and Aravind, 2012). Indeed, other than the PARPs and Neurl4-like ARTs, the eukaryotes also possess several mono-ARTs which are nested within the radiation of bacterial toxin ARTs. Thus, on more than three occasions eukaryotes appear to have recruited the toxin ART/PARP domains as protein-modifying enzymes, with the event giving rise to the PARPs probably happening before the LECA (**Figure 4**). While in bacteria these enzymes appear to largely function as toxins, in eukaryotes they appear to have been utilized to post-translationally modify proteins and provide an additional level of coding information (Koch-Nolte et al., 2008; Laing et al., 2011). Beyond the events spawning pathways that are widespread in eukaryotes, polymorphic and related toxin systems also appear to have contributed to the origin of signaling systems unique to certain lineages, such as metazoans. In addition to ARTs, other bacterial toxin domains utilizing NAD as a substrate have also been recruited to metazoan signaling. The ADP-ribosyl cyclase domain was previously observed only in animals (in the CD38 and CD157 proteins) and generates two messenger molecules, namely cyclic ADP ribose (cADPr) and nicotinic acid adenine dinucleotide phosphate (NAADP), respectively, from NAD and NADP (Guse and Lee, 2008). The former two nucleotides function as messenger molecules that induce calcium signaling pathways via the ryanodine receptors (Guse and Lee, 2008). The discovery of the ADP-ribosyl cyclase as a toxin domain in bacterial polymorphic toxins provides a potential explanation for the sudden origin of this signaling enzyme in animals (Zhang et al., 2012). Additionally, fungi too appear to have independently acquired this domain from bacteria, suggesting that it might have been recruited on more than one occasion in eukaryotic evolution (Zhang et al., 2012).

The Teneurin/Odd Oz proteins found in metazoans and choanoflagellates function as developmental regulators with a potential role in cell-surface adhesion in diverse processes such as cell migration, neuronal path finding and fasciculation, gonad development, and basement membrane integrity (Minet et al., 1999; Silva et al., 2011). These proteins appear to have been derived from a complete bacterial polymorphic toxin, with both the N-terminal RHS/YD repeats, which form a stalk and the C-terminal toxin domain that is a derived version of the HNH/EndoVII fold (Zhang et al., 2012). While the C-terminal toxin domain has lost its active site residues in the animal lineages, it is cleaved and secreted as a potential neuromodulator (Qian et al., 2004). On the other hand the N-terminal RHS repeats appear to play a role in adhesion between different Teneurin/Odd molecules, which is a key aspect of their cell-cell signaling function (Silva et al., 2011). Other than the toxin domains, certain other domains in eukaryotic signaling pathways have also been acquired from bacterial polymorphic toxin systems. The hedgehog signaling pathway is a eukaryotic signaling pathway initiated by the hedgehog proteins, which undergo autoproteolytic cleavage to release signaling messengers (Ingham et al., 2011). The HINT domain, which catalyzes this autoproteolytic cleavage in the eukaryotic hedgehog proteins, is likely to have been derived from the HINT domains commonly found in bacterial polymorphic toxins, where they apparently facilitate the autoproteolytic release of the C-terminal toxin domain into target cells (Zhang et al., 2011). In metazoans, hedgehog activates a down-stream signaling cascade in target cells to activate the transcription factor Gli (Ingham et al., 2011). The Suppressor of Fused (SuFu) protein tethers the Gli in the cytoplasm in the absence of the hedgehog signal to prevent constitutive activation. This SuFu protein of the animal hedgehog pathway also has its origin in bacterial polymorphic toxin systems, where members of the SuFu superfamily function as immunity proteins that neutralize a structurally diverse range of toxin domains (Zhang et al., 2011).

The eukaryotic ubiquitin system: origin and elaboration

One of the most remarkable features of eukaryotes is the ubiquitin system, which comprises of several parallel enzymatic cascades which ligate Ubiquitin or an Ubl protein to target proteins, typically on a lysine residue (Hochstrasser, 2009). These cascades are typified by an E1 enzyme, which activates the Ub/Ubl terminal COOH group by adenylation and trans-thiolation to transfer it to and E2 enzyme. The E2 enzyme may then either directly or via an E3 enzyme transfer the Ub/Ubl to the target protein. In eukaryotes, such modifications often target proteins for degradation via the proteasomal system, where the Ub/Ubl is first cleaved off and released by a JAB domain metallopeptidase (Kerscher et al., 2006). In addition to proteasomal degradation, Ub/Ubl modifications also alter the interactions, localization and biochemistry of the target proteins and are modulated by a series of peptidases (DUBs) that debiquitinate them (Burrows and Johnston, 2012). Until recently it was thought that the Ub-system was a purely eukaryotic innovation. However, multiple studies have shown that the antecedents of the Ub-system first emerged in prokaryotes as part of a dramatic radiation of Ubls and E1-like enzyme in operons for the biosynthesis of cofactors (e.g. thiamin and molybdopterin), cysteine, and peptide secondary metabolites such as siderophores, antibiotics/toxins and small molecule signals (Burroughs et al., 2011, 2012). A subset of these operons is highly mobile (i.e., widespread dispersal across distant lineages) and evolved features characteristic of the eukaryotic Ub-systems, namely the presence of E2 and sometimes E3 enzymes and the deubiquitinating JAB peptidase (Burroughs et al., 2011). The fact that these operons are mobile, and usually tend to couple the ubiquitinating enzymes with deubiquitinating JAB peptidases, presents parallels to the R-M systems (Iyer et al., 2006c). Like

them these systems combine opposing actions in the modifying and de-modifying enzymes, and have no links to the metabolic enzymes that are typical of the operons with E1-like enzymes and Ubls that synthesize small molecule. Hence, we posit that these are potential selfish elements that act like the R-Ms, but at the protein level, by possibly destabilizing proteins through transfer of the Ubl and restoring the original protein by removal of the Ubl by the JAB peptidase. Since these Ub-like systems are closest to the eukaryotic versions, it is very likely that they were derived from them. On account of their mobility they are seen in several bacteria and certain archaea (e.g., the *Caldiarchaeum*) (Iyer et al., 2006c; Burroughs et al., 2011; Nunoura et al., 2011); hence, it is possible that eukaryotes might have acquired the precursor of their Ub-system either from their archaeal precursor or from endosymbiotic bacteria (**Figure 4**).

The only DUB that is consistently observed in prokaryotic Ub-like systems is the JAB peptidase domain (Iyer et al., 2006c; Burroughs et al., 2011). Eukaryotes, however, possess several other DUBs, most of which belong to the papain-like peptidase fold and a few to the Zincin-like metallopeptidase fold (Iyer et al., 2004). Interestingly, papain-like peptidases (e.g., Otu-like peptidase domain) and Zincin-like metallopeptidases are frequently found among the toxin domains of effectors delivered by a range of intra-cellular bacteria (Loureiro and Ploegh, 2006). These were previously thought to be lateral transfers from hosts to their endo-symbionts/parasites, which are used to interfere with the host Ub-system (Lomma et al., 2010; Schmitz-Esser et al., 2010). However, recent studies on polymorphic toxin systems suggest that such peptidase domains are far more widely distributed in bacterial toxins and often among toxins of free-living bacteria deployed in inter-bacterial conflicts (Zhang et al., 2012). Hence, it seems more likely that they first emerged in bacteria as part of the polymorphic and related secreted toxin systems and were acquired by eukaryotes and recruited as DUBs in course of the development of the mitochondrial endosymbiosis (**Figure 4**). Not surprisingly, these DUB-like peptidase domains are common among intra-cellular bacteria such as *Wolbachia*, *Rickettsia* and *Odyssella*, which are closely related to the mitochondrial precursors (**Figure 2**). Indeed, these DUBs probably originally emerged as part of the strategy utilized by these bacterial endosymbionts/pathogens that countered the immunity mechanism based on ubiquitination of target proteins. Interestingly, several of these papain-like DUB domains are also related to polyprotein-processing peptidases of eukaryotic RNA viruses and retroviruses (Iyer et al., 2004). It is conceivable that the emergence of the Ub-system in eukaryotes also provided a means for RNA viruses to escape constraints placed by the eukaryotic mRNA cap on internal translation initiation, by simply enabling translation of polyproteins that are then processed by the DUB peptidases. In course of viral evolution many of the DUB domains were probably incorporated into their own polyproteins to allow auto-proteolytic processing.

Executers of apoptosis: multiple independent recruitments of domains from prokaryotic conflict systems

One of the simplest counter-pathogen strategies is regulated cell death or apoptosis, in which a cell might sacrifice its own fitness and prevent the pathogen from replicating within it. This typically works in situations where the inclusive fitness accrued from saving kin from infection might contribute to fixation of altruistic behaviors such as apoptosis (Aravind et al., 2009). Such mechanisms are likely to be further enhanced with the emergence of colonial or multicellular organization. Some of the simplest programmed cell death systems seen in bacteria are constituted by intra-genomic selfish elements. For example, in *Escherichia coli* a defective prophage produces a toxin known as Lit with a zincin-like metallopeptidase domain to cleave the elongation factor Tu and kill the cell when infected by the phage T4, thereby preventing the further spread of T4 to remaining cells in the colony (Snyder, 1995). Likewise, under conditions of starvation, when resources are limiting, chromosomally encoded toxin-antitoxin systems, such as the entericidin locus, mediate cell death in bacteria like *E. coli* and allow certain cells to survive and grow at the expense of kin that have undergone cell death (Bishop et al., 1998). Thus, the principle of the use of toxins as mediators of programmed cell death appears to be an ancient one (Jensen and Gerdes, 1995; Bishop et al., 1998). Although eukaryotes lack conventional toxin-antitoxin systems, the executioners of apoptosis resemble the prokaryotic toxins from these and other conflict systems in that they cleave or modify specific target proteins or permeabilize membranes in the cell committed to apoptosis. These effectors have been best studied in the animal lineage and include membrane-permeability regulators (the BCL2 superfamily), DNA-cleaving enzymes (e.g., the DNA fragmentation factor/CIDE), DNA-modifying enzymes (e.g., pierisin) and peptidases (e.g., the caspases) (Chou et al., 1999; Lugovskoy et al., 1999; Kanazawa et al., 2001; Riedl and Salvesen, 2007). Investigation into the provenance of these proteins has revealed multiple ancient connections to bacterial toxin systems. The core helical domain of the BCL2 superfamily (the first 6 helices) is specifically related to the translocation (T) domain of several host-directed toxins from distantly related bacteria such as the diphtheria, botulinum, tetanus and *Vibrio* toxins (Chou et al., 1999). The T-domain undergoes a pH induced conformational change to assume a BCL2-like structure, inserts into the endosomal membrane and transfers the catalytic domain of the toxin into host cytoplasm. Given its sudden emergence in metazoans, it is likely that it was derived from a bacterial toxin and recruited as regulator of the permeability of the mitochondrial membrane. In metazoans these domains diversified into anti-apoptotic versions, which prevent the release of cytochrome C from mitochondria and pro-apoptotic versions which foster its release (Chou et al., 1999; Riedl and Salvesen, 2007). From animals, the BCL2 superfamily was secondarily acquired by large DNA viruses that infect them, such as herpesviruses, poxviruses, iridoviruses and asfarviruses, and used as an anti-apoptotic effector to prevent hosts from using cell death as a defense against them (Iyer et al., 2006b). The T-domain of bacterial toxins appears to have been independently transferred to the fungus *Metarhizium*, where it appears to be utilized in multiple toxins directed against the insect host.

Among catalytic effectors of apoptosis, in metazoans the DFF/CIDE endonuclease catalyzes the genome fragmentation of DNA that is typical of apoptosis (Lugovskoy et al., 1999; Riedl and

Salvesen, 2007). Structural studies had revealed that this domain contains an endonuclease domain of the HNH/EndoVII fold, but its origins remained unclear (Lugovskoy et al., 1999). Recent analysis of the bacterial polymorphic toxins revealed that a subset of them contains a toxin nuclease domain, which shares unique sequence signatures with the DFF/CIDE endonuclease domain to the exclusion of other representatives of HNH/EndoVII fold (Zhang et al., 2012). Here again, the relative abundance of the HNH/EndoVII fold among polymorphic and related toxin domains compared to its lone presence in DFF/CIDE, which is restricted to metazoans, points to an origin for the latter from a representative in the bacterial toxin systems. Pierisin-type ARTs are unusual enzymes that mediate apoptosis (thus far only known from lepidopterans) by ADP-ribosylating the N2 atom of guanine in DNA (Kanazawa et al., 2001). The lepidopteran pierisin-like ARTs are specifically related to the ART toxin domains found in certain bacterial polymorphic toxins and insecticidal toxins of insect pathogens, such *as Bacillus sphaericus* (Orth et al., 2011). This suggests that they were probably laterally transferred into lepidopterans from a bacterial symbiont or parasite, followed by their reuse as an apoptotic effector. In all the above examples the natural action of the bacterial toxins in disrupting or killing animal cells appears to have been harnessed as a mechanism to execute apoptosis.

Caspase-like peptidases are the central executers of apoptosis throughout eukaryotes and have been demonstrated to play a central role in cell death in animals, fungi, plants, and certain other eukaryotes (Aravind and Koonin, 2002; Riedl and Salvesen, 2007). Prior evolutionary analysis of the caspase-like superfamily revealed that they first diversified in bacteria into several clades such as the metacaspases, paracaspases and numerous other bacteria-specific lineages (Aravind and Koonin, 2002). Metacaspases were transferred to eukaryotes prior the LECA and are found in most eukaryotes (**Figure 4**). Subsequently, in the animal lineage and in dictyostelid slime molds metacaspases were displaced by a second acquisition from bacteria, the paracaspases, which then radiated in animal to give rise to the classical caspases (Aravind and Koonin, 2002). This phyletic pattern suggests that paracaspases are effectively functionally comparable to metacaspases, as they have displaced them on more than one occasion. Interestingly, several bacteria, particularly endosymbiotic/parasitic alphaproteobacteria (e.g., *Agrobacterium, Labrenzia, Bradyrhizobium*) encode metacaspases and paracaspases with N-terminal signal peptides that are likely to be secreted into their hosts (Aravind and Koonin, 2002). Hence, these peptidases were possibly first used in regulating endoparasite/symbiont-host conflicts to modulate the immune response and cell death in favor of the intra-cellular bacterium. Consistent with this, recent studies in humans have shown that the paracaspase modulates the T-cell-dependent immune response by cleaving A20, a deubiquitinating enzyme involved in the process, and is required for prevention of cell death in diffuse large B cell lymphoma (Coornaert et al., 2008; Ferch et al., 2009). This suggests that caspase-like peptidases might have been acquired on multiple occasions in eukaryotic evolution from endosymbiotic bacteria, which were probably utilizing them to regulate the survival of their host cells. On a similar note, the GIMAP/AIG1-like GTPases, which are deployed by certain endo-symbionts/parasites (**Figure 2**), could have given rise to the eukaryote representatives of this clade which are known to modulate both apoptosis and the immune response.

Thus, protein domains that originally diversified in prokaryotic conflict systems both as toxin and also as potential modulators of host defensive responses have had a notable effect on the evolution of apoptosis.

Origin of antigen receptor diversification mechanisms and mutagenic immunity mechanisms

Despite the enormous disparities in the immune systems of different eukaryotes, there are a few common strategies that are observed across most of them. These include the use of a relatively small number of families of protein domains as antigen receptors. Diversification of antigen receptors in most eukaryotes is a passive process of sequence divergence, probably under positive selection, within families of lineage-specifically expanded proteins (e.g., LRR proteins). However, in both jawed and jawless vertebrates two distinct and directed mechanisms for their diversification have been observed, namely recombination and active mutagenesis, which result in different populations of lymphocytes expressing different types of antigen receptors (Pancer and Cooper, 2006; Schatz and Swanson, 2011). In both jawed and jawless vertebrates the process of directed mutagenesis by DNA cytosine deaminases of the AID-APOBEC superfamily is utilized (Rogozin et al., 2007). Such mutagenesis is used either as a trigger for antigen gene-conversion, or for hypermutation or for antibody class-switching. Additionally, certain representatives of the AID-APOBEC family of cytosine deaminases are also major line of defense against retroviruses by mutagenizing their genomes by cytosine deamination (Chiu and Greene, 2006). Although AID-APOBEC-like deaminases were, until recently, thought to be restricted to vertebrates, sensitive sequence analysis showed that more divergent members exist in nematodes, cnidarians and several distantly related algal groups. Identification of these sequences helped establish that the fast-evolving AID-APOBEC deaminases have their ultimate origin in the toxin domains of polymorphic and related secreted bacterial toxins (Iyer et al., 2011b). Indeed, effectors with toxin domains most closely related to the AID-APOBEC deaminases are observed in the *Wolbachia* endosymbiont of the moth *Cadre cautella* and the plant pathogen *Pseudomonas brassicacearum* (Iyer et al., 2011b). Thus, these mutagenic deaminase domains, which were originally part of toxins deployed by bacteria, appear to have provided the basis for the unique mechanism for antigen receptor diversification in vertebrates. However, their role in anti-retroviral response suggests that they were probably initially recruited merely as mutagenic enzymes that targeted viruses (i.e., similar to the original toxin role but merely directed at viruses). Interestingly, several filamentous fungi show a lineage-specific expansion of related nucleic acid deaminases that also appear to have been derived from toxin domains of bacterial provenance (Iyer et al., 2011b). It is conceivable that these play a similar role as the counter-retroviral deaminases in potentially mutating cytoplasmic parasitic elements or preventing anastomosis by unrelated hyphae.

In jawed vertebrates, antibody and T-cell receptor diversity is generated by the action (V-D-J and V-J recombination) of a dedicated recombination apparatus comprised of two proteins Rag1 and Rag2, of which Rag1 is the catalytic subunit of the recombinase (Schatz and Swanson, 2011). The origin of the Rag1 recombinase in animals had remained mysterious until it was shown that their recombinase domain is related to the transposase domain of a distinct class of eukaryotic transposons known as the Transib elements (Kapitonov and Jurka, 2005; Panchin and Moroz, 2008). This transposase domain contains a distinctive version of the RNAseH fold and cleaves sites associated with the termini of these transposons, which show sequence relationships to V-D-J and V-J recombination sites. Thus, the Rag1 recombinase appears to have evolved from a "domesticated" selfish element whose recombinase domain and terminal recognition sites were reused as a mechanism to generate diversity. Indeed, domestication of selfish elements for generation of diversity in host-pathogen interfaces is a general phenomenon, which is not restricted to the animal immune system: In certain caudate bacteriophages, the mutagenic reverse transcriptase of an integrated retroelement has been shown to play a role in creating sequence diversity in a tail-fiber-associated protein (Medhekar and Miller, 2007). This allows the bacteriophages to recognize a changing landscape of cell-surface proteins on their hosts.

Was the origin of the eukaryotic nucleus-related to inter-organismal and intra-genomic conflicts?

As the endosymbiotic model for eukaryogenesis involves juxtaposition of two distinct genomes in the same cell, it implies an increased scope for genetic conflicts between the genomes and the intra-genomic selfish elements contained by them. Indeed, different scenarios exploiting such conflicts have been proposed. One of these argues that the mobile self-splicing group-II introns from the alphaproteobacterial mitochondrial progenitor invaded and proliferated in the progenitor of the nuclear genome (Koonin, 2006). As a consequence there was selection for the nuclear membrane as a physical barrier to protect unspliced intron-containing transcripts from the translation apparatus. This hypothesis posits that the pre-LECA eukaryotes were enormously enriched in introns (Koonin, 2006) as a consequence of reduced selection due to decreased effective population sizes (Lynch, 2007). However, direct evidence for highly intron-rich pre-LECA genomes is lacking based on available genomes and with the current data it is not possible to distinguish between: (1) the early proliferation of introns in eukaryotes being a consequence of the emergence of a protective barrier of the nucleus and (2) the nucleus being a consequence of the selective pressure imposed by intron proliferation. Moreover, there is little evidence for extensive proliferation of group-II introns in any prokaryotic lineage. In an alternative hypothesis, greater alignment of the genetic interests of the genomes of the pro-mitochondrion and the nucleus is likely to have happened with the transfer of genes, including those encoding ribosomal proteins, from the former genome to the latter (Jekely, 2008). This is likely to have resulted in chimeric ribosomes in the cytoplasm with potentially deleterious effects for both genomes. This hypothesis presents the nucleus as a physical barrier to prevent such chimerism and might also effectively explain the origin of the nucleolus, another defining feature of eukaryotes. It should be noted that nucleus-like structures have convergently evolved in certain representatives of the clade of bacteria uniting the planctomycetes, chlamydiae and verrucomicrobia (McInerney et al., 2011). In these cases there is no evidence for deleterious effects arising from intra-genomic selfish elements like group-II introns or ribosomal chimerism. Indeed alternative selective pressures could have facilitated nucleogenesis.

One key feature of bacterial endo-symbionts/parasites is their deployment of toxin/effector systems that contain nuclease and nucleic acid deaminase domains, both from polymorphic and host-directed toxin systems (Iyer et al., 2011b; Zhang et al., 2011, 2012). These are observed in a variety of extant endo-symbionts/parasites such as *Wolbachia, Rickettsia, Orientia, Odyssella, Legionella, Amoebophilus, and Protochlamydia* (**Figure 2**). Indeed, such genome-targeting toxins are likely to play a role in the chromosomal disruptions produced by *Wolbachia* in the process of regulating sex-specific survival and killing of incompatible hosts (Duron, 2008). Interestingly, a key nuclear pore component, Nup96/98, has an autoproteolytic ZU5 domain (Mans et al., 2004). ZU5 domains appear to have originated in bacterial cell-surface proteins, such as polymorphic toxins, and play a role in the autoproteolytic processing of toxins on the cell-surface [ZU5 domains were also secondarily acquired again from bacterial sources to give rise to the animal apoptosis regulator PIDD (Riedl and Salvesen, 2007; Zhang et al., 2012)]. It is possible that this key nuclear pore component was derived from a toxin system of the ancestral endosymbiont. Thus, it is likely that nucleic-acid-targeting toxins were deployed by the mitochondrial progenitor, which could have threatened the integrity of the DNA of the nuclear genome precursor. Hence, the nucleus was probably selected for, as a physical barrier to minimize this threat. In this scenario, once the initial endosymbiotic association between the mitochondrial precursor and the archaeon was underway, the selective pressure from the DNA-targeting toxins of the mitochondrial precursor favored the emergence of the nucleus very early in the development of the association. The early presence of the nucleus then favored the development of several characteristics of eukaryotes, including those that have been noted in the other hypotheses: (1) it would have allowed transfer of alphaproteobacterial ribosomal genes to the nuclear genome, as chimerism could be avoided due to presence of an additional compartment (Jekely, 2008), eventually leading the origin of the nucleolus. (2) It allowed retroelements associated with group-II introns to proliferate in nucleus (Koonin, 2006). This not only gave rise to introns but also the telomerase (Aravind et al., 2006). (3) The telomerase in turn facilitated the origin of multiple linear chromosomes, whose expression could now be coordinated as they were contained within the nuclear compartment. (4) Linear chromosomes, together with the nucleus, probably selected against the prokaryotic pumping mechanisms for chromosome segregation based on HerA-FtsK-like ATPases, and instead favored a cytoskeleton-based mechanism, which allowed for fixation of the microtubular apparatus. (5) The stabilization of multiple linear chromosomes contained with a nucleus also probably allowed for increased genome sizes in eukaryotes, as it removed the

constraints coming from containing the entire genome on a large circular chromosome segregated by the ATPase pumps.

In conclusion a number of mechanistically distinct scenarios support a role for organismal and genomic conflict systems in eukaryotic nucleogenesis. Further investigations of alternative scenario presented here might provide a new handle to understand key events in eukaryogenesis.

GENERAL CONCLUSIONS

In the above discussion, we provide a series of examples from across the eukaryotic phyletic spread for how the interplay between lateral gene flow inter-organismal, inter-genomic, and intra-genomic conflicts has shaped the evolution of numerous functional systems (**Figure 4**). These examples are by no means meant to be exhaustive—rather, they were chosen in order to provide a glimpse of the sheer variety of biological systems that are affected by the evolutionary contributions from such systems. One key theme that emerges from the above discussion is that domain families gained through lateral transfer in course of intimate inter-organismal associations, such as symbiosis and parasitism, can notably determine the very nature of these interactions. This is strikingly illustrated by the case of apicomplexan adhesion molecules implicated in host interaction: here manifold domains were acquired by the parasites via lateral transfer from their hosts, spawning unique "animal-like" interfaces for interacting with the host (**Figure 1**). The other recurrent theme, which transcends various biological systems, is how proteins/protein domains originally emerging in the context of various biological conflicts were recycled as regulatory molecules (**Figure 4**). Of these host-directed toxins, and the toxins, immunity proteins, structural modules and secretory components from bacterial polymorphic toxin systems have a distinct life beyond their locus of provenance in eukaryotic regulatory and defense systems (Iyer et al., 2011b; Zhang et al., 2011, 2012). We outline numerous occasions where these components were incorporated into regulatory systems of eukaryotes, and sometimes might have played a major role in the very origin of these systems. This process appears to be constantly on-going, all the way from the origin of eukaryotes to the terminal tips of the eukaryotic tree (**Figure 4**). The reason why proteins derived from biological conflict systems appear to be recruited for other functions might be attributed to the consequences of natural selection. Not surprisingly, toxin-immunity systems used in inter-organismal conflict have a large effect on the fitness of both the organisms producing toxins and those defending against them, thereby escalating an arms race situation. Many of the conflict systems deployed by bacteria might even function at the interface of symbiotic and parasitic interactions of bacteria and eukaryotes, thereby developing adaptations to effectively target components of eukaryotic systems. Toxins and immunity proteins of intra-genomic selfish elements are also under multiple levels of selection that foster their diversification. At one level they are under selection to evade host resistance to function effectively as addictive agents. At another level many of them might also be under selection to function as effective stress response mechanisms that allow their host genomes to survive adverse conditions. Consequently, there are strong selective pressures for constant diversification of toxins and the corresponding immunity proteins in various conflict systems. Hence, these biological conflicts could have functioned as evolutionary "nurseries" for innovations in both prokaryotic and eukaryotic proteins. Hence, lateral gene flow from symbionts, parasites and other modes of DNA uptake (Gladyshev et al., 2008; Nikoh et al., 2008) has enabled eukaryotes to have access to and import a "readymade" set of molecular innovations from such biological conflict systems. When recruited in non-conflict biological contexts, they can in turn spur the emergence of new interactions in eukaryotic systems. Thus, number of key eukaryotic innovations can be traced back to the above-described players in biological conflict systems, such as secondary metabolism operons, R-M, polymorphic and host-directed toxins systems, anti-phage systems, phage counter-restriction strategies, and mobile elements. These systems appear to have particularly expanded in bacteria on account of the presence of operons, extensive lateral transfer with several modes of DNA uptake and recombination, perhaps combined with high effective population sizes (Lynch, 2007). Thus, organismal and genomic conflicts as the basis for major molecular innovations, which in turn might facilitate major evolutionary transitions, can be considered a general evolutionary principle.

ACKNOWLEDGMENTS

The authors' research is supported by the intramural funds of the US Department of Health and Human Services (National Library of Medicine, NIH).

REFERENCES

Adachi, J., and Hasegawa, M. (1992). *MOLPHY: Programs for Molecular Phylogenetics.* Tokyo: Institute of Statistical Mathematics.

Aepfelbacher, M., Aktories, K., and Just, I. (2000). *Bacterial Protein Toxins.* Berlin, New York: Springer.

Allis, C. D., Jenuwein, T., Reinberg, D., and Caparros, M. (2006). *Epigenetics.* New York, NY: Cold Spring Harbor Laboratory Press.

Alouf, J. E., and Popoff, M. R. (2006). *The Comprehensive Sourcebook of Bacterial Protein Toxins.* Amsterdam, Boston: Elsevier Academic Press.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Ame, J. C., Spenlehauer, C., and De Murcia, G. (2004). The PARP superfamily. *Bioessays* 26, 882–893.

Aminov, R. I., and Mackie, R. I. (2007). Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol. Lett.* 271, 147–161.

Anantharaman, V., and Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* 4, R81.

Anantharaman, V., Iyer, L. M., Balaji, S., and Aravind, L. (2007). Adhesion molecules and other secreted host-interaction determinants in Apicomplexa: insights from comparative genomics. *Int. Rev. Cytol.* 262, 1–74.

Aoki, S. K., Poole, S. J., Hayes, C. S., and Low, D. A. (2011). Toxin on a stick: modular CDI toxin delivery systems play roles in bacterial competition. *Virulence* 2, 356–359.

Aravind, L., Abhiman, S., and Iyer, L. M. (2011). Natural history of the eukaryotic chromatin protein methylation system. *Prog. Mol. Biol. Transl. Sci.* 101, 105–176.

Aravind, L., Anantharaman, V., and Venancio, T. M. (2009). Apprehending multicellularity: regulatory networks, genomics, and evolution. *Birth Defects Res. C Embryo Today* 87, 143–164.

Aravind, L., and Iyer, L. M. (2012). The HARE-HTH and associated

domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* 11, 119–131.

Aravind, L., Iyer, L. M., and Koonin, E. V. (2006). Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr. Opin. Struct. Biol.* 16, 409–419.

Aravind, L., and Koonin, E. V. (2002). Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* 46, 355–367.

Aravind, L., Makarova, K. S., and Koonin, E. V. (2000). Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* 28, 3417–3432.

Arredondo, S. A., Cai, M., Takayama, Y., Macdonald, N. J., Anderson, D. E., Aravind, L., Clore, G. M., and Miller, L. H. (2012). Structure of the Plasmodium 6-cysteine s48/45 domain. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6692–6697.

Babu, M. M., Iyer, L. M., Balaji, S., and Aravind, L. (2006). The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 34, 6505–6520.

Barry, S. M., and Challis, G. L. (2009). Recent advances in siderophore biosynthesis. *Curr. Opin. Chem. Biol.* 13, 205–215.

Batut, J., Andersson, S. G., and O'callaghan, D. (2004). The evolution of chronic infection strategies in the alpha-proteobacteria. *Nat. Rev. Microbiol.* 2, 933–945.

Bertelli, C., Collyn, F., Croxatto, A., Ruckert, C., Polkinghorne, A., Kebbi-Beghdadi, C., Goesmann, A., Vaughan, L., and Greub, G. (2010). The waddlia genome: a window into chlamydial biology. *PLoS ONE* 5:e10890. doi: 10.1371/journal.pone.0010890

Bestor, T. H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 326, 179–187.

Bhattacharya, D., Yoon, H. S., and Hackett, J. D. (2004). Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays* 26, 50–60.

Bishop, R. E., Leskiw, B. K., Hodges, R. S., Kay, C. M., and Weiner, J. H. (1998). The entericidin locus of *Escherichia coli* and its implications for programmed bacterial cell death. *J. Mol. Biol.* 280, 583–596.

Bork, P. (1993). The modular architecture of a new family of growth regulators related to connective tissue growth factor. *FEBS Lett.* 327, 125–130.

Bowen, D., Rocheleau, T. A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., and Ffrench-Constant, R. H. (1998). Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* 280, 2129–2132.

Bradley, P. J., and Sibley, L. D. (2007). Rhoptries: an arsenal of secreted virulence factors. *Curr. Opin. Microbiol.* 10, 582–587.

Brady, S. F., Chao, C. J., and Clardy, J. (2004). Long-chain N-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* 70, 6865–6870.

Burroughs, A. M., Iyer, L. M., and Aravind, L. (2011). Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. *Mol. Biosyst.* 7, 2261–2277.

Burroughs, A. M., Iyer, L. M., and Aravind, L. (2012). The natural history of ubiquitin and ubiquitin-related domains. *Front. Biosci.* 17, 1433–1460.

Burrows, J. F., and Johnston, J. A. (2012). Regulation of cellular responses by deubiquitinating enzymes: an update. *Front. Biosci.* 17, 1184–1200.

Burt, A., and Trivers, R. (2006). *Genes in Conflict: the Biology of Selfish Genetic Elements.* Cambridge, MA: The Belknap Press of Harvard University Press.

Cascales, E., Buchanan, S. K., Duche, D., Kleanthous, C., Lloubes, R., Postle, K., Riley, M., Slatin, S., and Cavard, D. (2007). Colicin biology. *Microbiol. Mol. Biol. Rev.* 71, 158–229.

Chiu, Y. L., and Greene, W. C. (2006). APOBEC3 cytidine deaminases: distinct antiviral actions along the retroviral life cycle. *J. Biol. Chem.* 281, 8309–8312.

Chopin, M. C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Curr. Opin. Microbiol.* 8, 473–479.

Chou, J. J., Li, H., Salvesen, G. S., Yuan, J., and Wagner, G. (1999). Solution structure of BID, an intracellular amplifier of apoptotic signaling. *Cell* 96, 615–624.

Citarelli, M., Teotia, S., and Lamb, R. S. (2010). Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evol. Biol.* 10, 308.

Collingro, A., Toenshoff, E. R., Taylor, M. W., Fritsche, T. R., Wagner, M., and Horn, M. (2005). '*Candidatus Protochlamydia amoebophila*', an endosymbiont of Acanthamoeba spp. *Int. J. Syst. Evol. Microbiol.* 55, 1863–1866.

Coornaert, B., Baens, M., Heyninck, K., Bekaert, T., Haegman, M., Staal, J., Sun, L., Chen, Z. J., Marynen, P., and Beyaert, R. (2008). T cell antigen receptor stimulation induces MALT1 paracaspase-mediated cleavage of the NF-kappaB inhibitor A20. *Nat. Immunol.* 9, 263–271.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* 14, 892–893.

Dao, D. N., Kessin, R. H., and Ennis, H. L. (2000). Developmental cheating and the evolutionary biology of Dictyostelium and Myxococcus. *Microbiology* 146(Pt 7), 1505–1512.

Dawkins, R., and Krebs, J. R. (1979). Arms races between and within species. *Proc. R. Soc. Lond. B Biol. Sci.* 205, 489–511.

De Souza, R. F., and Aravind, L. (2012). Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. *Mol. Biosyst.* 8, 1661–1677.

De Souza, R. F., Iyer, L. M., and Aravind, L. (2010). Diversity and evolution of chromatin proteins encoded by DNA viruses. *Biochim. Biophys. Acta* 1799, 302–318.

Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A., and Moran, N. A. (2009). *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9063–9068.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154, S164–S177.

Dlakic, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* 17, 799–808.

Doolittle, W. F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 14, 307–311.

Duron, O. (2008). Insights beyond Wolbachia-Drosophila interactions: never completely trust a model: insights from cytoplasmic incompatibility beyond Wolbachia-Drosophila interactions. *Heredity (Edinb)* 101, 473–474.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211.

Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., Bryant, D., Steel, M. A., Lockhart, P. J., Penny, D., and Martin, W. (2004). A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* 21, 1643–1660.

Felsenstein, J. (1989). PHYLIP–Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166.

Feng, F., Yang, F., Rong, W., Wu, X., Zhang, J., Chen, S., He, C., and Zhou, J. M. (2012). A Xanthomonas uridine 5'-monophosphate transferase inhibits plant immune kinases. *Nature* 485, 114–118.

Ferch, U., Kloo, B., Gewies, A., Pfander, V., Duwel, M., Peschel, C., Krappmann, D., and Ruland, J. (2009). Inhibition of MALT1 protease activity is selectively toxic for activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* 206, 2313–2320.

Fields, S. D., and Rhodes, R. G. (1991). Ingestion and retention of Chroomonas spp. (cryptophyceae) by *Gymnodinium acidotum* (dinophyceae). *J. Phycol.* 27, 525–529.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211–D222.

Gabaldon, T., and Huynen, M. A. (2007). From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput. Biol.* 3:e219. doi: 10.1371/journal.pcbi.0030219

Gajewski, S., Comeaux, E. Q., Jafari, N., Bharatham, N., Bashford, D., White, S. W., and Van Waardenburg, R. C. (2012). Analysis of the active-site mechanism of tyrosyl-DNA phosphodiesterase I: a member of the phospholipase D superfamily. *J. Mol. Biol.* 415, 741–758.

Galun, E. (2003). *Transposable elements: A Guide to the Perplexed and The Novice: With Appendices on RNAi, Chromatin Remodeling and*

Gene Tagging. Dordrecht, Boston: Kluwer Academic.

Georgiades, K., Madoui, M. A., Le, P., Robert, C., and Raoult, D. (2011). Phylogenomic analysis of Odyssella thessalonicensis fortifies the common origin of Rickettsiales, Pelagibacter ubique and *Reclimonas americana* mitochondrion. *PLoS ONE* 6:e24857. doi: 10.1371/journal.pone.0024857

Gilbert, R. J. (2002). Pore-forming toxins. *Cell. Mol. Life Sci.* 59, 832–844.

Gladyshev, E. A., Meselson, M., and Arkhipova, I. R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science* 320, 1210–1213.

Goff, L. J., and Coleman, A. W. (1995). Fate of parasite and host organelle dna during cellular transformation of red algae by their parasites. *Plant Cell* 7, 1899–1911.

Gommers-Ampt, J. H., and Borst, P. (1995). Hypermodified bases in DNA. *FASEB J.* 9, 1034–1042.

Grewal, S. I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev.* 20, 134–141.

Guse, A. H., and Lee, H. C. (2008). NAADP: a universal Ca2+ trigger. *Sci. Signal.* 1, re10.

Hauk, G., and Bowman, G. D. (2011). Structural insights into regulation and action of SWI2/SNF2 ATPases. *Curr. Opin. Struct. Biol.* 21, 719–727.

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.

Heaton, B. E., Herrou, J., Blackwell, A. E., Wysocki, V. H., and Crosson, S. (2012). Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *J. Biol. Chem.* 287, 12098–12110.

Hochstrasser, M. (2009). Origin and function of ubiquitin-like proteins. *Nature* 458, 422–429.

Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24, 2780–2781.

Hota, S. K., and Bartholomew, B. (2011). Diversity of operation in ATP-dependent chromatin remodelers. *Biochim. Biophys. Acta* 1809, 476–487.

Huang, J., and Gogarten, J. P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8, R99.

Hurst, L. D., Atlan, A., and Bengtsson, B. O. (1996). Genetic Conflicts. *Q. Rev. Biol.* 71, 317–364.

Hurst, M. R., Glare, T. R., and Jackson, T. A. (2004). Cloning *Serratia entomophila* antifeeding genes–a putative defective prophage active against the grass grub *Costelytra zealandica*. *J. Bacteriol.* 186, 5116–5128.

Ingham, P. W., Nakano, Y., and Seger, C. (2011). Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* 12, 393–406.

Ishikawa, K., Fukuda, E., and Kobayashi, I. (2010). Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems. *DNA Res.* 17, 325–342.

Iyer, L. M., Abhiman, S., and Aravind, L. (2008a). MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Dir.* 3, 8.

Iyer, L. M., Anantharaman, V., Wolf, M. Y., and Aravind, L. (2008b). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* 38, 1–31.

Iyer, L. M., Abhiman, S., and Aravind, L. (2011a). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 101, 25–104.

Iyer, L. M., Zhang, D., Rogozin, I. B., and Aravind, L. (2011b). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* 39, 9473–9497.

Iyer, L. M., Abhiman, S., De Souza, R. F., and Aravind, L. (2010). Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res.* 38, 5261–5279.

Iyer, L. M., Abhiman, S., Maxwell Burroughs, A., and Aravind, L. (2009). Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Mol. Biosyst.* 5, 1636–1660.

Iyer, L. M., and Aravind, L. (2011). Insights from the architecture of the bacterial transcription apparatus. *J. Struct. Biol.* doi: 10.1016/j.jsb.2011.12.013. [Epub ahead of print].

Iyer, L. M., Babu, M. M., and Aravind, L. (2006a). The HIRAN domain and recruitment of chromatin remodeling and repair activities to damaged DNA. *Cell Cycle* 5, 775–782.

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006b). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184.

Iyer, L. M., Burroughs, A. M., and Aravind, L. (2006c). The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. *Genome Biol.* 7, R60.

Iyer, L. M., Koonin, E. V., and Aravind, L. (2004). Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. *Cell Cycle* 3, 1440–1450.

Jekely, G. (2008). Origin of the nucleus and Ran-dependent transport to safeguard ribosome biogenesis in a chimeric cell. *Biol. Dir.* 3, 31.

Jensen, R. B., and Gerdes, K. (1995). Programmed cell death in bacteria: proteic plasmid stabilization systems. *Mol. Microbiol.* 17, 205–210.

Johnson, M. D., Oldach, D., Delwiche, C. F., and Stoecker, D. K. (2007). Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* 445, 426–428.

Kall, L., Krogh, A., and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.* 35, W429–W432.

Kanazawa, T., Watanabe, M., Matsushima-Hibiya, Y., Kono, T., Tanaka, N., Koyama, K., Sugimura, T., and Wakabayashi, K. (2001). Distinct roles for the N- and C-terminal regions in the cytotoxicity of pierisin-1, a putative ADP-ribosylating toxin from cabbage butterfly, against mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2226–2231.

Kapitonov, V. V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* 3:e181. doi: 10.1371/journal.pbio.0030181

Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V., and Menard, R. (1999). Conservation of a gliding motility and cell invasion machinery in Apicomplexan parasites. *J. Cell Biol.* 147, 937–944.

Kappe, S. H., Noe, A. R., Fraser, T. S., Blair, P. L., and Adams, J. H. (1998). A family of chimeric erythrocyte binding proteins of malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1230–1235.

Kaslow, D. C., Quakyi, I. A., Syin, C., Raum, M. G., Keister, D. B., Coligan, J. E., Mccutchan, T. F., and Miller, L. H. (1988). A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* 333, 74–76.

Keeling, P. J. (2004). Diversity and evolutionary history of plastids and their hosts. *Am. J. Bot.* 91, 1481–1493.

Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* 22, 159–180.

Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* 29, 3742–3756.

Koch-Nolte, F., Kernstock, S., Mueller-Dieckmann, C., Weiss, M. S., and Haag, F. (2008). Mammalian ADP-ribosyltransferases and ADP-ribosylhydrolases. *Front. Biosci.* 13, 6716–6729.

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Dir.* 1, 22.

Korch, S. B., Contreras, H., and Clark-Curtiss, J. E. (2009). Three *Mycobacterium tuberculosis* Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. *J. Bacteriol.* 191, 1618–1630.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.

Laing, S., Unger, M., Koch-Nolte, F., and Haag, F. (2011). ADP-ribosylation of arginine. *Amino Acids* 41, 257–269.

Laneve, P., Altieri, F., Fiori, M. E., Scaloni, A., Bozzoni, I., and Caffarelli, E. (2003). Purification, cloning, and characterization of XendoU, a novel endoribonuclease involved in processing of intron-encoded small nucleolar RNAs in *Xenopus laevis*. *J. Biol. Chem.* 278, 13026–13032.

Lassmann, T., Frings, O., and Sonnhammer, E. L. (2009). Kalign2, high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* 37, 858–865.

Leander, B. S., and Keeling, P. J. (2003). Morphostasis in alveolate evolution. *Trends Ecol. Evol.* 18, 395–402.

Leander, B. S., Lloyd, S. A., Marshall, W., and Landers, S. C. (2006). Phylogeny of marine Gregarines (Apicomplexa)–Pterospora, Lithocystis and Lankesteria–and the

origin(s) of coelomic parasitism. *Protist* 157, 45–60.

Lee, W., Van Baalen, M., and Jansen, V. A. (2012). An evolutionary mechanism for diversity in siderophore-producing bacteria. *Ecol. Lett.* 15, 119–125.

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Dreze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525.

Levine, N. D. (1988). *The Protozoan Phylum Apicomplexa.* Boca Raton, FL: CRC.

Lilley, D. M., and White, M. F. (2000). Resolving the relationships of resolving enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9351–9353.

Liras, P., and Demain, A. L. (2009). Chapter 16. Enzymology of beta-lactam compounds with cephem structure produced by actinomycete. *Meth. Enzymol.* 458, 401–429.

Lomma, M., Dervins-Ravault, D., Rolando, M., Nora, T., Newton, H. J., Sansom, F. M., Sahr, T., Gomez-Valero, L., Jules, M., Hartland, E. L., and Buchrieser, C. (2010). The *Legionella pneumophila* F-box protein Lpp2082 (AnkB) modulates ubiquitination of the host protein parvin B and promotes intracellular replication. *Cell. Microbiol.* 12, 1272–1291.

Loureiro, J., and Ploegh, H. L. (2006). Antigen presentation and the ubiquitin-proteasome system in host-pathogen interactions. *Adv. Immunol.* 92, 225–305.

Lugovskoy, A. A., Zhou, P., Chou, J. J., Mccarty, J. S., Li, P., and Wagner, G. (1999). Solution structure of the CIDE-N domain of CIDE-B and a model for CIDE-N/CIDE-N interactions in the DNA fragmentation pathway of apoptosis. *Cell* 99, 747–755.

Luhn, K., Wild, M. K., Eckhardt, M., Gerardy-Schahn, R., and Vestweber, D. (2001). The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. *Nat. Genet.* 28, 69–72.

Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006). Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J. Biol. Chem.* 281, 9385–9392.

Lynch, M. (2007). *The Origins of Genome Architecture.* Sunderland, MA: Sinauer Associates.

Mak, A. N., Lambert, A. R., and Stoddard, B. L. (2010). Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* 18, 1321–1331.

Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Dir.* 6, 38.

Mans, B. J., Anantharaman, V., Aravind, L., and Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3, 1612–1637.

Martin, W., and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41.

Maynard Smith, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution.* Oxford, New York: W. H. Freeman Spektrum.

McGuckin, M. A., Linden, S. K., Sutton, P., and Florin, T. H. (2011). Mucin dynamics and enteric pathogens. *Nat. Rev. Microbiol.* 9, 265–278.

McInerney, J. O., Martin, W. F., Koonin, E. V., Allen, J. F., Galperin, M. Y., Lane, N., Archibald, J. M., and Embley, T. M. (2011). Planctomycetes and eukaryotes: a case of analogy not homology. *Bioessays* 33, 810–817.

Medhekar, B., and Miller, J. F. (2007). Diversity-generating retroelements. *Curr. Opin. Microbiol.* 10, 388–395.

Minet, A. D., Rubin, B. P., Tucker, R. P., Baumgartner, S., and Chiquet-Ehrismann, R. (1999). Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J. Cell Sci.* 112(Pt 12), 2019–2032.

Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M., and Fukatsu, T. (2008). Wolbachia genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res.* 18, 272–280.

Nowotny, M. (2009). Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* 10, 144–151.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G. J., Hattori, M., Kanai, A., Atomi, H., Takai, K., and Takami, H. (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223.

Oborník, M., Janouškovec, J., Chrudimský, T., and Lukeš, J. (2009). Evolution of the apicoplast and its hosts: from heterotrophy to autotrophy and back again. *Int. J. Parasitol.* 39, 1–12.

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P. E., Claverie, J. M., and Raoult, D. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* 2:e76. doi: 10.1371/journal.pgen.0020076

Orth, J. H., Schorch, B., Boundy, S., Ffrench-Constant, R., Kubick, S., and Aktories, K. (2011). Cell-free synthesis and characterization of a novel cytotoxic pierisin-like protein from the cabbage butterfly Pieris rapae. *Toxicon* 57, 199–207.

Palmer, J. D. (2003). The symbiotic birth and spread of plastids: how many times and whodunit? *J. Phycol.* 39, 4–12.

Pancer, Z., and Cooper, M. D. (2006). The evolution of adaptive immunity. *Annu. Rev. Immunol.* 24, 497–518.

Panchin, Y., and Moroz, L. L. (2008). Molluscan mobile elements similar to the vertebrate Recombination-Activating Genes. *Biochem. Biophys. Res. Commun.* 369, 818–823.

Patthy, L. (1999). *Protein Evolution.* London: Blackwell Publishing.

Pei, J., Sadreyev, R., and Grishin, N. V. (2003). PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* 19, 427–428.

Penz, T., Horn, M., and Schmitz-Esser, S. (2010). The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" encodes an afp-like prophage possibly used for protein secretion. *Virulence* 1, 541–545.

Pisani, D., Cotton, J. A., and McInerney, J. O. (2007). Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* 24, 1752–1760.

Pradel, G., Hayton, K., Aravind, L., Iyer, L. M., Abrahamsen, M. S., Bonawitz, A., Mejia, C., and Templeton, T. J. (2004). A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J. Exp. Med.* 199, 1533–1544.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Qian, X., Barsyte-Lovejoy, D., Wang, L., Chewpoy, B., Gautam, N., Al Chawaf, A., and Lovejoy, D. A. (2004). Cloning and characterization of teneurin C-terminus associated peptide (TCAP)-3 from the hypothalamus of an adult rainbow trout *(Oncorhynchus mykiss)*. *Gen. Comp. Endocrinol.* 137, 205–216.

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329.

Riedl, S. J., and Salvesen, G. S. (2007). The apoptosome: signalling platform of cell death. *Nat. Rev. Mol. Cell Biol.* 8, 405–413.

Rivera, M. C., and Lake, J. A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155.

Rochat, H., and Martin-Eauclaire, M.-F. (2000). *Animal Toxins: Facts and Protocols.* Basel Boston, MA: Birkhauser Verlag.

Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* 8, 647–656.

Rosenberg, H. F. (2008). RNase A ribonucleases and host defense: an evolving story. *J. Leukoc. Biol.* 83, 1079–1087.

Rouhiainen, L., Paulin, L., Suomalainen, S., Hyytiainen, H., Buikema, W., Haselkorn, R., and Sivonen, K. (2000). Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90. *Mol. Microbiol.* 37, 156–167.

Ruprich-Robert, G., and Thuriaux, P. (2010). Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases. *Nucleic Acids Res.* 38, 4559–4569.

Russell, A. B., Hood, R. D., Bui, N. K., Leroux, M., Vollmer, W., and Mougous, J. D. (2011). Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* 475, 343–347.

Salgado, P. S., Koivunen, M. R., Makeyev, E. V., Bamford, D. H., Stuart, D. I., and Grimes, J. M. (2006). The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* 4:e434. doi: 10.1371/journal.pbio.0040434

Samel, S. A., Marahiel, M. A., and Essen, L. O. (2008). How to tailor non-ribosomal peptide products–new clues about the structures and mechanisms of modifying enzymes. *Mol. Biosyst.* 4, 387–393.

Santos, J. M., and Soldati-Favre, D. (2011). Invasion factors are coupled

to key signalling events leading to the establishment of infection in apicomplexan parasites. *Cell. Microbiol.* 13, 787–796.

Sapp, J. (2007). "Mitochondria and their host: morphology to molecular phylogeny," in *Origin of Mitochondria and Hydrogenosomes*, eds W. F. Martin and M. Müller (Berlin, Heidelberg: Springer), 57–83.

Sassera, D., Beninati, T., Bandi, C., Bouman, E. A., Sacchi, L., Fabbi, M., and Lo, N. (2006). *'Candidatus Midichloria mitochondrii'*, an endosymbiont of the tick Ixodes ricinus with a unique intramitochondrial lifestyle. *Int. J. Syst. Evol. Microbiol.* 56, 2535–2540.

Schatz, D. G., and Swanson, P. C. (2011). V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* 45, 167–202.

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and Horn, M. (2010). The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J. Bacteriol.* 192, 1045–1057.

Schwarz, S., West, T. E., Boyer, F., Chiang, W. C., Carl, M. A., Hood, R. D., Rohmer, L., Tolker-Nielsen, T., Skerrett, S. J., and Mougous, J. D. (2010). Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog.* 6:e1001068. doi: 10.1371/journal.ppat.1001068

Schwefel, D., Frohlich, C., Eichhorst, J., Wiesner, B., Behlke, J., Aravind, L., and Daumke, O. (2010). Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2). *Proc. Natl. Acad. Sci. U.S.A.* 107, 20299–20304.

Silva, J. P., Lelianova, V. G., Ermolyuk, Y. S., Vysokov, N., Hitchen, P. G., Berninghausen, O., Rahman, M. A., Zangrandi, A., Fidalgo, S., Tonevitsky, A. G., Dell, A., Volynski, K. E., and Ushkaryov, Y. A. (2011). Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12113–12118.

Skippington, E., and Ragan, M. A. (2011). Lateral genetic transfer and the construction of genetic exchange communities. *FEMS Microbiol. Rev.* 35, 707–735.

Smith, J. M., and Price, G. R. (1973). The logic of animal conflict. *Nature* 246, 15–18.

Snyder, L. (1995). Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.* 15, 415–420.

Soding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248.

Soldati-Favre, D. (2008). Molecular dissection of host cell invasion by the apicomplexans: the glideosome. *Parasite* 15, 197–205.

Stoddard, B. L. (2005). Homing endonuclease structure and function. *Q. Rev. Biophys.* 38, 49–95.

Stwora-Wojczyk, M. M., Kissinger, J. C., Spitalnik, S. L., and Wojczyk, B. S. (2004). O-glycosylation in *Toxoplasma gondii*: identification and analysis of a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Int. J. Parasitol.* 34, 309–322.

Sumby, P., and Smith, M. C. (2002). Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3. *Mol. Microbiol.* 44, 489–500.

Thomas, J. H. (2006). Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res.* 16, 1017–1030.

Vainio, S., Genest, P. A., Ter Riet, B., Van Luenen, H., and Borst, P. (2009). Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol. Biochem. Parasitol.* 164, 157–161.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999). *Essentials of Glycobiology*. New York, NY: Cold Spring Harbor Laboratory Press.

Vivier, E., and Desportes, I. (1990). "Phylum Apicomplexa," in *Handbook of Protoctista*, eds L. Margulis, J. O. Corliss, M. Melkonian, and D. J. Chapman (Boston, MA: Jones and Bartlett Publishers), 549–573.

Walsh, C. (2003). *Antibiotics: Actions, Origins, Resistance.* Washington, DC: ASM Press.

Werren, J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 2), 10863–10870.

Wiesner, J., and Vilcinskas, A. (2010). Antimicrobial peptides: the ancient arm of the human immune system. *Virulence* 1, 440–464.

Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., and Orth, K. (2009). AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science* 323, 269–272.

Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., and Takenaka, M. (2011). PPR proteins network as site-specific RNA editing factors in plant organelles. *RNA Biol.* 8, 67–70.

Zhang, D., De Souza, R. F., Anantharaman, V., Iyer, L. M., and Aravind, L. (2012). Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Dir.* doi: 10.1186/1745-6150-7-18. [Epub ahead of print].

Zhang, D., Iyer, L. M., and Aravind, L. (2011). A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Res.* 39, 4532–4552.

Zhao, L., Bonocora, R. P., Shub, D. A., and Stoddard, B. L. (2007). The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J.* 26, 2432–2442.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 May 2012; accepted: 13 June 2012; published online: 29 June 2012.
Citation: Aravind L, Anantharaman V, Zhang D, de Souza RF and Iyer LM (2012) Gene flow and biological conflict systems in the origin and evolution of eukaryotes. Front. Cell. Inf. Microbio. 2:89. doi: 10.3389/fcimb.2012.00089

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 21 August 2012
doi: 10.3389/fcimb.2012.00110

Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms

Claire Bertelli and Gilbert Greub*

Center for Research on Intracellular Bacteria, Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Hua Xie, Meharry Medical College, USA
Mickael Boyer, Danone Research, France

****Correspondence:***
Gilbert Greub, Institute of Microbiology, Rue du Bugnon 48, 1011 Lausanne, Switzerland.
e-mail: gilbert.greub@chuv.ch

Based on Darwin's concept of the tree of life, vertical inheritance was thought to be dominant, and mutations, deletions, and duplication were streaming the genomes of living organisms. In the current genomic era, increasing data indicated that both vertical and lateral gene inheritance interact in space and time to trigger genome evolution, particularly among microorganisms sharing a given ecological niche. As a paradigm to their diversity and their survival in a variety of cell types, intracellular microorganisms, and notably intracellular bacteria, were considered as less prone to lateral genetic exchanges. Such specialized microorganisms generally have a smaller gene repertoire because they do rely on their host's factors for some basic regulatory and metabolic functions. Here we review events of lateral gene transfer (LGT) that illustrate the genetic exchanges among intra-amoebal microorganisms or between the microorganism and its amoebal host. We tentatively investigate the functions of laterally transferred genes in the light of the interaction with their host as they should confer a selective advantage and success to the amoeba-resisting microorganisms (ARMs).

Keywords: amoeba, intracellular bacteria, giant virus, gene transfer, evolution

INTRODUCTION

For many years following the publication of "*The origin of species*" by Charles Darwin (Darwin, 1859), the evolutionary history of the living organisms was represented by structures of trees that represent their common descent. However, this representation ignores the significance and the importance of LGT that allowed ancestral prokaryotes, and further on unicellular eukaryotes, to rapidly increase their genetic variability at a much faster rate than allowed by vertical inheritance, duplications and mutations (Keeling and Palmer, 2008; Lopez and Bapteste, 2009). It was progressively accepted that LGT have contributed to shape bacterial, archaeal, and eukaryotic genomes rending difficult to represent the vertical evolutionary history of these organisms (Andersson, 2005; Lopez and Bapteste, 2009; Olendzenski and Gogarten, 2009; Raoult, 2010b; Danchin and Rosso, 2012). From 1975, several new illustrations of evolution such as a "network-like representation," a "reticulated tree," or a "ring of life" have been proposed to account for the importance of horizontal transfers (Paz and Espinosa, 2010). More recently, to incorporate the theories of multiplicity and de-novo creation of genes, the "rhizome of life" was proposed as a representation of the evolution of species and the chimerism of bacterial genomes (Raoult, 2010b; Merhej et al., 2011).

In the light of the genome sequencing era, a growing number of whole genome analyses assessed the importance of lateral transfer in the constitution of gene repertoire (Doolittle, 1999; Doolittle and Bapteste, 2007) that reflects the organism lifestyle. Symbiotic and parasitic microorganisms, considered as extreme specialists, were shown to undergo genome reduction and have small gene repertoires due to their dependence on multiple host cell factors (McCutcheon and Moran, 2012). On the contrary, amoeba-resisting microorganisms (ARMs), which include both viruses and bacteria, often exhibit larger genomes than their mammalian-infecting relatives (Moliner et al., 2010). Amoebae, as a reservoir of numerous microorganisms sharing a sympatric lifestyle, i.e., microorganisms living in a community within amoebae, were proposed to bring these latter in close contact and facilitate genetic exchanges (Greub and Raoult, 2004; Moliner et al., 2010; Raoult, 2010a; Thomas and Greub, 2010). Indeed, microorganisms in large communities and sharing an ecological niche are more prone to genetic exchanges than isolated populations (Merhej et al., 2009; Raoult, 2010a). In this review, we summarize the recent findings on LGT in amoeba-infecting microorganisms highlighting the complex composite nature of their gene repertoire. Moreover, the function of exchanged genes is discussed in the context of symbiosis or host-pathogen interaction.

AMOEBAE AND THEIR MICROBIAL HOSTS

AMOEBAE AS AN EVOLUTIONARY NICHE

The term free-living amoebae comprises more than 15,000 species (Adl et al., 2007) forming a heterogeneous group of phylogenetically distantly related protists that are widespread in water and soil ecosystems and display similar ecological characters. These unicellular eukaryotes were recently classified into two main suprakingdom-level groups; (i) the Excavata notably comprising the *Andalucia, Jakoba* (Jakobids), *Naegleria, Sawyeria, Vahlkampfia* (Heterolobosea) as well as the parasites *Trypanosoma, Leishmania* (Euglenozoa) and (ii) the Amoebozoa comprising among others *Acanthamoeba, Hartmannella, Vannella, Dictyostelium* and the medically important parasite *Entamoeba* (Hampl et al., 2009; Pawlowski and Burki, 2009).

Most amoebae live under the form of a trophozoite that replicates by binary fission, but in unfavorable conditions they can differentiate into a dormant form, the cyst. This latter is resistant to harsh conditions such as high temperature, desiccation, pH, and saline stress as well as disinfection processes (Thomas et al., 2004). Phagocytic amoebae graze on various microorganisms free-living or established in biofilms, including algae, bacteria, yeasts, and viruses (Rodriguez-Zaragoza, 1994). Microorganisms are phagocytosed and normally follow the endocytic pathway to be degraded in acidic phagolysosomes by a number of hydrolases (Greub and Raoult, 2004). However, several giant viruses and bacteria have evolved strategies to escape degradation, hence their naming as ARMs. They live symbiotically within their host or replicate in vacuoles before lysing the amoeba.

Taking profit from these characteristics, Rowbotham (1983) used cultures of amoebal cells to grow *Legionella* species. Since then, amoebal co-culture has become a method of choice to retrieve new microorganisms able to resist and grow in these professional phagocytes. This method uses amoebae as a cell background to inoculate environmental or medical samples, in order to retrieve ARMs (Lienard and Greub, 2011). However, the almost uniform use of *Acanthamoeba castellanii* (Thomas et al., 2006; Corsaro et al., 2009) and *A. polyphaga* (Greub et al., 2004b; Pagnier et al., 2008) largely biases and underestimates the diversity of known ARMs. Diversifying the species of amoeba used in co-culture experiments is required to improve our understanding of the pool of amoebal symbionts and parasites.

Concisely, amoebae can act as a replicative niche and a reservoir of ARMs that are established in water and soil environments (Greub and Raoult, 2004). As shown in **Figure 1**, amoebae may hide several ARMs in their cytoplasm or more commonly in phagocytic vacuoles. The cyst may function as an armor to protect internalized microorganism from difficult external conditions as well as disinfection procedures. Moreover, the development of strategies to resist microbicidal effectors by ARMs may help selecting virulence traits enabling to survive in the macrophages, the first line of human defense, as it is the case for *Mycobacteria*, *Legionella*, *Parachlamydia* (Greub, 2009; Lamoth and Greub, 2009; Salah et al., 2009), and Mimivirus (Ghigo et al., 2008). More importantly, amoebae were lately suggested as a place that favor genetic exchanges by bringing in close vicinity ARMs (Moliner et al., 2010; Thomas and Greub, 2010; Merhej et al., 2011). The recent sequencing of some amoebal genomes such as *Entamoeba histolytica, E. dispar, Dictyostelium discoideum,* and *A. castellanii* provide the opportunity to highlight the first hints on the genetic exchanges between the amoebae and their intracellular microbes.

AMOEBA-RESISTING VIRUSES

The search for amoeba-resisting viruses (ARVs), i.e., viruses able to replicate alone or in combination with others within amoeba, started with the discovery and the sequencing of Mimivirus in the early 2000s (La Scola et al., 2003; Raoult et al., 2004), which raised an extraordinary interest. Within less than a decade, the known complexity of ARVs was boosted by (i) the discovery of the small virophage Sputnik (La Scola et al., 2008), (ii) the description of Marseillevirus and Lausannevirus, two large viruses encoding histone-like proteins (Boyer et al., 2009; Thomas et al., 2011), as well as (iii) the publication of *Megavirus chilensis* (Fischer et al., 2010; Arslan et al., 2011), the largest identified virus harboring a 1.26 Mb genome (**Table 1**). According to new systematic searches, ARVs seem to be fairly common in the environment (La Scola et al., 2010).

Most ARVs possess pseudo-icosahedral capsids hiding intricate large dsDNA genomes that pushed forward the recognized limits of viral genome size. Therefore, they were called (i) "giant

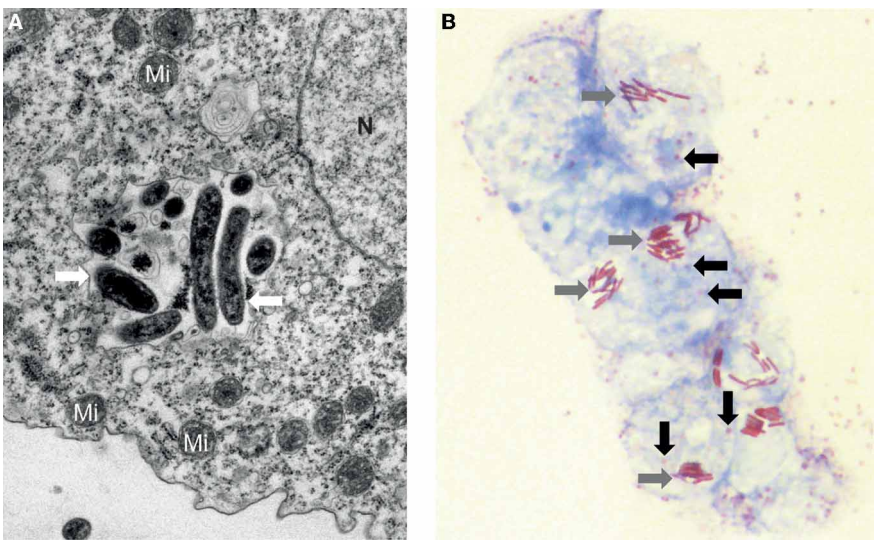


FIGURE 1 | Amoebae may hide several ARMs. (A) Electron microscopy of an inclusion containing a mixed population of microorganisms (white arrow) recovered from a water-humidifier co-cultured in an amoeba of the species *Acanthamoeba castellanii.* Mi: Mitochondrion, N: Nucleus. Magnification 10,000×. **(B)** Photonic microscopy of *Legionella spp.* and *Lausannevirus* in *A. castellanii.* Several amoebae contain simultaneously both the giant virus (black arrow) and the rod-shaped bacteria *Legionella* (grey arrow). Gimenez staining, Magnification 1000×.

Table 1 | Amoeba-resisting viruses with a publicly available genome sequence.

Family	Microorganism	Genome	Host	References
Mimiviridae	Mimivirus	1.18 Mb	*Acanthamoeba*	Raoult et al., 2004
Mimiviridae	Megavirus chilensis	1.26 Mb	*Acanthamoeba*	Arslan et al., 2011
Mimiviridae*	*Cafeteria roenbergensis* virus	0.730 Mb	*C. roenbergensis*	Fischer et al., 2010
–	Sputnik	0.018 Mb	*Acanthamoeba*	La Scola et al., 2008
Marseilleviridae	Marseillevirus	0.368 Mb	*Acanthamoeba*	Boyer et al., 2009
Marseilleviridae	Lausannevirus	0.346 Mb	*Acanthamoeba*	Thomas et al., 2011

*Classification proposed by Colson et al. (2011).

viruses" (La Scola et al., 2003; Raoult et al., 2004), a concept previously used for large algal DNA viruses (Van Etten and Meints, 1999), or (ii) "giruses" (Legendre et al., 2012). Another interesting girus is *Cafeteria roenbergensis* virus (CroV) that infects a marine phagocytic flagellate belonging to the *Chromalveolata* (Fischer et al., 2010). CroV was not demonstrated to replicate in amoebae, but it is worth discussing as it shares many similarities with other ARVs. At the opposite, Sputnik is a 18 kb virophage able to replicate in *A. castellanii* only when Mimivirus, or its close relative Mamavirus, co-infects the amoeba (La Scola et al., 2008). Sequences homologous to Sputnik were detected in the environmental dataset of the Global Ocean Survey, suggesting that it may represents a new virus family, but its classification is currently unclear.

Mimivirus, Megavirus, CroV, Marseillevirus, and Lausannevirus belong to the monophyletic class of nucleo-cytoplasmic large DNA virus (NCLDV) (Iyer et al., 2006). They were classified into two main families of NCLDV, the Mimiviridae and the Marseilleviridae (**Table 1**). NCLDVs only share a core genome of 30–47 genes (Iyer et al., 2006; Yutin et al., 2009). Thus, core genes only represent a minor fraction of the gene repertoire, whereas ORFans, i.e., genes that do not have homologs in other organisms, and dispensable genes, i.e., genes present in two or more NCLDVs, are the major constituent of these viral genomes.

The reconstruction of deep phylogenetic relationship from viral sequences is controversial because of the rapid evolutionary rate of viruses and the presence of numerous horizontal transfers (Moreira and Brochier-Armanet, 2008), in particular from host genomes. Nevertheless, several evidences suggest that these giruses have evolved from a common ancestor. The two families of distant giruses, *Marseilleviridae* and *Mimiviridae*, harbor an unusual genomic repertoire that includes genes for protein translation, a hallmark of cellular organisms (Raoult et al., 2004; Boyer et al., 2009; Arslan et al., 2011). Moreover, the presence of tRNA synthetases in some viruses and their absence in others supports the idea that all *Mimiviridae* evolved by reductive evolution from a common ancestor, potentially a cellular ancestor: four tRNA synthetases homologs have been found in Mimivirus and Megavirus, an additional one in both CroV and Megavirus, and two additional ones are present in Megavirus only (Ghigo et al., 2008). A further example of lineage-specific deletion is given by the DNA photolyase: CroV possesses two intact copies, Mimivirus harbors fragmented ortholog remnant of one of them, and finally Megavirus encodes one intact ortholog and one ortholog split in two parts by a transposase (Ghigo et al., 2008).

The amazing diversity in genome size and gene repertoire among these phylogenetically related viruses questions the respective importance of both LGTs and vertical inheritance in evolution. The large differences observed in genome size (0.018–1.3 Mb) would imply either an extensive genome growth via LGTs or a divergent reductive evolution in the different phyla. An increased propensity to acquire genes of foreign origin surely accounts for such differences in genome size (Monier et al., 2007) and some authors even consider viruses as "bags of genes" (Hendrix et al., 2000; Moreira and Lopez-Garcia, 2005). However, it is questionable whereas LGTs are sufficient to explain such a large variation. Moreover, core genes seem to have originated from different kingdom, including eukaryotes, bacteria, and bacteriophages (Koonin and Yutin, 2010). These observations are in agreement with the scenario of a bacteriophagic origin of NCLDV (Koonin and Yutin, 2010). At this stage, a mix of genes from very different eukaryotic and bacterial organisms were acquired concurrently to the loss of phage genes except those essential for genome replication and virion formation (Koonin and Yutin, 2010). The following section attempts to provide an overview of the extent of genetic exchanges documented in ARVs.

AMOEBA-RESISTING BACTERIA

A large variety of amoeba-resisting bacteria (ARBs) have been isolated using amoebal co-culture or directly retrieved from their host by amoebal enrichment (Greub et al., 2004b; Horn and Wagner, 2004; Lienard and Greub, 2011). In addition, many microorganisms have been shown to survive *in vitro* in amoebae such as different *Burkholderia, Coxiella burnetii,* a strain of *E. coli*, *Francisella tularensis, Helicobacter pylori, Listeria monocytogenes, Porphyromonas gingivalis,* and *Vibrio cholerae* (Greub and Raoult, 2004; Wagner et al., 2006). As shown in **Figure 2**, these encompass various clades scattered through the prokaryotic phylogeny, including members of the *Actinobacteria, Bacteroidetes, Chlamydiales, Firmicutes*, and different subdivisions of *Proteobacteria* (α, β, γ, ε). Although ARBs are found in most major taxonomic phyla, only few major groups of bacteria have been studied more extensively, including *Mycobacteria*, *Chlamydia*-related bacteria, *Rickettsia* and *Legionellae* and will be the focus of this review (**Table 2**).

Actinobacteria: The genus *Mycobacteria* comprises many bacteria such as *M. tuberculosis* and *M. leprae* that are major threat to human health and 19 different species have been

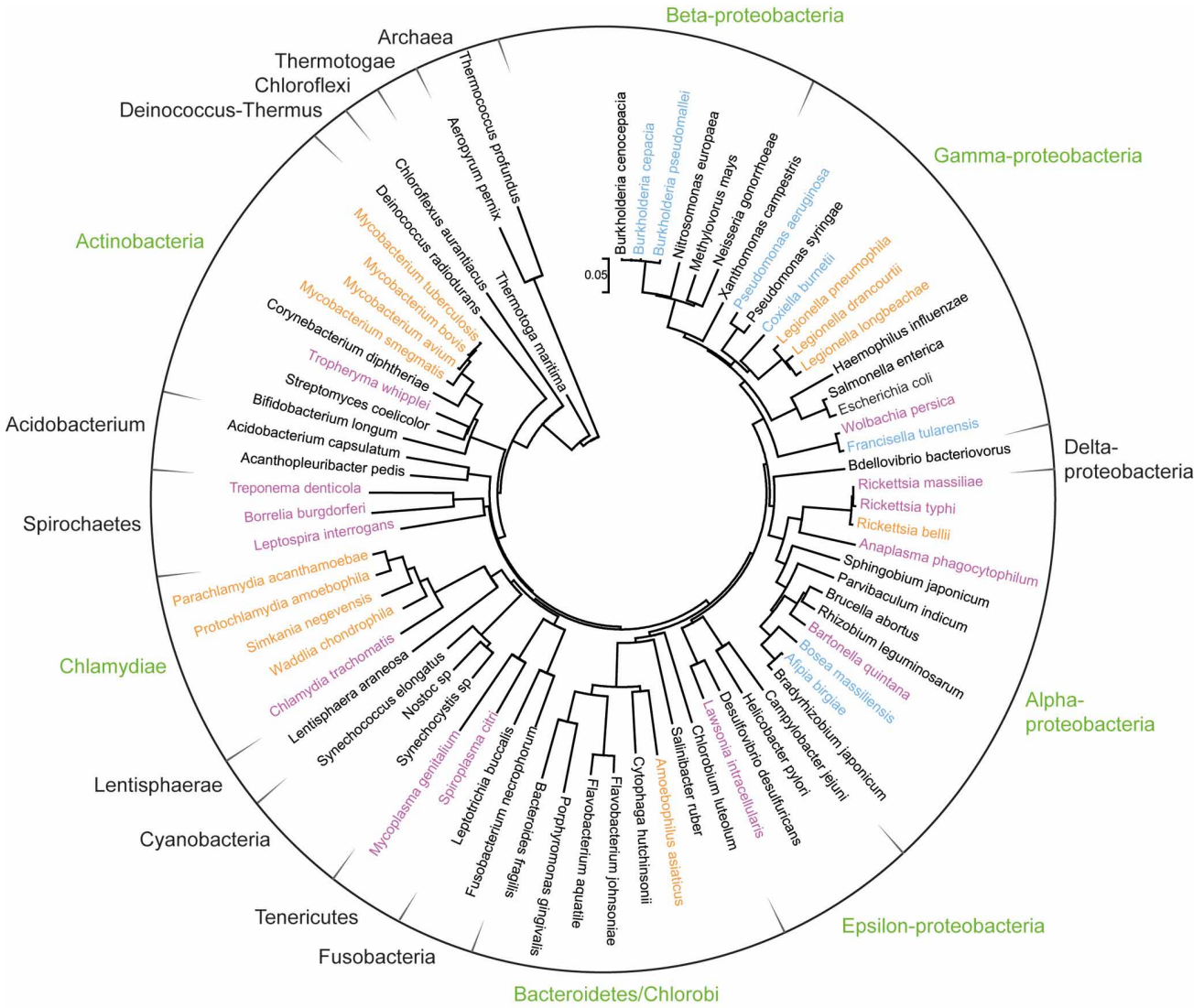


FIGURE 2 | Phylogenetic positioning of amoeba-resisting bacteria. The ribosomal RNA small subunit (16S) of Bacteria and two Archaea was aligned with Muscle and a neighbor-joining tree was subsequently built with Mega 5 (complete deletion, gamma distributed). The tree was rooted between Archaea and Bacteria. Bacterial phyla that comprise amoeba-resisting bacteria are indicated in green. Bacteria isolated from amoeba or growing within amoeba are highlighted in orange whereas those shown to resist amoebal phagocytosis *in vitro* are shown in blue. Finally other intracellular and fastidious bacteria are shown in purple.

sequenced to date. Most *Mycobacteria*, including members of the nontuberculous and the tuberculous complex groups, have been shown to survive and grow in various amoebae such as *Acanthamoeba sp., Dictyostelium discoideum,* and *Tetrahymena pyriformis* (Thomas and McDonnell, 2007; Mba Medie et al., 2011). It was suggested that adaptations for the intra-amoebal survival and multiplication may have facilitated the virulence toward mammalian cells (Molmeret et al., 2005).

Chlamydiales: The *Chlamydiales* order includes important human and animal pathogens such as *Chlamydia trachomatis*. A large number of amoebal symbionts called "*Chlamydia*-related bacteria" have been retrieved and shown to infect various organisms, including free-living amoebae, arthropods, insects as well as vertebrates (Corsaro and Greub, 2006; Horn, 2008). Four of them have been fully sequenced (**Table 2**), two additional strains have been published as draft genomes and new representative of the *Parachlamydiaceae* and *Criblamydiaceae* families are currently being sequenced (Horn et al., 2004; Greub et al., 2009; Bertelli et al., 2010; Collingro et al., 2011).

Alpha-proteobacteria: Highly pathogenic representatives of *Rickettsia* are typically transmitted through arthropods (Merhej and Raoult, 2011) but some *Rickettsia*-like endosymbionts were observed in *Acanthamoeba* (Fritsche et al., 1999). *Rickettsia bellii,* a species that diverged early in evolution, was shown to survive for three weeks in *A. polyphaga* (Ogata et al., 2006). *Odyssella thessalonicensis* is a strict intracellular bacteria isolated from an

Table 2 | Selected amoeba-resisting bacteria with a publicly available genome sequence.

Classification	Microorganism	Genome	Host	References
Actinobacteria	*Mycobacterium avium*	4.83 Mb	*Acanthamoeba*	Li et al., 2005
	Mycobacterium marinum	6.64 Mb	*Acanthamoeba*	Stinear et al., 2008
	Mycobacterium smegmatis	6.99 Mb	*Acanthamoeba*	Fleischmann et al., 2006
	Mycobacterium tuberculosis	4.41 Mb	*Acanthamoeba*	Cole et al., 1998
Bacteroidetes	*Amoebophilus asiaticus*	1.88 Mb	*Acanthamoeba*	Schmitz-Esser et al., 2010
	Porphyromonas gingivalis	2.34 Mb	*A. castellanii*[a]	Nelson et al., 2003
	Flavobacterium johnsoniae	6.10 Mb	*A. polyphaga*[a]	McBride et al., 2009
Chlamydiales	*Parachlamydia acanthamoebae*	3.07 Mb	*A. castellanii*	Collingro et al., 2011
	Protochlamydia amoebophila	2.41 Mb	*A. castellanii*	Horn et al., 2004
	Simkania negevensis	2.50 Mb	*A. castellanii*[a]	Collingro et al., 2011
	Waddlia chondrophila	2.12 Mb	*A. castellanii*	Bertelli et al., 2010
α-proteobacteria	*Rickettsia bellii*	1.52 Mb	*A. polyphaga*	Ogata et al., 2006
	Odyssella thessalonicensis	2.85 Mb[b]	*Acanthamoeba*	Georgiades et al., 2011
γ-proteobacteria	*Legionella drancourtii*	4.16 Mb[b]	*A. polyphaga*	Gimenez et al., 2011
	Legionella longbeachae	4.08 Mb	*Acanthamoeba*	Kozak et al., 2010
	Legionella pneumophila	3.50 Mb	*Acanthamoeba*	Cazalet et al., 2004; Chien et al., 2004

[a] *Shown to grow in amoebae, but maybe not the natural host.*

[b] *Unfinished genome sequence.*

air conditioning system in Greece (Doolittle and Bapteste, 2007). An unfinished genome sequence has been released, but unfortunately, little information is known on the genetic characteristics of this bacterium.

Gamma-proteobacteria: *Legionellae* are distributed worldwide and commonly found in water environments where they are major components of the biofilms (Rogers et al., 1994). Since the first culture of *L. pneumophila*, the causative agent of the Legionnaire's disease, within amoeba (Rowbotham, 1980), more than fifty strains of *Legionella* have been discovered. Although they are sometimes considered as all growing in amoebae, only a few have been shown to grow within amoebae (Rowbotham, 1980; Neumeister et al., 1997; La Scola et al., 2004) and the genomes of three species have been sequenced.

Geographic distribution and biology

ARBs are distributed worldwide and have been principally isolated from samples of aquatic environments where they are predominantly found in biofilms. The numerous and diverse bacteria entertain close relationships in biofilms where conjugation and transformation occur frequently enabling genetic exchanges (Molin and Tolker-Nielsen, 2003). The current explosion of new genome sequences significantly contributed to understand the mechanisms underlying genome evolution of the ARBs (Darby et al., 2007; Horn, 2008; Moran et al., 2008; Moya et al., 2008) and should enable us to conduct more extensive studies on genetic exchanges between ARMs, their biology and their interactions with the host cell. Indeed, ARBs show major differences in the mechanisms of host cell interaction. As an example, they have developed numerous ways to escape phagocytosis in amoebae: some escape the phagocytic pathway and replicate within the host cell cytoplasm (Birtles et al., 2000; Horn et al., 2001) whereas others block the maturation of phagolysosomes at various stages and replicate in host-derived vacuoles (Greub and Raoult, 2002; Isberg et al., 2009).

Genomic features

All the aforementioned ARBs share some similarities in their genomic characteristics. It has been shown that free-living bacteria tend to have larger genomes than the intracellular specialists that undergo genome reduction (Merhej et al., 2009). Although they are considered as specialized bacteria, ARBs harbor larger genomes compared to related organisms infecting human or other vertebrates (Moliner et al., 2010). For example, *Chlamydia*-related bacteria have genomes ranging from 2.1 to 3 Mb, i.e., twice two three times larger than classical *Chlamydiae*.

The genomes of *Rickettsia* and *Chlamydia*-related bacteria appear to be extensively shuffled by rearrangements, insertions and deletions and thus exhibit a very limited colinearity compared to related genomes (Ogata et al., 2006; Bertelli et al., 2010; Collingro et al., 2011). However, the signal may be dispersed by (i) the large difference in size that mimics numerous large insertions and (ii) the fact that these bacteria are more distantly related to one another than the closest relatives among them. To corroborate this hypothesis, the four strain of *L. pneumophila* fully sequenced to date are highly conserved with a single inversion of 250 kb taking place in strain Lens compared to the three other strains (Gomez-Valero et al., 2009).

As expected for strictly intracellular bacteria, they overall lack several global pathways or key enzymes for the synthesis of nucleotides, amino-acids, and cofactors (Chien et al., 2004; Horn et al., 2004; Bertelli et al., 2010; Schmitz-Esser et al., 2010; Collingro et al., 2011). In *Chlamydiales,* the pattern of missing pathways varies in each organism suggesting that these organisms have evolved by reductive evolution from an ancestor with enhanced if not complete biosynthetic capabilities. Interestingly, although *L. pneumophila* is known to be auxotrophic for several

amino acids such as cysteine, methionine, phenylalanine, and tyrosine, genes necessary for their biosynthesis were found (Chien et al., 2004). On the contrary, key features for the host pathogen interaction are highly conserved, as for example the complete type III secretion system of *Chlamydia* or the type IV Dot/Icm system in *Legionella* used to translocate effectors into the host cell (Bertelli et al., 2010; Moliner et al., 2010).

In summary, ARBs exhibit interesting genome characteristics compared to related microorganisms that do not infect amoebae: larger genomes, more genetic rearrangements, and more extensive metabolic capabilities. Globally, most ARBs harbor a largely diverse gene repertoire that indicates the occurrence of numerous lateral gene transfers (LGTs) detailed below.

GENETIC EXCHANGES

MECHANISMS OF GENE TRANSFER

Several mechanisms allow genetic exchanges among organisms from the various domains of life (Paz and Espinosa, 2010). Briefly, virus-mediated transfers occur via transduction from phages (prokaryotes) or transposon (prokaryotes and eukaryotes). Prokaryotes use transformation, i.e., DNA uptake from the environment, or conjugation, i.e., DNA exchange through pilus-like systems, to transfer gene fragments or plasmids. Prokaryote to eukaryote transfers may arise as the result of the ingestion of cells, a process called phagotrophism, or symbiogenesis, i.e., the establishment of a permanent association.

Such events were notably involved in the formation of the animal lineage through the primary symbiosis forming the primary eukaryotic cell. Some authors hypothesized that Archaebacteria and an Actinobacteria interacted in the primary symbiotic event leading to the creation of mitochondria, a hypothesis that was soon rejected (Cavalier-Smith, 2002). Others suggested that an archaebacteria was invaded by a second prokaryote related to α-*proteobacteria* that had a bacteriovory ability similar to *Bdellovibrio* (Davidov et al., 2006; Cox et al., 2008). Georgiades et al. (2011) proposed that mitochondria are more related to *Rickettsiales* and the bacterium *Pelibacter ubique*. Recently, it was also noticed that mitochondria may be sister to *Rhizobiales, Rhodobacterales,* or *Rickettsiales* suggesting an eventual chimeric origin of mitochondria (Atteia et al., 2009; Abhishek et al., 2011; Georgiades and Raoult, 2011). Similarly, the symbiosis of a cyanobacterium with this primary eukaryotic cell was at the basis of the establishment of the chloroplast 1.2 billion years ago (Dyall et al., 2004). However, some lateral transfer events or chimeric events may have occurred with an ancestral *Chlamydiales*, as suggested by the high proportion of plant-like genes in chlamydiae (Brinkman et al., 2002).

Lateral transfers occur with varying frequency, magnitude and resulting fitness, which modulates the establishment of the transferred domain in the population. LGTs are frequent in prokaryotes and transferred genes can become integrated rapidly in the population thanks to short generation times (Paz and Espinosa, 2010). On the contrary, although increasingly reported, gene transfers involving eukaryotic species are still currently underrepresented in the litterature (Keeling and Palmer, 2008; Andersson, 2009). This is in part due to the large number of available prokaryotic genome sequence facilitating such analyses, compared to the currently restricted number of published sequence of unicellular eukaryotes.

MOBILE GENETIC ELEMENTS

Mobile genetic elements (MGEs) are essential actors and markers of non vertical genome evolution (Frost et al., 2005). Indeed, they often encode mechanisms to spread in new host and form sites for the preferential acquisition of exogenic sequences. Insertions sequences (IS) and transposases, mobile endonucleases of the HNH family and DNA methylases that may form restriction/modification systems, as well as inteins are typical examples of MGEs commonly found in bacteria and archaea. As such, they were found in several, if not all, ARBs and we only provide a few examples below. Interestingly, they were also identified in several ARVs and we report them here in more details.

The genome sequence of *M. tuberculosis* harbor many IS ($n = 54$) that are preferentially integrated in intergenic regions, close to tRNAs (Cole et al., 1998), whereas *M. smegmatis* and *M. marinum* present fewer IS (Stinear et al., 2008). Similarly, *Chlamydia*-related bacteria all encode numerous entire or remnants of transposases, whereas their relatives *Chlamydia* have little if no trace of invasion by such mobile elements (Bertelli et al., 2010). Among ARBs, *A. asiaticus* is an interesting case as it exhibits a massive proliferation of MGEs (up to 24% of the genome), including abundant IS, but its genome seems relatively stable (Schmitz-Esser et al., 2010). An interesting example of composite transposon is found in *A. asiaticus* where the cluster of gene for lasso peptide synthesis is flanked by two IS, suggesting its ability to be mobilized for genetic transfer (Schmitz-Esser et al., 2010). The genomes of *Legionellae* exhibited a large plasticity indicated by the presence of numerous mobile elements and three plasmids in *L. pneumophila* (Chien et al., 2004; Gomez-Valero et al., 2009; Cazalet et al., 2010).

HNH and restriction-like endonucleases have been found in Marseillevirus ($n = 10$), Lausannevirus ($n = 8$), Megavirus ($n = 6$) and Mimivirus ($n = 3$) (Raoult et al., 2004; Boyer et al., 2009; Arslan et al., 2011; Thomas et al., 2011). In addition, a phage-type endonuclease discovered in Megavirus and two HNH flanking a prophage gene in Mimivirus might have been laterally acquired from prophage genomes.

Inteins are typical selfish elements that mediate protein splicing to trigger their excision from the protein precursor (Perler et al., 1994). They often contain a homing endonuclease (HE) that cuts double-stranded genomic DNA to integrate itself in the corresponding gene of non-infected organisms (Gogarten et al., 2002). Giruses are invaded by inteins of different types and at different locations (**Table 3**).

The eukaryotic-like DNA polymerase B of Mimivirus encodes an intein and its likely functional HE that is most closely related to extremophile archaea (Ogata et al., 2005a). Marseillevirus and Lausannevirus present two highly similar inteins and their HE in orthologous genes, suggesting that these MGEs were acquired by their ancestor and are now evolving differently toward degradation (Thomas et al., 2011). Megavirus and CroV also possess respectively two and four inteins that were until now not studied in detail. As it is often the case, all viral inteins are integrated in highly conserved genes related to replication, transcription, or

Table 3 | Inteins and their homing endonucleases in ARVs.

Microorganism	Infected gene	Intein insertion site	Homing endonuclease*
Mimivirus	DNA polymerase B	Tli Pol-2	Complete
Megavirus	DNA polymerase B	Tli Pol	n.a.
Megavirus	DNA-directed RNA polymerase beta subunit	RPB2	n.a.
CroV	DNA polymerase B	Tli Pol-2	Remnant
CroV	Ribonucleoside-diphosphate reductase, alpha subunit	RIR1-h	Remnant
CroV	DNA-directed RNA polymerase beta subunit	RPB2-c	None
CroV	DNA Topoisomerase IIA	Top2-a	Complete
Lausannevirus	D6/D11-like helicase	-	Complete
Lausannevirus	Ribonucleoside- diphosphate reductase, alpha subunit	RIR1	Remnant
Marseillevirus	D6/D11-like helicase	-	Complete
Marseillevirus	Ribonucleoside-diphosphate reductase, alpha subunit	RIR1	Remnant

** Remnant is indicated if one or several conserved blocks (C-D-E-H) of the homing endonuclease is missing.*

"n.a." Information not available.

DNA metabolism. These conserved genes constitutes preferential target as they are essential to the virus. These observations are in agreement with the hypothesis that viruses might play a central role in the transmission of inteins across species (Pietrokovski, 1998) and suggest that these mobile elements had or still have the ability to excise, spread an insert into new hosts. However, since the various ARVs present inteins in orthologous genes, further studies are required to investigate if these MGEs have spread in giruses sharing a sympatric lifestyle or if they were acquired by common ancestors.

The role and the exact origin of these MGEs in amoeba-infecting viruses are in most cases unclear. MGEs are often colocalized with genes of putative bacterial origin (Filee and Chandler, 2010). Like in bacteria, MGEs could thus be involved in promoting or facilitating lateral transfers, which may in turn provide some selective advantage to the ARVs by facilitating the acquisition of new advantageous genes.

EXTENT OF GENETIC EXCHANGES IN ARVs

The role of genetic exchanges has been a matter of intense debate since the publication of the Mimivirus (Raoult et al., 2004; Moreira and Lopez-Garcia, 2005; Ogata et al., 2005b). The first analyses suggested massive LGT from various origins, including members of the four domains of life. It was suggested that the genomes of giant viruses infecting protists are largely affected by LGTs and non-orthologous gene displacements (Filee et al., 2007; Fischer et al., 2010; Koonin and Yutin, 2010). Some authors even regarded these viruses as "bags of genes" thus suggesting that the amount of genes transferred from amoebae to viruses overweight the flux of genes in the opposite direction (Moreira and Lopez-Garcia, 2005; Moreira and Brochier-Armanet, 2008). It was shown that genes with an anomalous nucleotide composition may lead to the overestimation of LGTs (Monier et al., 2007), and family specific genes do not present an accelerated evolutionary rate and were not laterally exchanged (Ogata and Claverie, 2007). Several studies thus used BLAST-based searches and phylogenetic methods to investigate the occurrence of lateral transfers from host to virus, virus to host as well as between viruses and other microorganisms (**Table 4**). Unfortunately, the use of different cutoffs renders some results difficult to compare.

Lateral gene transfers with eukaryotes

The increasing availability of genome sequences from viral host enabled to investigate the occurrence of LGTs between NCLDVs and eukaryotes, and more specifically their host, using BLAST-based searches (Filee et al., 2008). Mimivirus had the lowest proportion of genes (3%, $n = 30$) from potential eukaryotic origin. A preliminary analysis of Megavirus, the only other sequenced representative of the Mimiviridae, showed that only 17% of the 258 genes of Megavirus with no obvious homolog in Mimivirus match against the nr database (Arslan et al., 2011). However, the authors report no special affinity with potential donors and do not report any information on the taxonomic classification these donors. By contrast, Marseillevirus showed the highest propensity ($n = 24$, 5.6%) to acquire genes from its host (Boyer et al., 2010; Filee and Chandler, 2010). These results challenge the theory of the viruses as "bags of genes" that would have gained genomic content by acquiring genes from diverse sources since the divergence of the last common NCLDV ancestor. In this case, we would expect larger genomes to have acquired more genes by LGTs, which is not observed here. On another hand, some virus families may have gained over time a higher tendency to acquire genes, which would bias the estimation of LGT propensity.

Potential gene transfers were not extensively studied for Lausannevirus, but its ubiquitin encoding gene was shown to present a best hit against *A. castellanii,* suggesting a transfer between the virus and its host (Thomas et al., 2011). By comparing Mimivirus and *Entamoeba histolytica*, 5 genes (5%) were potentially acquired from its amoebal host, providing that the *Entamoeba* genome is representative of the natural host, *A. castellanii*, whose genome was not available at that time (Andersson, 2009). Finally, Colson et al. (Colson et al., 2011) reported that *A. castellanii* encodes a homolog to the major capsid of CroV. They hypothesized that *A. castellanii* might represent an ancient host for CroV itself or for its ancestor which then specialized to infect flagellate protists such as *Cafeteria roenbergensis.*

Table 4 | Events of lateral gene transfer with amoeba-resisting viruses.

Microorganism	Viruses[a]		Host		Eukaryota		Bacteria		Archaea	
	B	P	B	P	B	P	B	P	B	P
Mimivirus	45	4	5	12–13	30	60	96	29	n.a.	1
Marseillevirus	59	51	n.a.	25	70	85	57	49	2	n.a.
Lausannevirus	2	n.a.	1	n.a.	2	n.a.	7	n.a.	-	n.a.

The table present the number of genes potentially laterally transferred as retrieved by BLAST-based (B) and by phylogeny-based (P) methods.

[a] *In Megavirus, 44 genes were reported to match against non-viral sequences in the nr database, but no information on taxonomic classification is available.*

In Mimivirus, phylogenetic analyses confirmed the limited number of LGTs with eukaryotes ($n = 60$, 6%) and with its host (n is unkown, ∼10% of 126 ORFs with eukaryotic homologs) (Moreira and Brochier-Armanet, 2008). However, these numbers do not take into account genes shared only by the virus and its amoebal host, rending impossible the phylogenetic reconstruction of an evolutionary history. In addition, Moreira et al. reported a few phylogenies supporting a transfer with the amoebae *Naegleria* and *Sawyeria*. Interestingly, large virus-like organisms that might be related to Mimiviridae were described in the cytoplasm of such amoebae (Schuster and Dunnebacke, 1974).

The histone-like proteins of Marseillevirus and Lausannevirus represent another intriguing case of potential lateral transfer. Heliothis zea virus, Bracovirus, and the ostreid herpesvirus were already shown to encode histone-like proteins probably acquired from their hosts (Cheng et al., 2002; Gad and Kim, 2008; De Souza et al., 2010). However, both Marseilleviridae harbor orthologs of as many as three histone-like proteins, two of them encoding for histone doublets, the third one being the fusion of a histone fold with an unknown domain (Thomas et al., 2011). A phylogenetic reconstruction clustered the various histone domains with different eukaryotic and archaeal histones, thus raising questions on their origin and their potential acquisition in a single or multiple events of lateral transfer. The function of these histone-like proteins has not been clearly shown yet. These histones were detected in the viral particle where they may help packaging DNA (Boyer et al., 2009), but they may as well have a role in modifying the chromatin structure of the host genome, a hypothesis that remains to be tested.

Lateral gene transfers with prokaryotes

Only few genes show some evidence of lateral transfer with archaea. A phylogenetic study suggested the archaeal origin of the DNA-directed RNA polymerase of Mimivirus (Moreira and Brochier-Armanet, 2008). In addition, two genes of Marseillevirus exhibited best BLAST hits against archaeal genes but this potential relationship were not further validated by phylogenetic analyses (Boyer et al., 2009). Thus, giruses and archaea do not seem to undergo significant lateral exchanges, probably, to some extent, because they share a less sympatric lifestyle. However, these results might be biased by the current paucity of archaeal genomes in sequence databases.

ARVs show more extensive potential LGTs with bacteria. BLAST-based methods identified respectively 96 genes (10%) in Mimivirus, 57 genes (13%) in Marseillevirus, and 7 additional genes in Lausannevirus that may have been exchanged with bacteria (Filee et al., 2008; Boyer et al., 2009; Thomas et al., 2011). Moreover, phylogenetic reconstructions confirmed 29 cases of gene transfer in Mimivirus and 49 in Marseillevirus (Moreira and Brochier-Armanet, 2008; Boyer et al., 2009). The analysis of LGTs in all NCLDVs showed that bigger genomes have a higher propensity to acquire genes from bacteria (La Scola et al., 2010). Interestingly, viruses infecting host that do not graze on bacteria exhibit less genes of potential bacterial origin (La Scola et al., 2010), suggesting that host grazing on microorganisms provide a favorable niche for genetic exchanges.

A few cases of lateral transfers between amoeba-resisting viruses and bacteria were documented. Moreira et al. (Moreira and Brochier-Armanet, 2008) reported the clustering of two Mimivirus genes to *Legionella pneumophila* and *Campylobacter spp.*, two bacterial species able to infect amoebae. Moreover, the dUTPase of Lausannevirus exhibited highest similarity to *Candidatus* Amoebophilus asiaticus, a symbiont of *Acanthamoeba* (Thomas et al., 2011). However, phylogenetic reconstruction did not confirm the clustering of both microorganisms. This may be due to the short length of the protein leaving only few phylogenetically informational sites. Finally, no phylogeny showed a cluster of Marseillevirus and ARBs such as *Legionella* or *Parachlamydia* (Boyer et al., 2009).

Viral genome extremities are hotspots for gene exchange

Genes potentially acquired by LGT from bacteria were shown to cluster at the extremities of Mimivirus linear genome, whereas they are more scattered throughout the genomes of other NLCDV (Filee et al., 2007, 2008). On the contrary, NCLDV core genes and genes of eukaryotic origin are centered on the genome sequence. Interestingly, Megavirus and Mimivirus are largely collinear in the central genomic region and exhibit a single inversion and a translocation (Arslan et al., 2011). In *Mimiviridae*, this is not correlated to a decrease in sequence conservation or to an enrichment in transposases in these regions. Similarly, Marseillevirus, and Lausannevirus show only a few inversions between 150 and 350 kb (Thomas et al., 2011). In contrast, the genomic extremities show an almost total loss of colinearity. A similar feature was observed in poxviruses (Esteban and Hutchinson, 2011), and, as suggested by Arslan et al. (Arslan et al., 2011) this might reflect some similarities in the system of genome replication, and an eventual coupling of replication and recombination that would favor the rearrangements, insertion, or deletion of genes at the extremities of the viral chromosomes.

EXTENT OF GENETIC EXCHANGES IN ARBs

Intracellular bacteria are thought to evolve by genome reduction rather than by acquisition of new genes (Moran, 2002). Individual genome publications as well as more detailed phylogenetic studies explored the events of LGT between bacteria, between bacteria and eukaryotes as well as with other microorganisms. Again, the various methods and cutoffs used make it difficult to directly compare the extent of genetic exchanges in each bacterial phyla. However, they provide an essential knowledge to appreciate the importance of such transfer events and the large diversity in the potential couples of donor-acceptor as shown by some well studied examples in **Table 5**.

Lateral gene transfers with eukaryotes

In *A. asiaticus*, five genes exhibiting typical eukaryotic domains were identified as laterally transferred with eukaryotes (Schmitz-Esser et al., 2010). Two corresponding phylogenies clustered *Amoeobophilus* with the amoeba *D. discoideum*. However, the precise identity of the donor remains unknown due to the limited availability of eukaryotic homologs and especially *Acanthamoeba*, the natural host of this bacterium.

Little evidence was found for recent LGTs in classical *Chlamydia* (Dalevi et al., 2002) and in *Pr. amoebophila* (Horn et al., 2004). Following the analysis of the first *Chlamydiales* genomes (Stephens et al., 1998; Horn et al., 2004), several studies identified plant genes of chlamydial origin and reported the importance of chlamydial genes in the establishment of plant plastid functions (Brinkman et al., 2002; Huang and Gogarten, 2007; Moustafa et al., 2008; Suzuki and Miyagishima, 2010). A recent study including newly sequenced *Chlamydia*-related bacteria demonstrated that 53 genes were transferred from *Chlamydiales* to plants ($n = 31$), to a subgroup of plants ($n = 7$) or in an unknown direction ($n = 9$) (Collingro et al., 2011). These genes encode a variety of functions listed by decreasing importance: carbohydrate metabolism, energy production, lipid metabolism, and translation. The central metabolic functions encoded by these genes support an essential contribution of *Chlamydiales* to plant genomes.

L. pneumophila and *L. longbeachae* harbor respectively 30 and 70 proteins with highest similarity to eukaryotic proteins (Cazalet et al., 2004; Kozak et al., 2010). Thomas et al. (Thomas and Greub, 2010) reported that out of 30 eukaryotic-like proteins of *L. pneumophila,* 8 were phylogenetically related to proteins encoded in ongoing amoebal genomes and expressed sequence tags. Interestingly, *L. pneumophila* likely acquired from a protist LegS2, a homolog to the sphingosine-1-phosphate lyase (SPL) that is highly conserved in eukaryotes (Degtyar et al., 2009). This proteins harbor an extra C-terminal domain, absent from its eukaryotic homologs, that is used to trigger its translocation into the host cells using the type IV Icm/Dot secretion system and to target it to the mitochondria. This demonstrates the ability of *Legionella* to alter proteins of eukaryotic origin to better use its host. In addition, Degtyar et al. (2009) denoted that the pattern of presence/absence of effector-encoding genes does not correlate with the *Legionella* phylogenetic tree of the genus, suggesting that these genes were acquired through a massive lateral transfer and lost during evolution.

In a recent study on the occurrence LGTs between *L. drancourtii* and *P. acanthamoebae*, we showed that three proteins of *L. drancourtii* clustered with eukaryotes in phylogenies: a keto acid dehydrogenase, a hypothetical protein and the 7-dehydrocholesterol reductase (Gimenez et al., 2011). This latter was previously shown to be present in *C. burnettii,* two *Chlamydia*-related bacteria (*P. amoebophila* and *P. acanthamoebae*) and Mimivirus but absent from other *Legionellae* (Moliner et al., 2009; Thomas and Greub, 2010). Moliner et al. suggested that the 7-dehydrocholesterol reductase had been acquired by a chlamydial ancestor from viridiplantae and further transferred to other intracellular bacteria. Thank to the availability of sequences from the amoeba *Naegleria gruberi* that clustered with ARBs, Thomas and Greub (Thomas and Greub, 2010) proposed that this gene had been directly exchanged between an amoeba and intracellular bacteria.

Another example of multiple lateral exchanges is the ADP/ATP translocase that is present only in some intracellular bacteria, green plants, and algae plastids (Winkler, 1976; Greub and Raoult, 2003). The first study considered that these genes were transferred from plants to *Rickettsiae* and *Chlamydiae* (Wolf et al., 1999). Subsequently, Amiri et al. (2003) suggested that these genes were of rickettsial origin. However, detailed phylogenetic analyses suggested an early gene duplication in *Chlamydiae*, an exchange between *Chlamydiae* and *Rickettsiae,* and a transfer from *Chlamydiae* to plants (Greub and Raoult, 2003; Linka et al., 2003; Schmitz-Esser et al., 2004). Based on 16S rRNA divergence

Table 5 | Examples of lateral gene transfer with amoeba-resisting bacteria.

Gene	Function	Partners of LGT	References
legS2	Sphingosine-1-phosphate lyase	*L. pneumoniae*-Protist	Degtyar et al., 2009
dhcR7, dwf	7-dehydrocholesterol reductase	*Legionellae*-Amoeba	Moliner et al., 2009; Thomas and Greub, 2010
tlc, ntt	ADP/ATP translocase, Nucleotides transporter	*Chlamydiales-Rickettsiales*-Plants	Greub and Raoult, 2003; Linka et al., 2003; Schmitz-Esser et al., 2004
ralF	Sec7 domain-containing protein	Eukaryota-*Legionella-Rickettsia*	Cox et al., 2004
tra	Type IV secretion system	Unknown-*Rickettsia Proteobacteria-Chlamydiales*	Gillespie et al., 2010; Greub et al., 2004a

and fossile estimates, this transfer from *Chlamydiae* to plants was dated to 1 billion years ago (Greub and Raoult, 2003). Interestingly, in *A. asiaticus* the single ADP/ATP translocase is flanked by two nearly identical IS, suggesting it has been acquired by transposon mediated transfer.

Lateral gene transfers with prokaryotes

The analysis of *A. asiaticus* genome revealed that 54 genes showed bidirectional best BLAST hits with organisms outside the phylum *Bacteroidetes* (Schmitz-Esser et al., 2010). Among them, 37 had a stable and well supported position in phylogenetic trees indicating they have been acquired laterally. Those represent ancient transfer events whose direction cannot be unambiguously determined. A large marjority of those genes are shared with other amoeba-associated bacteria, and in particular with *Rickettsiae*.

Only few examples of LGTs were reported in the *Chlamydiales*. Based on the presence of conserved indels, three genes were proposed to be exchanged with *Archaea* (*glmU*) and *Actinobacteria* (*murA* and *glyA*), respectively (Griffiths and Gupta, 2002, 2006). Moreover, sets of proteins unique to different chlamydial family or members were investigated by BLAST leading to the discovery of 33 cases of putative gene loss and transfer (Griffiths et al., 2006). In an extensive study of the cross-talk between *P. acanthamoebae* and *L. drancourtii*, 7 genes were likely involved in a direct LGT event (Gimenez et al., 2011). Moreover, 18 tree topologies suggested a transfer from *P. acanthamoebae* to an ancestor of the *Legionellae*. In addition, 4 topologies clustered various members of the *Chlamydiales* and the *Legionellales* indicating probable ancient exchanges between ancestors of these two otherwise distantly related clades.

Coscolla et al. (Coscolla et al., 2011) extensively studied the importance of LGT with other bacteria in the constitution of the *L. pneumophila* pangenome. A significant proportion (18%, $n = 704$) of the 3846 genes forming the pangenome were likely transferred with the following bacterial phyla in decreasing order of importance: β-*proteobacteria* ($n \approx 200$), α-*proteobacteria* ($n > 100$), *Actinobacteria* ($n \approx 100$), *Acidobacteria* and *Bacteroidetes*, followed by *Cyanobacteria*, *Firmicutes,* and *Chlamydiales*.

BLAST-based searches of *Rickettsia bellii* proteome highlighted respectively 72 and 22 proteins with a best hit among *Legionellae* and *Parachlamydiaceae* (Ogata et al., 2006). Similar searches with the proteomes of related α-proteobacteria suggested that *R. bellii* and *R. felis* were significantly enriched in sequences homologous to *Legionellae* and *Parachlamydiaceae* (8.8 and 8.2%, respectively) than other bacteria that do not belong to the *Rickettsiales* order (*Pelagibacter ubique*: 2.5%, *Mesorzizobium loti*: 0.9%, *Brucella melitensis, Caulobacter crescentus*). Further BLAST and phylogenetic analyses highlighted respectively 6 and 3 genes likely transferred laterally with *Legionellaceae* and *Parachlamydiaceae*, respectively. Among these is the Sec7 domain-containing protein that is homologous to RalF protein of *L. pneumophila*. This protein is secreted into the host cytosol where it helps recruiting ADP-ribosylation factors to the replicative vacuole (Nagai et al., 2002). Sec7 protein was suggested to be transferred from eukaryotes to bacteria and then in a secondary event between *Legionella* and *Rickettsia* (Cox et al., 2004).

A type four secretion system (T4SS) similar to an F-like conjugation system is encoded by *tra* genes in the genomes of *Rickettisa* (Ogata et al., 2005c, 2006). This system may translocate effectors into the host or mediate DNA transfer among bacteria (Christie, 2001; Ding et al., 2003). Its conservation among *Rickettsiales* and phylogenetic reconstructions suggests an ancestral acquisition of the rickettsial T4SS from organisms that do not belong to α-*proteobacteria* (Gillespie et al., 2010). A highly similar system was identified in several *Chlamydiales;* on a genomic island of *P. amoebophila* (Greub et al., 2004a), as a partial operon in *P. acanthamoebae* (Greub et al., 2009) as well as on the plasmid of *S. negevensis* (Collingro et al., 2011). First proposed to be of proteobacterial origin (Greub et al., 2004a), new genomic information suggested that the T4SS was acquired by an ancestor of the *Chlamydia*-related bacteria and subsequently lost in the *Waddliaceae* family (Collingro et al., 2011). This type IV secretion system likely contributes to the genome evolution of these intracellular pathogens by allowing the formation of conjugative pilus and the transfer of DNA from the donor to the recipient cell.

EXCHANGED PROTEIN FUNCTIONS AND ORIGIN OF TRANSFER

ARBs and ARVs both share the extremely sympatric lifestyle of ARMs. Amoebae represent a specific niche with particular requirements and strong selection pressure to enable the survival and growth of microorganisms. In this constrained environment, the acquisition and the establishment in the population of particular genes that provide advantages to the resisting organism in the host-pathogen interaction are thus naturally favored. The analysis of both ARBs and ARVs suggested a link between the function of laterally transferred genes and the potential origin of transfer. The functions may be categorized into (i) core functions for replication, transcription, and translation, (ii) metabolic pathways, (iii) mobile elements and systems for DNA conjugation or effector translocation, and (iv) eukaryotic domain of unclear function that may help in interacting with host cell factors.

As microorganisms extremely dependant on the host machinery, giruses have preferentially acquired by horizontal transfer genes belonging to the first and second category compared to bacteria. However, a few examples discussed above underlined their ability to acquire MGEs and to encode several proteins containing eukaryotic domains such as those with ankyrin repeats or leucine-rich repeats (LRRs), which fall within the third and fourth categories. Interestingly, it was denoted both in Marseillevirus and Mimivirus that genes involved in translation were more likely acquired from amoebae (**Table 6**) (Moreira and Brochier-Armanet, 2008; Boyer et al., 2009). For genes belonging to other aforementioned categories, the established origin of transfer differs in the two viruses (**Table 6**). In Marseillevirus, those involved in signal transduction were acquired from other eukaryotes whereas defense and repair functions, notably encoded by nucleases, were of bacterial or bacteriophage origin (Boyer et al., 2009). Finally, core metabolic functions, protein and lipid modification or degradation were from mixed bacterial and eukaryotic origin. In Mimivirus, genes for tRNA modification, protein folding and molecular chaperones, lipid metabolism, as well as amino acid metabolism were

Table 6 | Function and origin of LGTs in amoeba-resisting viruses.

Function	Mimivirus			Marseillevirus		
	E	B	P	E	B	P
Translation	x[a]			x[a]		
tRNA modification	x					
Repair					x	x
Defense					x	x
Signal transduction				x		
Polysaccharide metabolism		x		x	x	
Nucleotide metabolism		x		x	x	
Amino acid metabolism	x			x	x	
Protein modification and degradation	x			x	x	
Lipid metabolism	x			x	x	

E, Eukaryote; B, Bacteria; P, Bacteriophage.

[a] *More precisely, acquired from host.*

more likely exchanged with eukaryotes. On the contrary, genes involved in nucleotide or polysaccharide metabolism were of bacterial origin.

Globally, this suggests the role of LGTs to exchange genes conferring selective advantages to the viruses in diverting their host to their own advantage. The assessment of the potential transfer origin relies too much on the availability of certain types of viral, bacterial, and host genome sequences in public databases. The current incomplete picture of donor-acceptor classification hampers the drawing of further relationships.

All ARBs encodes numerous MGEs as well as conserved system for the translocation of effectors or DNA (Lawley et al., 2003; Juhas et al., 2008). Moreover, a large number of eukaryotic-like proteins and proteins harboring eukaryotic domains were discovered in the genomes of many intracellular bacteria like *C. burnetii* (Seshadri et al., 2003), *Legionellae* (Gomez-Valero et al., 2009; Kozak et al., 2010), *Rickettsia* (Ogata et al., 2005c, 2006), *A. asiaticus, M. avium,* and *F. tularensis* (Schmitz-Esser et al., 2010) suggesting their importance in interacting with the host cell. In *L. pneumophila, R. bellii,* and *R. felis,* they represent respectively 3.5, 6.8, and 4.9% of the gene content. These proteins present eukaryotic domains for host-cell interaction such as Ankyrin, TPR/Sel1, LRR, Serine/Threonine protein kinases as well as other protein–protein interaction domains such as F-box and U-box that may interfere with the host ubiquitin system. F-box proteins were discovered as components of the SCF complex that mediates the ubiquitination of proteins targeted for proteolysis and U-box proteins were identified as ubiquitin-protein ligases (Kipreos and Pagano, 2000; Hatakeyama and Nakayama, 2003).

Little is known on the origin and the evolution of the above-mentioned eukaryotic domains in eukaryotes and their parasites. Convergent evolution might explain some homologies, but several studies pointed the exogenic origin of these genes that have been likely acquired laterally from eukaryotes during evolution. Pioneering studies suggested important functions for some of these domains to modulate the host cell mechanisms and enable the efficient replication of amoeba-resisiting microorganisms. An example is the Serine/Threonine protein kinase of *M. tuberculosis* PknG that was shown to prevent phagosome-lysosome fusion and thus to promote the survival of the bacterium (Walburger et al., 2004). In *L. pneumophila*, Al-Quadan and Kwaik (2011) studied the role of an effector that contains an F-box and two Ankyrin repeats in macrophages and in *D. discoideum* amoeba. This effector is translocated through the Dot/Icm system and functions as a linker to dock polyubiquitinated proteins on the vacuole containing the *Legionella*.

CONCLUSIONS

Intracellular microorganisms have long been considered as specialists with limited genomic repertoires and few genes of exogenic origin. On the contrary, ARVs exhibit the largest genomes in the viral world (Raoult et al., 2004; Arslan et al., 2011). Given the importance of these viruses in the debate about the origin of life and the evolution of viruses, the number of known and sequenced ARVs will undoubtedly enlarge rapidly. Moreover, several ARBs of distant families harbor larger genomes than their closest relative infecting mammalian cells (Moliner et al., 2010). As genomic data on ARMs accumulate, a large panel of evidence substantiates the great activity of these microorganisms in transferring genes laterally. The current sequencing of new unicellular eukaryotes will highlight the genetic exchanges occurring in this setting. Broader and more standardized studies are now required to assess if amoebae represent a niche more favorable to lateral gene exchanges compared to other ecological systems and microbial communities such as biofilms or rhizosphere.

It is now clear that the genomes of amoeba-infecting microorganisms are of composite nature as they harbor genes related to all different kingdom of life. Giant viruses and ARBs had the possibility to exchange genes with eukaryotic organisms as well as with other intracellular microorganisms. The flux of genes in multiple directions enables eukaryote–virus–bacteria interactions. This explicitly indicates the role of amoebae as an evolutionary crib for the emergence of new microorganisms. However, data are still lacking to infer the exact prevalence of host to microorganism or microorganism to host transfers. Similarly, until now only few

studies evidenced probable direct LGTs between intra-amoebal microorganisms and the prevalence of ARMs to ARMs transfer is presumably underestimated.

It is only recently that amoebal co-culture has developed more broadly as a tool to isolate new ARMs. The difficulties to grow and isolate ARMs, as well as the propensity to use *Acanthamoeba* strains only for co-culture, largely bias our knowledge on the diversity of ARMs and the extent of genetic exchanges occurring within amoebae. The analysis of the genomes of new amoebae other than *Acanthamoeba* and their resisting microorganisms will enable to address if LGTs occurs at a similar rate in *Acanthamoeba* as in other amoebae. The proximity of ARMs in amoebae certainly helps maintaining a close relationship between their genomes through LGT. In a direct experiment, Saisongkorh et al. (2010) recently suggested that *Bartonella* and *Rhizobium radiobacter* can conjugate and exchange a plasmid when co-cultured in *A. polyphaga*. Similar experiments involving the co-culture of two or more microorganisms within the same amoeba over a varying number of generation followed by the isolation of the microorganisms and their resequencing will surely help evidencing such events of lateral transfer.

Genes laterally transferred belong to two categories; (i) those of known function, mostly involved in core processes that clearly improve the abilities of the microorganism to replicate and spread into new hosts and (ii) those of mostly unknown function such as proteins bearing particular interaction domains that may be used in the corruption of the cell machinery by the intracellular microorganism. Further investigations of laterally acquired genes and mechanistic studies on their function should enhance our knowledge on the mechanisms implicated in host–microbe interactions and the evolutionary history of pathogenesis. In this setting, the chimerism of ARMs may be more related to their lifestyle than to their phylogenetic and evolutionary history. The genome sequencing of new amoebae, and especially *A. castellanii*, will also likely reveal to be strongly influenced by LGTs, thus further challenging our current Darwinian perception of eukaryotic evolution.

REFERENCES

Abhishek, A., Bavishi, A., and Choudhary, M. (2011). Bacterial genome chimaerism and the origin of mitochondria. *Can. J. Microbiol.* 57, 49–61.

Adl, S. M., Leander, B. S., Simpson, A. G., Archibald, J. M., Anderson, O. R., Bass, D., Bowser, S. S., Brugerolle, G., Farmer, M. A., Karpov, S., Kolisko, M., Lane, C. E., Lodge, D. J., Mann, D. G., Meisterfeld, R., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Smirnov, A. V., and Spiegel, F. (2007). Diversity, nomenclature, and taxonomy of protists. *Syst. Biol.* 56, 684–689.

Al-Quadan, T., and Kwaik, Y. A. (2011). Molecular characterization of exploitation of the polyubiquitination and farnesylation machineries of dictyostelium discoideum by the AnkB F-Box effector of *Legionella pneumophila*. *Front. Microbiol.* 2:23. doi: 10.3389/fmicb.2011.00023

Amiri, H., Karlberg, O., and Andersson, S. G. (2003). Deep origin of plastid/parasite ATP/ADP translocases. *J. Mol. Evol.* 56, 137–150.

Andersson, J. O. (2005). Lateral gene transfer in eukaryotes. *Cell. Mol. Life Sci.* 62, 1182–1197.

Andersson, J. O. (2009). Horizontal gene transfer between microbial eukaryotes. *Methods Mol. Biol.* 532, 473–487.

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and Claverie, J. M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17486–17491.

Atteia, A., Adrait, A., Brugiere, S., Tardif, M., Van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., Martin, W., Garin, J., Joyard, J., and Rolland, N. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. *Mol. Biol. Evol.* 26, 1533–1548.

Bertelli, C., Collyn, F., Croxatto, A., Ruckert, C., Polkinghorne, A., Kebbi-Beghdadi, C., Goesmann, A., Vaughan, L., and Greub, G. (2010). The Waddlia genome: a window into chlamydial biology. *PLoS ONE* 5:e10890. doi: 10.1371/journal.pone.0010890

Birtles, R. J., Rowbotham, T. J., Michel, R., Pitcher, D. G., Lascola, B., Alexiou-Daniel, S., and Raoult, D. (2000). '*Candidatus Odyssella thessalonicensis*' gen. nov., sp. nov., an obligate intracellular parasite of Acanthamoeba species. *Int. J. Syst. Evol. Microbiol.* 50(Pt 1), 63–72.

Boyer, M., Gimenez, G., Suzan-Monti, M., and Raoult, D. (2010). Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. *Intervirology* 53, 310–320.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M. G., Suzan-Monti, M., La Scola, B., Koonin, E. V., and Raoult, D. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21848–21853.

Brinkman, F. S., Blanchard, J. L., Cherkasov, A., Av-Gay, Y., Brunham, R. C., Fernandez, R. C., Finlay, B. B., Otto, S. P., Ouellette, B. F., Keeling, P. J., Rose, A. M., Hancock, R. E., Jones, S. J., and Greberg, H. (2002). Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res.* 12, 1159–1167.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* 52, 297–354.

Cazalet, C., Gomez-Valero, L., Rusniok, C., Lomma, M., Dervins-Ravault, D., Newton, H. J., Sansom, F. M., Jarraud, S., Zidane, N., Ma, L., Bouchier, C., Etienne, J., Hartland, E. L., and Buchrieser, C. (2010). Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS Genet.* 6:e1000851. doi: 10.1371/journal.pgen.1000851

Cazalet, C., Rusniok, C., Bruggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F., Kunst, F., Etienne, J., Glaser, P., and Buchrieser, C. (2004). Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* 36, 1165–1173.

Cheng, C. H., Liu, S. M., Chow, T. Y., Hsiao, Y. Y., Wang, D. P., Huang, J. J., and Chen, H. H. (2002). Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the Baculoviridae. *J. Virol.* 76, 9024–9034.

Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S. M., Asamani, G., Hill, K., Nuara, J., Feder, M., Rineer, J., Greenberg, J. J., Steshenko, V., Park, S. H., Zhao, B., Teplitskaya, E., Edwards, J. R., Pampou, S., Georghiou, A., Chou, I. C., Iannuccilli, W., Ulz, M. E., Kim, D. H., Geringer-Sameth, A., Goldsberry, C., Morozov, P., Fischer, S. G., Segal, G., Qu, X., Rzhetsky, A., Zhang, P., Cayanis, E., De Jong, P. J., Ju, J., Kalachikov, S., Shuman, H. A., and Russo, J. J. (2004). The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305, 1966–1968.

Christie, P. J. (2001). Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol. Microbiol.* 40, 294–305.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. 3rd., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail,

M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R. C., Read, T. D., Bavoil, P. M., Sachse, K., Kahane, S., Friedman, M. G., Rattei, T., Myers, G. S., and Horn, M. (2011). Unity in variety–the pan-genome of the chlamydiae. *Mol. Biol. Evol.* 28, 3253–3270.

Colson, P., Gimenez, G., Boyer, M., Fournous, G., and Raoult, D. (2011). The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. *PLoS ONE* 6:e18935. doi: 10.1371/journal.pone.0018935

Corsaro, D., Feroldi, V., Saucedo, G., Ribas, F., Loret, J. F., and Greub, G. (2009). Novel Chlamydiales strains isolated from a water treatment plant. *Environ. Microbiol.* 11, 188–200.

Corsaro, D., and Greub, G. (2006). Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. *Clin. Microbiol. Rev.* 19, 283–297.

Coscolla, M., Comas, I., and Gonzalez-Candelas, F. (2011). Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol. Biol. Evol.* 28, 985–1001.

Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R., and Embley, T. M. (2008). The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20356–20361.

Cox, R., Mason-Gamer, R. J., Jackson, C. L., and Segev, N. (2004). Phylogenetic analysis of Sec7-domain-containing Arf nucleotide exchangers. *Mol. Biol. Cell* 15, 1487–1505.

Dalevi, D. A., Eriksen, N., Eriksson, K., and Andersson, S. G. (2002). Measuring genome divergence in bacteria: a case study using chlamydian data. *J. Mol. Evol.* 55, 24–36.

Danchin, E. G. J., and Rosso, M. N. (2012). Lateral gene transfers have polished animal genomes: lessons from nematodes. *Front. Cell. Inf. Microbio.* 2:27. doi: 10.3389/fcimb.2012.00027

Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J., and Andersson, S. G. (2007). Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* 23, 511–520.

Darwin, C. R. (ed.). (1859). *On the Origin of Species*. London: John Murray.

Davidov, Y., Huchon, D., Koval, S. F., and Jurkevitch, E. (2006). A new alpha-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. *Environ. Microbiol.* 8, 2179–2188.

Degtyar, E., Zusman, T., Ehrlich, M., and Segal, G. (2009). A Legionella effector acquired from protozoa is involved in sphingolipids metabolism and is targeted to the host cell mitochondria. *Cell. Microbiol.* 11, 1219–1235.

De Souza, R. F., Iyer, L. M., and Aravind, L. (2010). Diversity and evolution of chromatin proteins encoded by DNA viruses. *Biochim. Biophys. Acta* 1799, 302–318.

Ding, Z., Atmakuri, K., and Christie, P. J. (2003). The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol.* 11, 527–535.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* 284, 2124–2129.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Dyall, S. D., Brown, M. T., and Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. *Science* 304, 253–257.

Esteban, D. J., and Hutchinson, A. P. (2011). Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution. *BMC Genomics* 12, 261.

Filee, J., and Chandler, M. (2010). Gene exchange and the origin of giant viruses. *Intervirology* 53, 354–361.

Filee, J., Pouget, N., and Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *BMC Evol. Biol.* 8, 320.

Filee, J., Siguier, P., and Chandler, M. (2007). I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet.* 23, 10–15.

Fleischmann, R. D., Dodson, R. J., Haft, D. H., Merkel, J. S., Nelson, W. C., and Fraser, C. M. (2006). The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD, USA.

Fischer, M. G., Allen, M. J., Wilson, W. H., and Suttle, C. A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19508–19513.

Fritsche, T. R., Horn, M., Seyedirashti, S., Gautom, R. K., Schleifer, K. H., and Wagner, M. (1999). *In situ* detection of novel bacterial endosymbionts of Acanthamoeba spp. phylogenetically related to members of the order Rickettsiales. *Appl. Environ. Microbiol.* 65, 206–212.

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732.

Gad, W., and Kim, Y. (2008). A viral histone H4 encoded by *Cotesia plutellae* bracovirus inhibits haemocyte-spreading behaviour of the diamondback moth, *Plutella xylostella*. *J. Gen. Virol.* 89, 931–938.

Georgiades, K., Madoui, M. A., Le, P., Robert, C., and Raoult, D. (2011). Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *0* mitochondrion. *PLoS ONE* 6:e24857. doi: 10.1371/journal.pone.0024857

Georgiades, K., and Raoult, D. (2011). The rhizome of *Reclinomonas americana*, *Homo sapiens*, *Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. *Biol. Dir.* 6, 55.

Ghigo, E., Kartenbeck, J., Lien, P., Pelkmans, L., Capo, C., Mege, J. L., and Raoult, D. (2008). Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathog.* 4:e1000087. doi: 10.1371/journal.ppat.1000087

Gillespie, J. J., Brayton, K. A., Williams, K. P., Diaz, M. A., Brown, W. C., Azad, A. F., and Sobral, B. W. (2010). Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect. Immun.* 78, 1809–1823.

Gimenez, G., Bertelli, C., Moliner, C., Robert, C., Raoult, D., Fournier, P. E., and Greub, G. (2011). Insight into cross-talk between intra-amoebal pathogens. *BMC Genomics* 12, 542.

Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* 56, 263–287.

Gomez-Valero, L., Rusniok, C., and Buchrieser, C. (2009). *Legionella pneumophila*: population genetics, phylogeny and genomics. *Infect. Genet. Evol.* 9, 727–739.

Greub, G. (2009). *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. *Clin. Microbiol. Infect.* 15, 18–28.

Greub, G., Collyn, F., Guy, L., and Roten, C. A. (2004a). A genomic island present along the bacterial chromosome of the Parachlamydiaceae UWE25, an obligate amoebal endosymbiont, encodes a potentially functional F-like conjugative DNA transfer system. *BMC Microbiol.* 4.

Greub, G., La Scola, B., and Raoult, D. (2004b). Amoebae-resisting bacteria isolated from human nasal swabs by amoebal coculture. *Emerg. Infect. Dis.* 10, 470–477.

Greub, G., Kebbi-Beghdadi, C., Bertelli, C., Collyn, F., Riederer, B. M., Yersin, C., Croxatto, A., and Raoult, D. (2009). High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. *PLoS ONE* 4:e8423. doi: 10.1371/journal.pone.0008423

Greub, G., and Raoult, D. (2002). Crescent bodies of *Parachlamydia acanthamoeba* and its life cycle within *Acanthamoeba polyphaga*: an electron micrograph study. *Appl. Environ. Microbiol.* 68, 3076–3084.

Greub, G., and Raoult, D. (2003). History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a Chlamydiales ancestor to plants 1 billion years ago. *Appl. Environ. Microbiol.* 69, 5530–5535.

Greub, G., and Raoult, D. (2004). Microorganisms resistant to free-living amoebae. *Clin. Microbiol. Rev.* 17, 413–433.

Griffiths, E., and Gupta, R. S. (2002). Protein signatures distinctive of chlamydial species: horizontal transfers of cell wall biosynthesis genes glmU from archaea to chlamydiae and murA between chlamydiae and Streptomyces. *Microbiology* 148, 2541–2549.

Griffiths, E., and Gupta, R. S. (2006). Lateral transfers of serine hydroxymethyltransferase (glyA) and UDP-N-acetylglucosamine enolpyruvyl transferase (murA) genes from free-living Actinobacteria to the parasitic chlamydiae. *J. Mol. Evol.* 63, 283–296.

Griffiths, E., Ventresca, M. S., and Gupta, R. S. (2006). BLAST screening of chlamydial genomes to identify signature proteins that are unique for the Chlamydiales, Chlamydiaceae, Chlamydophila and Chlamydia groups of species. *BMC Genomics* 7, 14.

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G., and

Roger, A. J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc. Natl. Acad. Sci. U.S.A.* 106, 3859–3864.

Hatakeyama, S., and Nakayama, K. I. (2003). U-box proteins as a new family of ubiquitin ligases. *Biochem. Biophys. Res. Commun.* 302, 635–645.

Hendrix, R. W., Lawrence, J. G., Hatfull, G. F., and Casjens, S. (2000). The origins and ongoing evolution of viruses. *Trends Microbiol.* 8, 504–508.

Horn, M. (2008). Chlamydiae as Symbionts in Eukaryotes. *Annu. Rev. Microbiol.* 62, 113–131.

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C. L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G. J., Droege, M., Frishman, D., Rattei, T., Mewes, H. W., and Wagner, M. (2004). Illuminating the evolutionary history of chlamydiae. *Science* 304, 728–730.

Horn, M., Harzenetter, M. D., Linner, T., Schmid, E. N., Muller, K. D., Michel, R., and Wagner, M. (2001). Members of the Cytophaga-Flavobacterium-Bacteroides phylum as intracellular bacteria of acanthamoebae: proposal of '*Candidatus Amoebophilus asiaticus*'. *Environ. Microbiol.* 3, 440–449.

Horn, M., and Wagner, M. (2004). Bacterial endosymbionts of free-living amoebae. *J. Eukaryot. Microbiol.* 51, 509–514.

Huang, J., and Gogarten, J. P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8, R99.

Isberg, R. R., O'Connor, T. J., and Heidtman, M. (2009). The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nat. Rev. Microbiol.* 7, 13–24.

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184.

Juhas, M., Crook, D. W., and Hood, D. W. (2008). Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell. Microbiol.* 10, 2377–2386.

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618.

Kipreos, E. T., and Pagano, M. (2000). The F-box protein family. *Genome Biol.* 1, REVIEWS3002.

Koonin, E. V., and Yutin, N. (2010). Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* 53, 284–292.

Kozak, N. A., Buss, M., Lucas, C. E., Frace, M., Govil, D., Travis, T., Olsen-Rasmussen, M., Benson, R. F., and Fields, B. S. (2010). Virulence factors encoded by *Legionella longbeachae* identified on the basis of the genome sequence analysis of clinical isolate D-4968. *J. Bacteriol.* 192, 1030–1044.

La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J. M., and Raoult, D. (2003). A giant virus in amoebae. *Science* 299, 2033.

La Scola, B., Birtles, R. J., Greub, G., Harrison, T. J., Ratcliff, R. M., and Raoult, D. (2004). *Legionella drancourtii* sp. nov., a strictly intracellular amoebal pathogen. *Int. J. Syst. Evol. Microbiol.* 54, 699–703.

La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., and Raoult, D. (2010). Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53, 344–353.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., and Raoult, D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104.

Lamoth, F., and Greub, G. (2009). Amoebal pathogens as emerging causal agents of pneumonia. *FEMS Microbiol. Rev.* 34, 260–280.

Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15.

Legendre, M., Arslan, D., Abergel, C., and Claverie, J. M. (2012). Genomics of Megavirus and the elusive fourth domain of Life. *Commun. Integr. Biol.* 5, 102–106.

Li, L., Bannantine, J. P., Zhang, Q., Amonsin, A., May, B. J., Alt, D., Banerji, N., Kanjilal, S., and Kapur, V. (2005). The complete genome sequence of *Mycobacterium avium subspecies paratuberculosis. Proc. Natl. Acad. Sci. U.S.A.* 102, 12344–12349.

Lienard, J., and Greub, G. (2011). "Discovering new pathogens: amoebae as tools to isolate amoeba-resisting microorganisms from environmental samples," in *Environmental Microbiology: Current Technology and Water Applications*, eds K. Sen and N. J. Ashbolt (Norfolk, VA: Caister Academic Press), 143–162.

Linka, N., Hurka, H., Lang, B. F., Burger, G., Winkler, H. H., Stamme, C., Urbany, C., Seil, I., Kusch, J., and Neuhaus, H. E. (2003). Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* 306, 27–35.

Lopez, P., and Bapteste, E. (2009). Molecular phylogeny: reconstructing the forest. *C. R. Biol.* 332, 171–182.

Mba Medie, F., Ben Salah, I., Henrissat, B., Raoult, D., and Drancourt, M. (2011). *Mycobacterium tuberculosis* complex mycobacteria as amoeba-resistant organisms. *PLoS ONE* 6:e20499. doi: 10.1371/journal.pone.0020499

McBride, M. J., Xie, G., Martens, E. C., Lapidus, A., Henrissat, B., Rhodes, R. G., Goltsman, E., Wang, W., Xu, J., Hunnicutt, D. W., Staroscik, A. M., Hoover, T. R., Cheng, Y. Q., and Stein, J. L. (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl. Environ. Microbiol.* 75, 6864–6875.

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26.

Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences. *Mol. Biol. Evol.* 28, 3213–3223.

Merhej, V., and Raoult, D. (2011). Rickettsial evolution in the light of comparative genomics. *Biol. Rev. Camb. Philos. Soc.* 86, 379–405.

Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Dir.* 4, 13.

Molin, S., and Tolker-Nielsen, T. (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Curr. Opin. Biotechnol.* 14, 255–261.

Moliner, C., Fournier, P. E., and Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol. Rev.* 34, 281–294.

Moliner, C., Raoult, D., and Fournier, P. E. (2009). Evidence that the intra-amoebal *Legionella drancourtii* acquired a sterol reductase gene from eukaryotes. *BMC Res. Notes* 2, 51.

Molmeret, M., Horn, M., Wagner, M., Santic, M., and Abu Kwaik, Y. (2005). Amoebae as training grounds for intracellular bacterial pathogens. *Appl. Environ. Microbiol.* 71, 20–28.

Monier, A., Claverie, J. M., and Ogata, H. (2007). Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* 8, 456.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586.

Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190.

Moreira, D., and Brochier-Armanet, C. (2008). Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol. Biol.* 8, 12.

Moreira, D., and Lopez-Garcia, P. (2005). Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science* 308, 1114. Author reply 1114.

Moustafa, A., Reyes-Prieto, A., and Bhattacharya, D. (2008). Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* 3:e2205. doi: 10.1371/journal.pone.0002205

Moya, A., Pereto, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9, 218–229.

Nagai, H., Kagan, J. C., Zhu, X., Kahn, R. A., and Roy, C. R. (2002). A bacterial guanine nucleotide exchange factor activates ARF on *Legionella phagosomes. Science* 295, 679–682.

Nelson, K. E., Fleischmann, R. D., Deboy, R. T., Paulsen, I. T., Fouts, D. E., Eisen, J. A., Daugherty, S. C., Dodson, R. J., Durkin, A. S., Gwinn, M., Haft, D. H., Kolonay, J. F., Nelson, W. C., Mason, T., Tallon, L., Gray, J., Granger, D., Tettelin, H., Dong, H., Galvin, J. L., Duncan, M. J., Dewhirst, F. E., and Fraser, C. M. (2003). Complete genome sequence of the oral pathogenic Bacterium *porphyromonas gingivalis* strain W83. *J. Bacteriol.* 185, 5591–5601.

Neumeister, B., Schoniger, S., Faigle, M., Eichner, M., and Dietz, K. (1997). Multiplication of different Legionella species in Mono Mac 6 cells and in *Acanthamoeba castellanii. Appl. Environ. Microbiol.* 63, 1219–1224.

Ogata, H., and Claverie, J. M. (2007). Unique genes in giant viruses: regular substitution pattern and anomalously short size. *Genome Res.* 17, 1353–1361.

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P. E., Claverie, J. M., and Raoult, D. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* 2:e76. doi: 10.1371/journal.pgen.0020076

Ogata, H., Raoult, D., and Claverie, J. M. (2005a). A new example of viral intein in Mimivirus. *Virol. J.* 2, 8.

Ogata, H., Raoult, D., and Claverie, J. M. (2005b). Response to Comment on "The 1.2-Megabase Genome Sequence of Mimivirus". *Science* 308, 1114.

Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P. E., Parinello, H., Claverie, J. M., and Raoult, D. (2005c). The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* 3:e248. doi: 10.1371/journal.pbio.0030248

Olendzenski, L., and Gogarten, J. P. (2009). Evolution of genes and organisms: the tree/web of life in light of horizontal gene transfer. *Ann. N.Y. Acad. Sci.* 1178, 137–145.

Pagnier, I., Raoult, D., and La Scola, B. (2008). Isolation and identification of amoeba-resisting bacteria from water in human environment by using an *Acanthamoeba polyphaga* co-culture procedure. *Environ. Microbiol.* 10, 1135–1144.

Pawlowski, J., and Burki, F. (2009). Untangling the phylogeny of amoeboid protists. *J. Eukaryot. Microbiol.* 56, 16–25.

Paz, Y. M. C. G., and Espinosa, A. (2010). Integrating horizontal gene transfer and common descent to depict evolution and contrast it with "common design". *J. Eukaryot. Microbiol.* 57, 11–18.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994). Protein splicing elements: inteins and exteins–a definition of terms and recommended nomenclature. *Nucleic Acids Res.* 22, 1125–1127.

Pietrokovski, S. (1998). Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* 7, 64–71.

Raoult, D. (2010a). Giant viruses from amoeba in a post-Darwinist viral world. *Intervirology* 53, 251–253.

Raoult, D. (2010b). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.

Rodriguez-Zaragoza, S. (1994). Ecology of free-living amoebae. *Crit. Rev. Microbiol.* 20, 225–241.

Rogers, J., Dowsett, A. B., Dennis, P. J., Lee, J. V., and Keevil, C. W. (1994). Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl. Environ. Microbiol.* 60, 1585–1592.

Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J. Clin. Pathol.* 33, 1179–1183.

Rowbotham, T. J. (1983). Isolation of *Legionella pneumophila* from clinical specimens via amoebae, and the interaction of those and other isolates with amoebae. *J. Clin. Pathol.* 36, 978–986.

Saisongkorh, W., Robert, C., La Scola, B., Raoult, D., and Rolain, J. M. (2010). Evidence of transfer by conjugation of type IV secretion system genes between Bartonella species and *Rhizobium radiobacter* in amoeba. *PLoS ONE* 5:e12666. doi: 10.1371/journal.pone.0012666

Salah, I. B., Ghigo, E., and Drancourt, M. (2009). Free-living amoebae, a training field for macrophage resistance of mycobacteria. *Clin. Microbiol. Infect.* 15, 894–905.

Schmitz-Esser, S., Linka, N., Collingro, A., Beier, C. L., Neuhaus, H. E., Wagner, M., and Horn, M. (2004). ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to chlamydiae and rickettsiae. *J. Bacteriol.* 186, 683–691.

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and Horn, M. (2010). The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J. Bacteriol.* 192, 1045–1057.

Schuster, F. L., and Dunnebacke, T. H. (1974). Virus-like particles and an unassociated infectious agent in amoebae of the genus Naegleria. *Ann. Soc. Belg. Med. Trop.* 54, 359–370.

Seshadri, R., Paulsen, I. T., Eisen, J. A., Read, T. D., Nelson, K. E., Nelson, W. C., Ward, N. L., Tettelin, H., Davidsen, T. M., Beanan, M. J., Deboy, R. T., Daugherty, S. C., Brinkac, L. M., Madupu, R., Dodson, R. J., Khouri, H. M., Lee, K. H., Carty, H. A., Scanlan, D., Heinzen, R. A., Thompson, H. A., Samuel, J. E., Fraser, C. M., and Heidelberg, J. F. (2003). Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5455–5460.

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q. X., Koonin, E. V., and Davis, R. W. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282, 754–759.

Stinear, T. P., Seemann, T., Harrison, P. F., Jenkin, G. A., Davies, J. K., Johnson, P. D., Abdellah, Z., Arrowsmith, C., Chillingworth, T., Churcher, C., Clarke, K., Cronin, A., Davis, P., Goodhead, I., Holroyd, N., Jagels, K., Lord, A., Moule, S., Mungall, K., Norbertczak, H., Quail, M. A., Rabbinowitsch, E., Walker, D., White, B., Whitehead, S., Small, P. L., Brosch, R., Ramakrishnan, L., Fischbach, M. A., Parkhill, J., and Cole, S. T. (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* 18, 729–741.

Suzuki, K., and Miyagishima, S. Y. (2010). Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. *Mol. Biol. Evol.* 27, 581–590.

Thomas, V., Bertelli, C., Collyn, F., Casson, N., Telenti, A., Goesmann, A., Croxatto, A., and Greub, G. (2011). Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ. Microbiol.* 13, 1454–1466.

Thomas, V., Bouchez, T., Nicolas, V., Robert, S., Loret, J. F., and Levi, Y. (2004). Amoebae in domestic water systems: resistance to disinfection treatments and implication in Legionella persistence. *J. Appl. Microbiol.* 97, 950–963.

Thomas, V., and Greub, G. (2010). Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. *Intervirology* 53, 254–267.

Thomas, V., Herrera-Rimann, K., Blanc, D. S., and Greub, G. (2006). Biodiversity of amoebae and amoeba-resisting bacteria in a hospital water network. *Appl. Environ. Microbiol.* 72, 2428–2438.

Thomas, V., and McDonnell, G. (2007). Relationship between mycobacteria and amoebae: ecological and epidemiological concerns. *Lett. Appl. Microbiol.* 45, 349–357.

Van Etten, J. L., and Meints, R. H. (1999). Giant viruses infecting algae. *Annu. Rev. Microbiol.* 53, 447–494.

Wagner, Y., Noack, B., Hoffmann, T., Jacobs, E., and Christian Luck, P. (2006). Periodontopathogenic bacteria multiply in the environmental amoeba *Acanthamoeba castellani*. *Int. J. Hyg. Environ. Health* 209, 535–539.

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G., and Pieters, J. (2004). Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* 304, 1800–1804.

Winkler, H. H. (1976). Rickettsial permeability. An ADP-ATP transport system. *J. Biol. Chem.* 251, 389–396.

Wolf, Y. I., Aravind, L., and Koonin, E. V. (1999). Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* 15, 173–175.

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6, 223.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 May 2012; paper pending published: 22 May 2012; accepted: 01 August 2012; published online: 21 August 2012.

*Citation: Bertelli C and Greub G (2012) Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. Front. Cell. Inf. Microbio. **2**:110. doi: 10.3389/fcimb.2012.00110*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 04 May 2012
doi: 10.3389/fcimb.2012.00060

How microbiology helps define the rhizome of life

Kalliopi Georgiades and Didier Raoult*

Faculté de Médecine La Timone, Unité de Recherche en Maladies Infectieuses Tropical Emergentes (URMITE), CNRS-IRD UMR 6236-198, Université de la Méditerranée, Marseille, France

Edited by:
Eugene V. Koonin, National Institutes of Health, USA

Reviewed by:
Richard Hayward, University College London (UCL) and Birkbeck, University of London, UK
Eric Bapteste, Centre national de la recherche scientifique, France

***Correspondence:**
Didier Raoult, Faculté de Médecine La Timone, Unité de Recherche en Maladies Infectieuses Tropical Emergentes (URMITE), CNRS-IRD UMR 6236-198, Université de la Méditerranée, 27, Bd Jean Moulin, 13385, Marseille cedex 5, France.
e-mail: didier.raoult@gmail.com

In contrast to the tree of life (TOF) theory, species are mosaics of gene sequences with different origins. Observations of the extensive lateral sequence transfers in all organisms have demonstrated that the genomes of all life forms are collections of genes with different evolutionary histories that cannot be represented by a single TOF. Moreover, genes themselves commonly have several origins due to recombination. The human genome is not free from recombination events, so it is a mosaic like other organisms' genomes. Recent studies have demonstrated evidence for the integration of parasitic DNA into the human genome. Lateral transfer events have been accepted as major contributors of genome evolution in free-living bacteria. Furthermore, the accumulation of genomic sequence data provides evidence for extended genetic exchanges in intracellular bacteria and suggests that such events constitute an agent that promotes and maintains all bacterial species. Archaea and viruses also form chimeras containing primarily bacterial but also eukaryotic sequences. In addition to lateral transfers, orphan genes are indicative of the fact that gene creation is a permanent and unsettled phenomenon. Currently, a rhizome may more adequately represent the multiplicity and *de novo* creation of a genome. We wanted to confirm that the term "rhizome" in evolutionary biology applies to the entire cellular life history. This view of evolution should resemble a clump of roots representing the multiple origins of the repertoires of the genes of each species.

Keywords: rhizome, genealogic tree, horizontal sequence transfer, recombination, orphan genes

SPECIATION

The definition of a species established by Mayr (Mayr, 1957), known as the "Biological Species Concept," postulates that species are groups of interbreeding natural populations that are reproductively isolated from other such groups. Therefore, sympatric speciation can occur only between closely related species, in which gene exchange is possible, whereas reproductive isolation is a process that can only occur if geographical separation partially or completely impedes gene flow. This definition establishes allopatric speciation as the norm (Mayr, 1957). Ecological changes or natural disasters lead to organismal isolation and bottlenecks, and therefore, to speciation. To this end, the main characteristic of speciation in eukaryotes is geographic isolation or allopatry, which restricts the capacity to exchange genes (Via, 2009); this feature is also observed in specialized bacteria (Doolittle and Papke, 2006).

However, a species is not a stable entity because species are continuously created and transformed. Due to their previous sympatric lifestyle, in which genetic exchanges were not limited, genomes are mosaics of sequences with different origins (Georgiades et al., 2011a,b). Indeed, the observations of extensive lateral sequence transfers in all organisms demonstrated that the genomes of all life forms are collections of genes with different evolutionary histories that cannot be represented by a single tree of life (TOL) (Koonin, 2009).

Many alternatives have been proposed to represent all the gene exchange events in organisms, such as networks (Kunin et al., 2005; Dagan et al., 2008; Halary et al., 2010; Beauregard-Racine et al., 2011; Kloesges et al., 2011; Popa et al., 2011), forests (Lopez and Bapteste, 2009) and bushes (Gould, 1987).

We sought to confirm that the term "rhizome" in evolutionary biology is the most suitable descriptor and to demonstrate that the entire history of cellular life is a rhizome (Raoult, 2010a; Merhej et al., 2011; Koonin, 2012).

LATERAL INHERITANCE

Lateral gene transfers (LGT) have been considered to be marginal phenomena, important under specific circumstances, which could be discarded in the study of organismal evolution (Koonin and Wolf, 2008). However, recent extensive comparisons of multiple whole genome sequences have revealed a vast and surprising variability in gene content, even among closely related-species (Mira et al., 2001; Berg and Kurland, 2002; Konstantinidis and Tiedje, 2004; Koonin and Wolf, 2008; Georgiades et al., 2011a,b). Comparative genomics revealed that most genes are susceptible to LGT (Sorek et al., 2007), although the tendency to undergo LGT is highly variable among genes (Nakamura et al., 2004; Cohen et al., 2008, 2011; Hao and Golding, 2008). LGT affects different classes of genes to different extents, but no single gene is completely immune to LGT (Koonin and Wolf, 2008). LGT may be mediated by the inheritance of a plasmid, the integration of a lysogenic phage into a chromosome, or by the insertion of a linear fragment into a chromosome (Kaper and Hacker, 1999). Complete genome sequences

highlight the confounding effects of lateral transfers in reconstructing the history of organismal evolution (Koonin et al., 1997, 2000; Doolittle, 1999; Nelson et al., 1999; Boucher et al., 2001; Gogarten et al., 2002; Zhaxybayeva and Gogarten, 2002). A practical necessity for detecting LGT is a species tree that depicts the phylogeny of the compared organisms in their entirety. Indeed, the most common practice used to detect LGT involves identifying the reliable discrepancies between the topologies of a gene tree and a species tree (Merhej et al., 2011). The species tree concept was validated by comparing the phylogenetic trees of sets of several hundred single-copy clusters of orthologous groups (COGs) from well-characterized, widespread bacterial groups, such as alphaproteobacteria, gamma-proteobacteria or Bacillus-Clostridium (Koonin and Wolf, 2008). However, LGT is a misleading definition because lateral inheritance may refer to sequences and not to entire or single genes (Chan et al., 2009, 2011; Merhej et al., 2011).

LATERAL SEQUENCE TRANSFERS IN EUKARYOTES AND THE HUMAN RHIZOME

Two elements participate in eukaryotic genetic change. The first is sexuality. In 1889, August Weismann proposed that sex evolved because of the advantage in creating variation among siblings (Weismann, 1889). Therefore, natural selection favors the parents who can produce a variety of offspring. A similar hypothesis was proposed by Darwin in his "Origin of Species" (Darwin, 1859), which states that all sexually reproducing organisms are derived from a common, single-celled eukaryotic ancestor. Many protists reproduce sexually, as do multicellular plants, animals, and fungi. A few species, such as Bdelloidea and some parthenocarpic plants have secondarily lost this feature (Letunic and Bork, 2006). Some species, such as arthropods, can reproduce sexually and asexually or undergo parthenogenesis, which is the development of embryos in the absence of male fertilization. *Wolbachia* are known to induce female parthenogenesis in infected arthropods (Renvoisé et al., 2011).

The second mechanism of genetic exchange in eukaryotes is lateral sequence transfer by infection or recombination. Sequence transfers from bacteria to eukaryotes have occurred due to the ancestral symbiotic events that led to the establishment of plant and animal lineages (Thomas and Greub, 2010). Amoebae have also played a significant role as a melting pot for genetic exchange (Ogata et al., 2006; Moliner et al., 2010; Raoult and Boyer, 2010). A dramatic lateral sequence transfer has been reported: nearly the entire *Wolbachia* genome was observed to be integrated into the host genome. Comparative genomic studies support the existence of progressive sequence transfers from *Wolbachia* to arthropods, insects, and nematodes (Dunning Hotopp et al., 2007; Nikoh et al., 2008; McNulty et al., 2010). In particular, *Drosophila ananassae* harbors the entire genome of a *Wolbachia* species (Callaway, 2007).

The human genome does not escape such mixtures and also is a mosaic. Evidence from a recent study supports the integration of parasitic DNA into the human genome. The authors revealed that *Trypanosoma cruzi* sequences were integrated into the genomes of patients from five families from different Brazilian ecosystems. An investigation of the role of saliva in the transmission of human herpesvirus 6 (HHV6) revealed that all nine submandibular glands and one of four parotid glands harbored the HHV6 genome and expressed the corresponding proteins (Fox et al., 1990). The consequences of parasite DNA acquisition, the vertical inheritance of integrated DNA and its subsequent drift may contribute to the ongoing genetic diversity and speciation in the human population (Hecht et al., 2010), which is indicative of a chimera (Raoult, 2011).

Genome mixing has also occurred between archaic hominins and modern humans. Indeed, evidence for a notable presence of a Neanderthal-derived *X* chromosome segment among all contemporary human populations outside Africa has been presented (Yotova et al., 2011). Another archaic hominin was recently discovered in the Denisovan cave in Siberia. This group is derived from a hominin migration out of Africa distinct from the Neanderthal ancestors and modern humans, as suggested by the highly divergent morphological features. Evidence suggests that this archaic human lived close in time and space with Neanderthals and modern humans and that its genome contributed 4–6% of its genetic material to the genomes of the present-day Malaysians (Krause et al., 2010; Reich et al., 2010), whereas Neanderthals contributed approximately 1–4% of their genetic material to modern Europeans (Abi-Rached et al., 2011). Furthermore, a recent analysis of the highly polymorphic human leukocyte antigen (HLA) class I revealed how modern humans acquired the HLA-B*73 allele in West Asia through mixing with Denisovans (Abi-Rached et al., 2011). All these surprising data suggest that the genetic constitution of modern humans may be the outcome of a mosaic of lineages from different times and geographic origins (Yotova et al., 2011), as defined by our vision of the rhizome (Raoult, 2011) (**Figure 1**).

THE MITOCHONDRIAL RHIZOME

Mitochondrial evolution has recently been demonstrated to constitute a rhizome (Georgiades and Raoult, 2011). The results of this study suggest that the origins of mitochondrial genes are not limited to *Rickettsiales* and that their creation did not occur in a single event but through multiple successive events. Contrary to what has been believed until now, recent evidence strongly suggests that mitochondria do not have a single common ancestor (Emelyanov, 2001) but likely numerous ancestors, including proto-*Rickettsiales*, proto-*Rhizobiales*, proto-alphaproteobacteria and current alphaproteobacterial species. The mosaicism of the mitochondrial genome is also discussed by Esser et al. (2007), however, in our study, the use of four different types of mitochondria from four different organisms (protozoa, yeast, louse, human) revealed that the mitochondria of different organisms comprise different elements, while the analysis of the louse multichromosomal mitochondrion (Shao et al., 2009) showed that the mitochondria creation model is not fixed: mitochondria do not have a stable or unique form, and thus their evolution cannot be the same. We conclude from these results that the TOL is not sufficient to explain the chimeric structure of mitochondrial genomes and that their evolution should be represented as a rhizome as well.

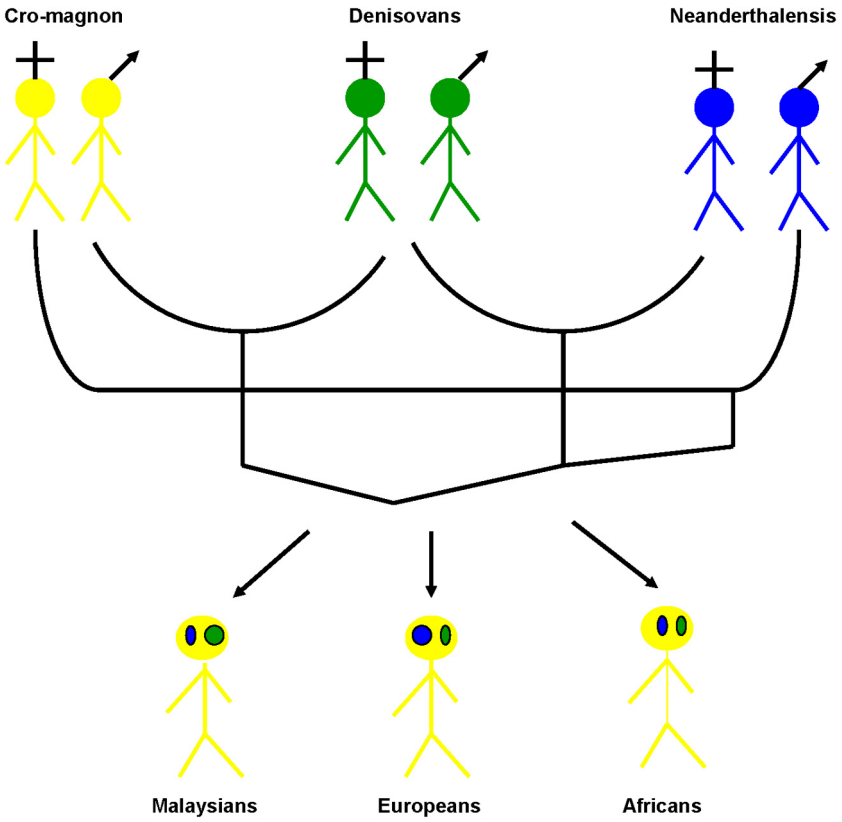


FIGURE 1 | The human rhizome. The current human genome harbors genes acquired vertically, predominantly from Cro-Magnon ancestors but also from Denisovans and Neanderthals. In addition to the Cro-Magnon genes, Europeans possess a few Denisovan genes, Africans possess a small and approximately similar amount of Denisovan and Neanderthal genes, and Malaysians possess few Neanderthal genes.

LATERAL SEQUENCE TRANSFERS IN BACTERIA AND THEIR RHIZOME

Lateral inheritance has been accepted as a major contributor of genome evolution in free-living bacteria since the Lederberg experiments on pneumococcus (Lederberg et al., 1951). Acquired genes play a major role in bacterial diversification by supplying previously unavailable traits, which permit the rapid exploitation of new environments (Ochman et al., 2005). Selfish and mobile genetic elements, such as virophages, plasmids and transposons, are the primary vehicles for lateral transfers among prokaryotes (Pace et al., 2008). The insertion sequences (IS) that were initially identified in the *Escherichia coli* galactose operon are also examples (Shapiro, 1969).

Conversely, the absence of evidence for lateral transfers has long been considered to be a fundamental characteristic of the genomic evolution in obligate intracellular bacteria. The isolated lifestyles of obligate intracellular bacteria were thought to reduce gene acquisition opportunities. Genetic exchanges have been judged to be negligible in these species because their intracellular environment is relatively constant and does not select for the genetic diversity promoted by more challenging environments (Moreno, 1998; Blanc et al., 2007a). These early views have progressively changed with the accumulation of genomic data, which have provided evidence for extended genetic exchanges in intracellular bacteria (Bordenstein and Reznikoff, 2005). Therefore, lateral sequence transfer is now viewed as an agent that promotes and maintains all bacterial species. Bacteria can acquire genes by several means, including conjugal gene transfer, phage-mediated insertion and native DNA from outside sources (Ochman et al., 2005). The process of gene loss has substantially contributed to the differences in the gene contents between the modern *Rickettsia* species. Additionally, the Typhus Group (TG) genomes were predominantly shaped by the reductive evolution from the ancestral *Rickettsia* genome.

However, the *R. bellii* genome includes numerous genes related to amoebal symbionts, perhaps due to ancient gene exchanges between an *R. bellii* ancestor and other amoebal hosts (Ogata et al., 2006). In a recent study, we provided evidence for lateral transfers between *Rickettsiales* species and other bacteria or other organisms. *R. canadensis* acquired genes from gamma-proteobacteria, *Bacteroidetes, R. akari,* and *R. felis* acquired genes from gamma- and delta-proteobacteria, and SFG species acquired genes from *Bacteroidetes* and *Cyanobacteria*, whereas the *R. belli* genome contained genes of a eukaryotic origin (Georgiades et al., 2011a). Finally, the plasmids in *R. felis* may have been acquired via conjugation (Ogata et al., 2005). The *R. felis* genome is actually a collection of genes and parts of genes with diverse evolutionary histories. The apparent horizontal transfer of different sized DNA segments corresponding to genes in various functional

categories has been detected, indicating that any segment of DNA may be horizontally transferred (Merhej et al., 2011). Other *Rickettsiales*, such as *Orientia* spp., also acquired genes by LGT from *Cyanobacteria*, protists, and viruses (Georgiades et al., 2011a). *Wolbachia* spp. also have highly recombinogenic genomes, as their sympatric lifestyle enables recombination among the intracellular bacterial community (Klasson et al., 2009). Although lateral transfers were thought to be rare phenomenon in obligate intracellular bacteria (Audic et al., 2007), the "mobilome," composed of mobile genetic elements, promotes horizontal genetic fluidity in *Rickettsia* species and has likely shaped the evolution of these genomes (Merhej and Raoult, 2011). Bacteria of the *Legionella* genus can also grow in amoebae, although a clear demonstration of intra-amoebal growth remains lacking for most of them (Thomas and Greub, 2010). Degtyar et al. demonstrated that the distribution of effector-encoding genes is highly variable in *Legionella* species: most genes of eukaryotic origin are present in different *Legionella* species, whereas others are specific to *L. pneumophila* (Degtyar et al., 2009). In another recent study, 557 laterally transferred genes were observed in *L. pneumophila*. Most of the transferred genes are part of the metabolism functional category. An exchange of genetic material with a common amoeba host most likely explains the multiple phylogenetic origins of a significant fraction of the *Legionella* genes (Coscolla et al., 2011). Other bacteria have also acquired genes through lateral transfers. Lawrence and Ochman (Lawrence and Ochman, 1997) proposed that at least 15% of the *E. coli* genome is atypical and may have arisen by recent lateral inheritance, while the diversity within the species *E. coli* and the overlap in gene content suggests a continuum rather than sharp species borders in the group of *Enterobacteriaceae* (Lukjancenko et al., 2010). Nelson et al. (Nelson et al., 1999) concluded that almost 25% of the *Thermotoga maritime* genes are most closely related to archaeal genes and have a history of gene transfer between these lineages (Gogarten et al., 2002). *Firmicutes, Bacteroidetes,* and gamma-proteobacteria species were observed to possess genes of *Rickettsiales* origin (Georgiades et al., 2011a). Approximately half of the species-specific genes in *Streptococcus* species have been proposed to be acquired by lateral transfer from diverse backgrounds. Specifically, multiple lateral sequence transfer events occurred during polyclonal infections among the nasopharyngeal *Streptococcus pneumoniae* strains recovered from a child suffering from chronic upper respiratory and middle-ear infections (Hiller et al., 2010). Finally, recent evidence has revealed the presence of horizontally transferred fragments of the human long interspersed nuclear element L1 in the genome of the strictly human pathogen *Neisseria gonorrhoeae* (Andersson and Seifert, 2011). Generally, it has been demonstrated that at least $81 \pm 15\%$ of the genes in each studied genome were involved in LGT at some point of their history even thought they can be vertically inherited after acquisition (Dagan et al., 2008). Taken together, these data suggest that the rhizome hypothesis is well suited for describing bacteria (Merhej et al., 2011) (**Figure 2**).

LATERAL SEQUENCE TRANSFERS IN ARCHAEA AND THEIR RHIZOME

Many Archaea inhabit extreme environments similar to those in which life originated. Although Archaea members may be seen as evolutionary relics of the earliest life forms, none of the organisms living today are primitive. All extant life forms are modern organisms that are well adapted to their ecological niches. Numerous authors have observed many horizontally transferred genes in Archaea, confirming that lateral sequence transfer is a wide-ranging phenomenon (Koonin and Galperin, 1997; Aravind et al., 1998). Koonin and Galperin (1997) observed large fractions of genes of bacterial or eukaryotic origins in Archaea genomes, suggesting a chimeric origin for Archaea. The percentages of horizontally transferred sequences in bacteria and Archaea are similar, ranging from 5% in *Methanococcus jannaschii* to 14% in *Aeropyrum pernix* (Garcia-Vallvé et al., 2000). Massive gene exchanges between the extremely thermophilic *Archaea* and the hyperthermophilic *Aquifex* have been suggested. Previous studies revealed that the *Aquifex* genome is a chimera that shares a large component with the *Archaea* genome in addition to the core gene set shared with the rest of the bacteria. Bacterial hyperthermophily has likely evolved secondarily within moderately thermophilic bacteria by the continuous acquisition of thermotolerance genes from pre-adapted hyperthermophiles, namely the

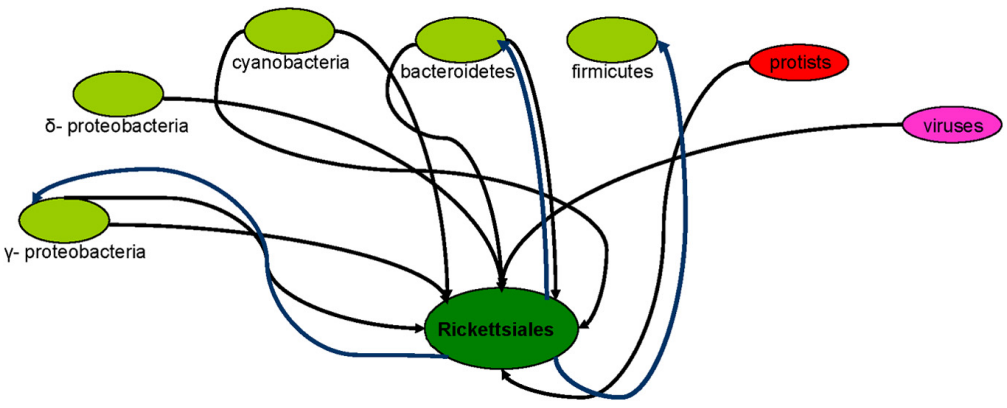


FIGURE 2 | The rhizome of bacteria—the Rickettsiales paradigm. Rickettsiales harbor genes from various bacterial species, protists, and viruses.

Archaea (Aravind et al., 1998). Finally, *M. jannaschii* protein sequences revealed that the number of proteins that are more similar to bacterial homologs significantly exceeds the number of those that are closer to eukaryotic homologs. The prevalence of the genes of bacterial origin in archaeal genomes suggests a genetic basis for the prokaryotic phenotype (Koonin and Galperin, 1997).

LATERAL SEQUENCE TRANSFERS IN VIRUSES AND THEIR RHIZOME

Viral genomes are not without lateral sequence transfers. A recently discovered virophage, Sputnik, was considered to be a vehicle that mediated lateral transfers between giant viruses (La Scola et al., 2008). The *Acanthamoeba polyphaga* mimivirus has unique features, including the presence of dsDNA, which was previously undocumented in viruses. Phylogenetic analyses identified a large number of bacterial homologs, suggesting an acquisition by lateral inheritance. Most of these genes were related to the orthologs in bacterial species, such as *L. pneumophila*, that are known to grow within amoebae. Mimiviruses likely acquired these genes from degraded or live bacteria sharing the same environment, particularly within amoebae (Moliner et al., 2010). In mammalian Herpesvirus, 141 putative transferred genes were identified, of which 91 were from gamma-herpesvirus, 42 from beta-herpesvirus, and 8 from alpha-herpesvirus, suggesting that gene acquisition in gamma-herpesvirus was more active than in the others (**Figure 3**). Although the functions of most of the putative transferred genes remain uncharacterized, many genes have been predicted to encode glycoproteins or membrane proteins (Fu et al., 2008). Furthermore, genomic analyses revealed that approximately 13% of the Herpesvirus proteins have clear sequence similarity to the human genome. The human homologs present in a large proportion of herpesvirus genomes, such as DNA polymerase and uracil-DNA glycosylase, are likely to have been acquired from a distant host by an ancestral Herpesvirus. Generally, gamma-herpesvirus genomes are particularly rich in genes that have a human counterpart (Holzerlandt et al., 2002).

Not only does the gene repertoire exhibit a substantial plasticity, but nearly all genes of all organisms, including Prokaryotes, Eukaryotes, Archaea and Viruses, have been exchanged or recombined at some point in time (Raoult, 2010b). Therefore, all life forms are mosaics that are part of the rhizome of life.

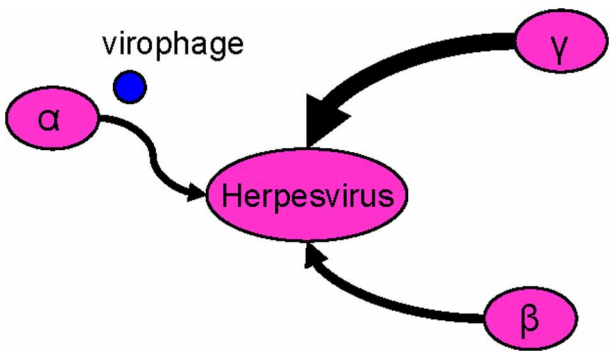


FIGURE 3 | The viral rhizome. The mammalian herpesvirus harbors genes from the α, β, and γ herpesvirus. Gene exchange often occurs by virophages.

INTRAGENIC RECOMBINATION

Other than lateral sequence transfer and recombination between various organisms, the origin of some genes cannot be simply explained because of the occurrence of intragenic recombination (Raoult, 2010a). Different varying sites recombine within the same gene locus, and intragenic recombination can generate new allelic variation at a locus. New gene products with potentially new properties will then arise (Watt, 1972). Chimeric gene sequences that could result from recombination between divergent alleles have been detected in several species (Kelly et al., 2009). A recent *R. felis* study revealed random transfers of DNA sequences in the *R. felis* genomes that occurred independently of gene function or sequence length. The apparent horizontal transfer of DNA segments with different sizes has been detected, thus indicating that any segment of DNA may be laterally transferred (Merhej et al., 2011).

Staphylococcus aureus expresses several proteases, including the thermolysin-like metalloprotease aureolysin. Sequence analyses revealed that the aur gene is present in two distinct types of related sequences and that it is very polymorphic. The gene trees constructed from aur and concatenated multilocus sequence typing (MLST) genes revealed several putative assortative recombination events, such as exchanges of the entire aur gene between divergent *S. aureus* lineages. Evidence for intragenic recombination events, such as exchanges of internal aur segments across aur genes, was also observed (Sabat et al., 2008). Moreover, results from Nicotiana illustrated that intragenic recombination may be a relatively common occurrence and can provide evidence for distinct parental contributions within diploid species of likely reticulate origin (Kelly et al., 2009). Bluetongue virus (BTV) is the pathogen that causes Bluetongue disease, which threatens sheep, deer, cattle, and goats. Mosaic viral genes have been observed to share a similar recombination event. Recombination can occur in BTVs and likely plays a role in the evolution and genetic diversity of the virus (He et al., 2010). However, the recombination rate of BTV is much lower than that of positive-sense RNA viruses, such as the foot and mouth disease virus, in which 10–20% of the viral genomes undergo recombination during a single replication cycle (Alejska et al., 2001). Intragenic recombination between two different existing alleles in a population can create new alleles. The role of this process in maintaining variation in a natural population has been investigated through the assumptions that one gene consists of two sites, each of which can mutate to an infinite number of unique alleles (Kimura and Crow, 1964; Strobeck and Morgan, 1978; Morgan and Strobeck, 1979). This process, like lateral sequence transfer, represents a challenge for the reconstruction of the evolutionary relationships between species because they cannot be represented adequately with classic trees (Kelly et al., 2009).

ORPHAN OR NEWLY CREATED GENES

Orphan genes constitute a class of lineage-specific genes that are not homologous to the sequences of other species (Fischer

and Eisenberg, 1999). They typically encode small proteins and have high non-synonymous substitution rates, but their functions remain unknown (Domazet-Loso and Tautz, 2003; Daubin and Ochman, 2004). Recently, a classification of ORFans has been proposed, dividing ORFans into singletons (unique predicted genes with no significant homolog), multipletons (orphan with one or more paralogs in its residing genome but none in other genomes), and lineage ORFans (orphan with homologs among a given taxonomic rank and none outside) (Boyer et al., 2010). Comparative analyses have revealed that new genes appear frequently in genomes and that useful genes can be retained, whereas purposeless genes are often removed from the genome. The current methods for identifying the newly acquired genes can be divided into three categories: compositional analysis, detection of phylogenetic anomalies, and comparison of genome content (Ochman et al., 2000; Nakamura et al., 2004; Gogarten and Townsend, 2005; Kuo and Ochman, 2009). Compositional analysis is the only method that does not rely on comparing sequences from multiple genomes. In short, this approach compares the sequence features, such as nucleotide composition and codon usage, of the genes within a genome. A direct genome alignment is the most frequently used approach for identifying new genes, in which the presence and absence of genes among the organisms is examined. This method entails comparing multiple related genomes and can provide the most direct evidence of genes that are gained or lost from a lineage (Kuo and Ochman, 2009). Each newly sequenced genome contains a significant number of ORFans (Toll-Riera et al., 2009). For example, in the 60 fully sequenced microbial genomes, 14% of the genes are species-specific orphans (Siew and Fischer, 2003), whereas 18% of the *Drosophila* genes are restricted to the *Drosophila* group (Zhang et al., 2007; Zhou et al., 2008). However, the origin of the orphan genes remains elusive (Merkeev and Mironov, 2008). One scenario proposes that they were derived from gene duplication events, in which one copy accumulated enough sequence changes so that the ancestral similarity is no longer detectable (Domazet-Loso and Tautz, 2003). Such ORFans were recently proposed to represent genes of viral or plasmid origin (Rocha et al., 2006), and some may correspond to genuine new genes formed *de novo* through diverse mechanisms of gene evolution (Boyer et al., 2010). This proposed mechanism has significantly impacted the process by which novel genes in mammals form, specifically in primates, in which 5.5% of orphan genes could have originated *de novo* from non-coding genomic regions (Toll-Riera et al., 2009). The formation of novel genes has also been described in *Drosophila* (Begun et al., 2006; Levine et al., 2006; Zhou et al., 2008) and *Saccharomyces cerevisiae* (Cai et al., 2008). To characterize the genetic basis of evolution and development in insects, the genomes of insect-specific proteomes were analyzed. The characterization of the proteomes-based on genome sequences provides a rapid method to approximate and update the putative proteomes as genome sequences become available. Using this approach, 50 insect-specific proteins were isolated, and many have been supported by experimental studies. The proteins-related to stress and immune responses constitute an extensive fraction of the proteins characterized in the insect-specific proteome. The presence of numerous insect-specific olfaction and cuticle development proteins underscores the significance of communication and adaptation to the environment during insect evolution (Zhang et al., 2007).

In a recent study by our laboratory, a small number of gene sequences in *Rickettsia* species were identified that did not match any database and may have resulted from de novo creation (Georgiades et al., 2011a). Indeed, 17 *Rickettsia* gene sequences do not possess homologs in the non-redundant (NR) database. The Ka/Ks ratio revealed that 15 of these sequences were either non-functional or had gained functionality later on. The probability of pseudogenization or a viral origin of these genes should not be excluded, however, because these genes were not found in the regions with traces of active or ancient integrated extrachromosomal elements, we strongly believe that they are novel genes (Georgiades et al., 2011a).

A study reconstructing the gene content of ancestral archaeal and proteobacterial genomes demonstrated that Archaea exhibit an abnormally high number of genesis events, particularly *Aeropyrum pernix*. The estimates for gene genesis also reveal at least 240 genes that originated at the branch leading to the Archaea. For Proteobacteria, this number was at least 320. Such genes can be considered typical of a taxon because they are unique and widespread within it (Snel et al., 2002).

Finally, the formation of new genes has been suggested to be essential to the viability of an organism. In the case of *Drosophila*, 59 *de novo* genes were observed to be as vitally essential as the old genes. The lethal phenotypes caused by the knockout of new genes suggested that the genes created *de novo* may integrate a vital pathway by interacting with existing genes, and this co-evolution may cause the new gene to become indispensable (Chen et al., 2010). Recently, an attempt to estimate the fraction of acquired genes that become dispensable and the features of those that are retained was made by examining the distribution of genes along an evolutionary lineage. Based on the numbers of the recent arrivals present in the *E. coli* K12 genome compared with the number of ancestral genes maintained by all members of this species, only 10–15% of the acquired genes were retained (Ochman and Davalos, 2006). Moreover, by establishing the origins of the genes acquired during the diversification of *E. coli* and *Shigella*, new genes that have distant homologs in other bacteria were acquired much more frequently but not retained as often as the acquired genes with no identifiable homologs (van Passel et al., 2008). These results suggest that genuinely novel genes, i.e., those that never conferred a function in a cellular genome, are more likely to persist in bacterial genomes (Kuo and Ochman, 2009).

In summary, gene creation is a continuous and unsettled phenomenon that is supported by the discovery of new genes that are permanently generated and are identified more frequently (Boyer et al., 2010; Raoult, 2010a). Genes created *de novo* illustrate the fact that nature is creative and not parsimonious. A rhizome is the most suitable representation for orphans; newly created genes should be represented as emerging roots on the rhizome.

THE TREE VERSUS THE RHIZOME OF LIFE

The TOL, initially elaborated by Darwin 150 years ago (Darwin, 1859), remains a biological fact for many supporters. For those

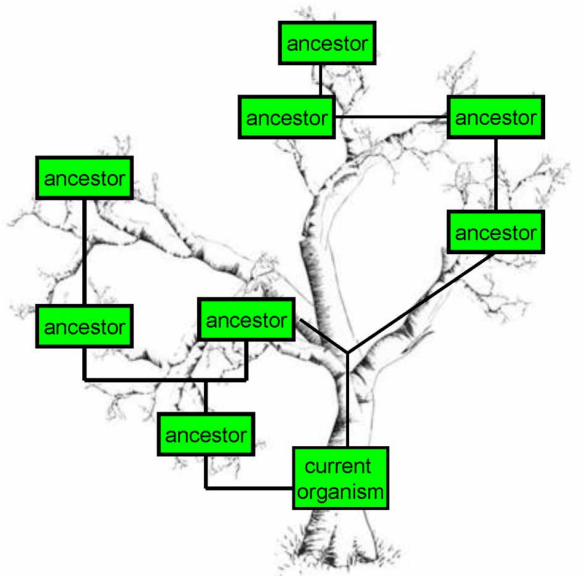


FIGURE 4 | Genealogic tree. A genealogic tree sample that harbors organisms with many ancestors closely or distantly related.

who question it, the TOL is nothing more than a scientific hypothesis not yet proven to be true. Darwin's theory reflects the religious and social beliefs of its time: the existence of one common ancestor at the origin of each branch is a direct resurgence of the Adam and Eve creation theory (Raoult, 2010b). The origin of life is a biblical definition (Penny, 2011) that reflects the "hierarchical natural order" originating from a single ancestor (Doolittle, 1999). Indeed, this hypothesis could be falsified and is based on the mixed message that classifications should be constructed as hierarchies because evolution is a branching process, and a hierarchical classification is a proof of a branching evolution (Doolittle and Bapteste, 2007). Nevertheless, the traditional biological explanations have mostly relied on the construction of a genealogical tree that describes lineages diverging from a last common ancestor. Such a hierarchy provides a dichotomous topology that structures biodiversity in the most informative way (Schliep et al., 2011). However, the robustness and universal scope of such tree-based evolutionary explanations has been recently questioned more frequently (Doolittle, 1999, 2009a,b; Bapteste et al., 2009; Dagan and Martin, 2009; Ragan and Beiko, 2009). For example, the ubiquity of lateral sequence transfers suggests that the TOL does not adequately represent prokaryote evolution (Bapteste and Boucher, 2008; Koonin and Wolf, 2008).

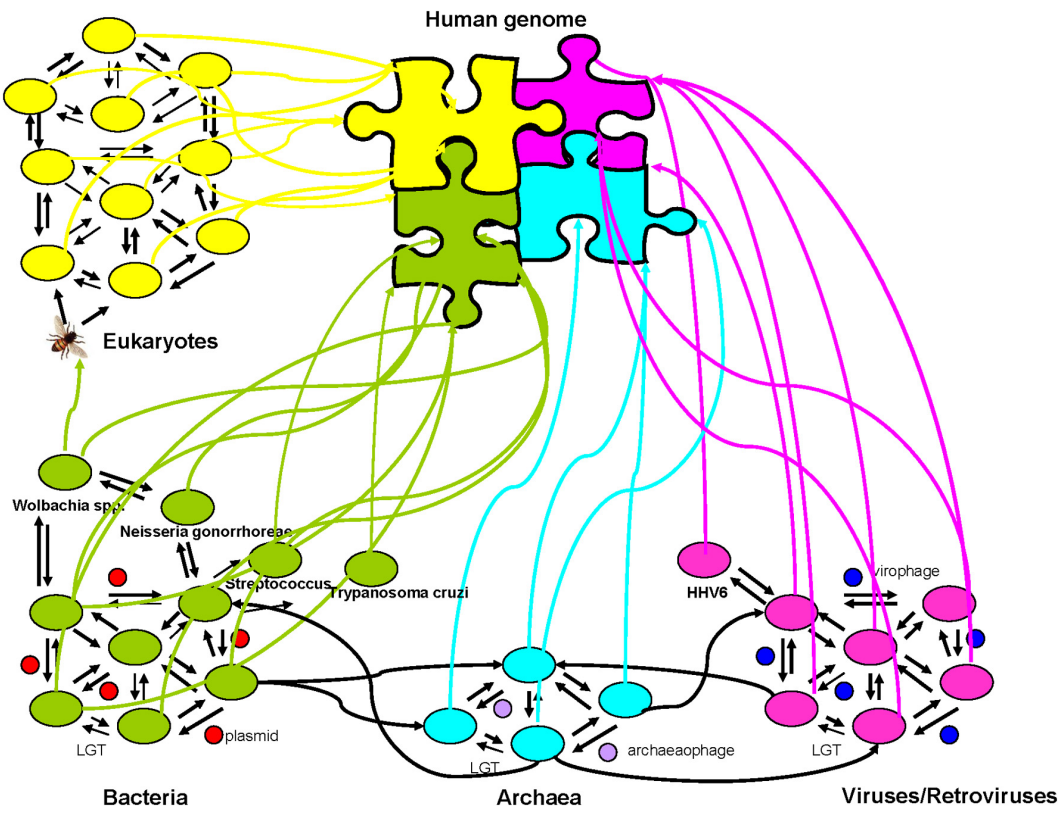


FIGURE 5 | The human genome as a mosaic and an ecosystem. Everything is a rhizome, in which lateral transfers occur between and among bacteria, viruses, archaea, and eukaryotes mediated by plasmids, virophages, and archaeophages. The modern human genome comprises vertically inherited elements from ancestors and from laterally inherited elements from other eukaryotes, bacteria, and viruses. The arrows identify the lateral sequence transfers, and the colorful arrows indicate the transfer events toward humans.

Bacterial classification implies that bacteria have a phylogeny and that the taxonomic system that functions adequately for the metazoans is also meaningful for the microbial world. Further, such a hierarchy and classification do not consider the numerous lateral sequence transfer events observed in all organisms. Up to 30% of the intra-species genome-to-genome variation in gene content results from lateral inheritance and gene loss. In some species, the pan-genome appears unlimited (Doolittle and Bapteste, 2007).

Human beings are complex ecosystems comprising more bacteria and viruses than eukaryotic cells in their mucosa, particularly in the gut. The human genome is a mosaic of genes with eukaryotic, bacterial, and viral origins. Bacteria harbor genes that originated from eukaryotic, viral, and archaeal organisms. Giant viruses also have chimeric genomes of different origins (Raoult, 2010a). Finally, genome analyses have revealed high proportions of newly generated ORFan genes (Fraser et al., 2000). An increasing number of scientists now argue that Darwin's TOL is best seen as an approximation to describe some parts of the living world but is less adequate elsewhere (e.g., viruses and prokaryotes). However, it is currently more complicated to place the root on the TOL (Ragan et al., 2009). A genealogic tree is a more adequate representation (**Figure 4**).

The theories of multiplicity and *de novo* creation need to be integrated in a post-Darwinian concept of the living species. When all these features are considered, the evolution of species seems more like a rhizome (Raoult, 2010a). A rhizome is a descriptive and epistemological model that does not organize the elements by a hierarchical coordination line, including a base and a root constituting the starting point of branches according to the well-known model of the Porphyrian tree (Eco, 1984). Instead, it is a model in which each element can affect or influence all the others (Deleuze and Guittari, 1972). A rhizome has a predominantly semiotic perspective. It is not easy to perceive things from the middle or from the top to the bottom. In contrast to trees, a rhizome connects any point with any other, and all of the features do not necessarily correspond to features of the same nature (Deleuze and Guittari, 1972). Consequently, this view of evolution considers the occurrence of multiplicities as emerging species grow from the rhizome with gene repertoires of various origins that allow for the multiplication and perpetuation of the species. Even the human genome is, and should be, viewed as a mosaic with eukaryotic, bacterial, archaeal, and viral genes that comprise an ecosystem that drives a network of gene exchange (**Figure 5**).

CONCLUSION

Darwin's TOL presents a single common ancestor on the root and different species on the major branches that separate and continuously diverge. However, the history of life cannot be attributed to a single ancestral species that yielded descendants that have adapted to their environment and developed into various species completely distinct and different from each other in such a short time. Furthermore, new species are continuously created and are not necessarily derived from other species (Raoult, 2010b). Furthermore, recent and massive gene transfer events have been identified in all living organisms. A rhizome could more adequately represent the multiplicities of genomes and their *de novo* creations. Emerging species grow from the rhizome with gene repertoires of various origins that allow for the multiplication of species under permissive environmental conditions (Raoult, 2010a). This view of evolution should resemble a clump of roots representing the multiple origins of the genetic repertoire of each species (Halary et al., 2010).

REFERENCES

Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizabeh, B., Luo, M., Plummer, F. A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S. G. E., Maiers, M., Guethlein, L. A., Tavoulakis, S., Little, A.-M., Green, R. E., Norman, P. J., and Parham, P. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334, 89–94.

Alejska, M., Kurzyniska-Kokorniak, A., Broda, M., Kierzek, R., and Figlerowicz, M. (2001). How RNA viruses exchange their genetic material. *Acta Biochim. Pol.* 48, 391–407.

Andersson, M. T., and Seifert, S. H. (2011). Opportunity and means: horizontal gene transfer from the human host to a bacterial pathogen. *MBio* 2, e00005–e00011.

Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R., and Koonin, E. V. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444.

Audic, S., Robert, C., Campagna, B., Parinello, H., Claverie, J. M., Raoult, D., and Drancourt, M. (2007). Genome analysis of *Minibacterium massiliensis* highlights the convergent evolution of water-living bacteria. *PLoS Genet.* 3:e138. doi: 10.1371/journal.pgen.0030138

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupré, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4, 34.

Bapteste, E., and Boucher, Y. (2008). Lateral gene transfer challenges principles of microbial systematics. *Trends Microbiol.* 16, 200–207.

Beauregard-Racine, J., Bicep, C., Schliep, K., Lopez, P., Lapointe, F. J., and Bapteste, E. (2011). Of woods and webs: possible alternatives to the tree of life for studying genomic fluidity in *Escherichia coli*. *Biol. Direct* 6, 39.

Begun, D. J., Lindfors, H. A., Thompson, M. E., and Holloway, A. K. (2006). Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags. *Genetics* 172, 1675–1681.

Berg, O. G., and Kurland, C. G. (2002). Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.* 19, 2265–2276.

Blanc, G., Ogata, H., Robert, C., Audic, S., Claverie, J. M., and Raoult, D. (2007a). Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res.* 17, 1657–1664.

Bordenstein, S. R., and Reznikoff, W. S. (2005). Mobile DNA in obligate intracellular bacteria. *Nat. Rev. Microbiol.* 3, 688–699.

Boucher, Y., Nesbo, C. L., and Doolittle, W. F. (2001). Microbial genomes: dealing with diversity. *Curr. Opin. Microbiol.* 4, 285–289.

Boyer, M., Madoui, M. A., Gimenez, G., La Scola, B., and Raoult, D. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4^{th} domain of life including giant viruses. *Plos One* 5:e15530. doi: 10.1371/journal.pone.0015530

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179, 487–496.

Callaway, E. (2007). Genomes within genomes. *Nature* 449, 6.

Chan, C. X., Beiko, R. G., Darling, A. E., and Ragan, M. A. (2009). Lateral transfer of genes and gene fragments in prokaryotes. *Genome Biol. Evol.* 1, 429–438.

Chan, C. X., Beiko, R. G., and Ragan, M. A. (2011). Lateral transfer of genes and gene fragments in *Staphylococcus* extends beyond mobile elements. *J. Bacteriol.* 193, 3964–3977.

Chen, S., Zhang, Y. E., and Long, M. (2010). New genes in *Drosophila*

quickly become essential. *Science* 330, 1682–1685.
Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to HGT. *Mol. Biol. Evol.* 28, 1481–1489.
Cohen, O., Rubinstein, N. D., Stern, A., Gophna, U., and Pupko, T. (2008). A likelihood framework to analyse phyletic patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3903–3911.
Coscolla, M., Comas, I., and Gonzales-Candelas, F. (2011). Quantifying non-vertical inheritance in the evolution of *Legionella pneumophila*. *Mol. Biol. Evol.* 28, 285–1001.
Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10039–10044.
Dagan, T., and Martin, W. (2009). Getting a better picture of microbial evolution en route to a network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2187–2196.
Darwin, C. (1859). *The Origin of Species.* London: Murray.
Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli. Genome Res.* 14, 1036–1042.
Degtyar, E., Zusman, T., Ehrlich, M., and Segal, G. (2009). A *Legionella* effector acquired from protozoa is involved in sphingolipids metabolism and is targeted to the host cell mitochondria. *Cell. Microbiol.* 11, 1219–1235.
Deleuze, G., and Guittari, F. (1972). *Rhizome: Introduction.* Paris, France: Ed de Minuit.
Domazet-Loso, T., and Tautz, D. (2003). An evolutionary analysis of orphan genes in *Drosophila. Genome Res.* 13, 2213–2219.
Doolittle, W. F. (1999). Phylogeneticc classification and the universal tree. *Science* 284, 2124–2129.
Doolittle, W. F. (2009a). Eradicating typological thinking in prokaryotic systematics and evolution. *Cold Spring Harb. Symp. Quant. Biol.* 74, 197–204.
Doolittle, W. F. (2009b). The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2221–2228.
Doolittle, W. F., and Bapteste, E. (2007). Pattern plurarism and the tree of life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.
Doolittle, W. F., and Papke, R. T. (2006). Genomics and the bacterial species problem. *Genome Biol.* 7, 116–123.
Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Munoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richard, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., and Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756.
Eco, U. (1984). "Dictionary vs. encyclopedia," in *Semiotics and the Philosophy of Language,* (Bloomington, USA: Indiana University Press), 46–87.
Emelyanov, V. V. (2001). *Rickettsiaceae, Rickettsia*-like endosymbionts and the origin of mitochondria. *Biosci. Rep.* 21, 1–17.
Esser, C., Martin, W., and Dagan, T. (2007). The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol. Lett.* 3, 180–184.
Fischer, D., and Eisenberg, D. (1999). Finding families for genomic ORFans. *Bioinformatics* 15, 759–762.
Fox, J. D., Briggs, M., Ward, P. A., and Tedder, R. S. (1990). Human herpesvirus 6 in salivary glands. *Lancet* 336, 590–593.
Fraser, C. M., Eisen, J. A., and Salzberg, S. L. (2000). Microbial genome sequencing. *Nature* 406, 799–803.
Fu, M., Denga, R., Wang, J., and Wang, X. (2008). Detection and analysis of horizontal gene transfer in herpesvirus. *Virus Res.* 131, 65–76.
Garcia-Vallvé, S., Romeu, A., and Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10, 1719–1725.
Georgiades, K., Merhej, V., El Karkouri, K., Raoult, D., and Pontarotti, P. (2011a). Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol. Direct* 6, 6.
Georgiades, K., Merhej, V., and Raoult, D. (2011b). The influence of Rickettsiologists on postmodern microbiology. *Front. Cell. Infect. Microbiol.* 1:8. doi: 10.3389/fcimb.2011.00008
Georgiades, K., and Raoult, D. (2011). The rhizome of *Reclimonas americana, Homo sapiens, Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. *Biol. Direct* 6, 55.
Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.
Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679.
Gould, S. J. (1987). The empire of the apes. *Nat. Hist.* 96, 20.
Halary, S., Leigh, J. W., Cheaib, B., Lopez, P., and Bapteste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 127–132.
Hao, W., and Golding, G. B. (2008). Uncovering rate variation of lateral gene transfer during bacterial genome evolution. *BMC Genomics* 9, 235.
He, C. Q., Ding, N. Z., He, M., Li, S. N., Wang, X. M., He, H. B., Liu, X. F., and Guo, H. C. (2010). Intragenic recombination as a mechanism of genetic diversity in bluetongue virus. *J. Virol.* 84, 11487–11495.
Hecht, M. M., Nitz, N., Araujo, P. F., Sousa, A. O., de Cássia Rosa, A., Gomes, D. A., Leonardecz, D., and Teixeira, A. R. L. (2010). Inheritance of DNA transferred from American trypanosomes to human hosts. *Plos One* 5:e9181. doi: 10.1371/journal.pone.0009181
Hiller, N. L., Ahmed, A., Powell, E., Martin, D. P., Eutsey, R., Earl, J., Janto, B., Boissy, R. J., Hogg, J., Barbadora, K., Sampath, R., Lonergan, S., Post, J. C., Hu, F. Z., and Ehrlich, G. D. (2010). Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfers during a chronic polyclonal pediatric infection. *Plos Pathog.* 6:e1001108. doi: 10.1371/journal.ppat.1001108
Holzerlandt, R., Orengo, C., Kellam, P., and Alba, M. M. (2002). Identification of new herpesvirus gene homologs in the human genome. *Genome Res.* 12, 1739–1748.
Kaper, J. and Hacker, J. (1999). *Pathogenicity Islands and Other Mobile Genetic Elements.* Washington, DC: ASM Press.
Kelly, L. J., Leitch, A. R., Clarkson, J. J., Hunter, R. B., Knapp, S., and Chasel, M. W. (2009). Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Mol. Biol. Evol.* 27, 781–799.
Kimura, M., and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49, 725–738.
Klasson, L., Westberg, J., Sapountzis, P., Naslund, K., Lutnaes, Y., Darby, A. C., Veneti, Z., Chen, L., Braig, H. R., Garrett, R., Bourtzis, K., and Andersson, S. G. E. (2009). The mosaic genome structure of the Wolbachia wRi strain infecting *Drosophila simulans*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5725–5730.
Kloesges, T., Popa, O., Martin, W., and Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.* 28, 1057–1074.
Konstantinidis, K. T., and Tiedje, L. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3160–3165.
Koonin, E. V. (2009). Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37, 1011–1034.
Koonin, E. V. (2012). "The postmodern state of evolutionary biology," in *The Logic of Chance: The Nature and Origin of Biological Evolution,* (Upper Saddle River, NJ: Pearson Education, Inc.), 403–404.
Koonin, E. V., Aravind, L., and Kondrashov, A. S. (2000). The impact of comparative genomics on our understanding of evolution. *Cell* 101, 573–576.
Koonin, E. V., Mushegian, A. R., Galperin, M. Y., and Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin of the archaea. *Mol. Microbiol.* 25, 619–637.
Koonin, E. V., and Galperin, M. Y. (1997). Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Dev.* 7, 757–763.
Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719.
Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., and Paabo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, 894–897.
Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C. A. (2005). The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* 15, 954–959.
Kuo, C. H., and Ochman, H. (2009). The fate of new bacterial genes. *FEMS Microbiol. Rev.* 33, 38–43.
La Scola, B., Desnues, C., Paguier, I., Robert, C., Barrassi, I., Fournous, G., Mechat, M., Suzan-Monti,

M., Forterre, P., Koonin, E., and Raoult, D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104.

Lawrence, J. G., and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383–397.

Lederberg, J., Lederberg, E. M., Zinder, N. D., and Lively, E. R. (1951). Recombination analysis of bacterial heredity. *Cold Spring Harb. Symp. Quant. Biol.* 16, 413–443.

Letunic, I., and Bork, P. (2006). Interactive tree of life. *Bioinformatics* 23, 127–128.

Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A., and Begun, D. J. (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9935–9939.

Lopez, P., and Bapteste, E. (2009). Molecular phylogeny: reconstructing the forest. *C. R. Biol.* 332, 171–182.

Lukjancenko, O., Wassenaar, T. M., and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* Genomes. *Microb. Ecol.* 60, 708–720.

Mayr, E. (1957). *The Species Problem*. Washington, DC: American Association for the Advancement of Science.

McNulty, S. N., Foster, J. M., Mitreva, M., Dunning Hotopp, J. C., Martin, J., Fischer, K., Wu, B., Davis, P. J., Kumar, S., Brattig, N. W., Slatko, B. E., Weil, G. J., and Fischer, P. U. (2010). Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One* 5:e11029. doi: 10.1371/journal.pone.0011029

Merhej, V., Notredame, C., Roeyr-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric *Rickettsia felis* paradigm demonstrates the random transfer of DNA sequences. *Mol. Biol. Evol.* 28, 3213–3223.

Merhej, V., and Raoult, D. (2011). Rickettsial evolution in the light of comparative genomics. *Biol. Rev.* 86, 379–405.

Merkeev, I. V., and Mironov, A. A. (2008). Orphan genes: function, evolution and composition. *Mol. Biol.* 42, 127–132.

Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596.

Moliner, C., Fournier, P. E., and Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol. Rev.* 34, 281–294.

Moreno, E. (1998). Genome evolution within the alpha Proteobacteria: why do some bacteria not possess plasmids and others exhibit more than one different chromosome? *FEMS Microbiol. Rev.* 22, 255–275.

Morgan, K., and Strobeck, C. (1979). Is intragenic recombination a factor in the maintenance of genetic variation in natural populations? *Nature* 277, 383–384.

Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* 36, 760–766.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Steward, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzbug, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.

Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M., and Fukatsu, T. (2008). *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiotic genes. *Genome Res.* 18, 272–280.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Ochman, H., Lerat, E., and Daubin, V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6595–6599.

Ochman, H., and Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science* 311, 1730–1733.

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P. E., Claverie, J. M., and Raoult, D. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *Plos Genet.* 5:e76. doi: 10.1371/journal.pgen.0020076

Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P. E., Parinello, H., Claverie, J. M., and Raoult, D. (2005). The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *Plos Biol.* 3:e248. doi: 10.1371/journal.pbio.0030248

Pace, J. K., Gilbert, C., Clark, M. S., and Feschotte, C. (2008). Repeated horizontal transfer of DNA transposon in mammals and other tetrapods. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17023–17028.

Penny, D. (2011). Darwin's theory of descent with modification versus the biblical tree of life. *PLoS Biol.* 9:e1001096. doi: 10.1371/journal.pbio.1001096

Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609.

Ragan, M. A., McInerney, J. O., and Lake, J. A. (2009). The network of life: genome beginnings and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2169–2175.

Ragan, M. A., and Beiko, R. G. (2009). Lateral genetic transfer: open issues. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2241–2251.

Raoult, D. (2010a). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Raoult, D. (2010b). *Dépasser Darwin*. France: Plon.

Raoult, D. (2011). A viral grandfather: genomics in 2010 contradict Darwin's vision of evolution. *Eur. J. Clin. Microbiol. Infect. Dis.* 30, 935–936.

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M., and Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060.

Renvoisé, A., Merhej, V., Georgiades, K., and Raoult, D. (2011). Intracellular *Rickettsiales*: onsights into manipulators of eukaryotic cells. *Trends Mol. Med.* 17, 573–583.

Rocha, M. T. G., Cooper, J. E., Smith, N. H., and Feil, E. J. (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* 239, 226–235.

Sabat, A. J., Wladyka, B., Kosowska-Shick, K., Grundmann, H., Maarten van Dijl, J., Kowal, J., Appelbaum, P. C., Dubin, A., and Hryniewicz, W. (2008). Polymorphism, genetic exchangeand intragenic recombinationof the aureolysin gene among *Staphylococcus aureus* strains. *BMC Microbiol.* 8, 129.

Schliep, K., Lopez, P., Lapointe, F. J., and Bapteste, E. (2011). Harvesting evolutionary signals in a forest of prokaryotic gene trees. *Mol. Biol. Evol.* 28, 1393–1405.

Shao, R., Kirkness, E. F., and Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 mini chromosomes in the human body louse *Pediculus humanus*. *Genome Res.* 19, 904–912.

Shapiro, J. A. (1969). Mutation caused by the insertion of genetic laterial into the galactose operon of *Escherichia coli*. *J. Mol. Biol.* 40, 93–105.

Siew, N., and Fischer, D. (2003). Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* 53, 241–251.

Snel, B., Bork, P., and Huynen, M. A. (2002). Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 12, 17–25.

Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., and Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452.

Strobeck, C., and Morgan, K. (1978). The effect of intragenic recombination on the number of alleles in a finite population. *Genetics* 88, 829–844.

Thomas, V., and Greub, G. (2010). Amoebae/amoebal symbionts genetic transfers: lessons from giant viruses neighbours. *Intervirology* 53, 254–267.

Toll-Riera, M., Bosch, N., Bellora, N., Castelo, R., Armengol, L., Estivill, X., and Alba, M. M. (2009). Origin of primate orphan genes: a comparative genomics approach. *Mol. Biol. Evol.* 26, 603–612.

van Passel, M. W., Marri, P. R., and Ochman, H. (2008). The emergence and fate of horizontally acquired genes in *Escherichia coli*. *PLoS Comput. Biol.* 4:e1000059. doi: 10.1371/journal.pcbi.1000059

Via, S. (2009). Natural selection in action during speciation.

Proc. Natl. Acad. Sci. U.S.A. 106, 9939–9946.

Watt, B. W. (1972). Intragenic recombination as a source of population genetic variability. *Am. Nat.* 106, 737–753.

Weismann, A. (1889). *Essays on Heredity and Kindred Biological Subjects.* Oxford, UK: Oxford University Press.

Yotova, V., Lefebvre, J. F., Moreau, C., Gbeha, E., Hovhannesyan, K., Bourgeois, S., Bédarida, S., Awevedo, L., Amorim, A., Sarkisian, T., Avogbe, P., Chabi, N., Hama Dicko, M., Sabiba Kou' Santa Amouzou, E., Sanni, A., Roberts-Thomson, J., Boettcher, B., Scott, R. J., and Labuda, D. (2011). An X-linked haplotype of Neanderthal origin is present among all non-African populations. *Mol. Biol. Evol.* 28, 1957–1962.

Zhang, G., Wang, H., Shi, J., Wang, X., Zheng, H., Wong, G. K. S., Clark, T., Wang, W., Wang, J., and Kang, L. (2007). Identification and characterization of insect-specific proteins by genome data analysis. *BMC Genomics* 8, 93.

Zhaxybayeva, O., and Gogarten, J. P. (2002). Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses. *BMC Genomics* 3, 4.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in *Drosophila*. *Genome Res.* 18, 1446–1455.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 January 2012; accepted: 16 April 2012; published online: 04 May 2012.

Citation: Georgiades K and Raoult D (2012) How microbiology helps define the rhizome of life. Front. Cell. Inf. Microbio. 2:60. doi: 10.3389/fcimb.2012.00060

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 31 October 2011
doi: 10.3389/fcimb.2011.00008

The influence of rickettsiologists on post-modern microbiology

Kalliopi Georgiades, Vicky Merhej and Didier Raoult*

Unité de Recherche en Maladies Infectieuses Tropical Emergentes, CNRS-IRD UMR 6236-198, Université de la Méditerranée, Marseille, France

Edited by:
D. Scott Merrell, Uniformed Services University, USA

Reviewed by:
Antje Flieger, Robert Koch Institute, Germany
Patrick Forterre, Université Paris-Sud, France

****Correspondence:***
Didier Raoult, Unité des Rickettsies URMITE, UMR CNRS 6236, IRD 198, Université de la Méditerranée, 27, Boulevard Jean Moulin, 13385 Marseille Cedex 05, France.
e-mail: didier.raoult@gmail.com

Many of the definitions in microbiology are currently false. We have reviewed the great denominations of microbiology and attempted to free microorganisms from the theories of the twentieth century. The presence of compartmentation and a nucleoid in Planctomycetes clearly calls into question the accuracy of the definitions of eukaryotes and prokaryotes. Archaea are viewed as prokaryotes resembling bacteria. However, the name archaea, suggesting an archaic origin of lifestyle, is inconsistent with the lifestyle of this family. Viruses are defined as small, filterable infectious agents, but giant viruses challenge the size criteria used for the definition of a virus. Pathogenicity does not require the acquisition of virulence factors (except for toxins), and in many cases, gene loss is significantly inked to the emergence of virulence. Species classification based on 16S rRNA is useless for taxonomic purposes of human pathogens, as a 2% divergence would classify all *Rickettsiae* within the same species and would not identify bacteria specialized for mammal infection. The use of metagenomics helps us to understand evolution and physiology by elucidating the structure, function, and interactions of the major microbial communities, but it neglects the minority populations. Finally, Darwin's descent with modification theory, as represented by the tree of life, no longer matches our current genomic knowledge because genomics has revealed the occurrence of *de novo*-created genes and the mosaic structure of genomes, the Rhizome of life is therefore more appropriate.

Keywords: definitions, prokaryotes, virus, archaea, metagenomics, tree of life, bacterial virulence factors, orphan genes

INTRODUCTION

Post-modern philosophy, also called the French theory (Wicks, 2003), states that the majority of theories, including scientific theories, are only based on meta-narratives expressing the influence of a culture at a given time. These scientific theories can be questioned when a change in techniques creates instability in the theory, as postulated by Karl Popper (Popper, 1959; Raoult, 2010a). In addition (and in the direction of post-modern theory), these theories can also be called into question due to an intellectual change of paradigm (Kuhn, 1962). The study of *Rickettsiae* has been challenging for the past few years because of the great difficulty in their handling. Moreover, the ancestors of *Rickettsiae* contributed to the birth of modern eukaryotes by transferring genes to the mitochondrion and the nucleus (Koonin, 2010; Renvoisé et al., 2011). During the explosion of microbial genetics, the study of *Rickettsia* did not benefit from the model of *Escherichia coli*, and rickettsiologists had to develop alternative approaches that did not include the common meta-narratives (Renvoisé et al., 2011). Among these approaches were those based on observations of the characteristics of intracellular bacteria, whose genomes and behaviors resemble those of viruses. Thus, *Rickettsia* has been classified as intermediate bacteria between the viruses and bacteria. Currently, the genomic revolution and "multiomics" have made it possible to analyze *Rickettsia* with many new tools (Bechah et al., 2010), and *Rickettsia* was among the species that were sequenced most quickly (Andersson et al., 1998; Ogata et al., 2001). This sequencing and generally all the work achieved by Rickettsiologists brought an important revision to the way of thinking with respect to *E. coli* and forced microbiologists to visualize the general theories concerning bacterial species in a different way, so several theories concerning bacteria had, or need, to be revised (Georgiades and Raoult, 2011a). In this work, our goal was to revise the overarching classifications and denominations used in microbiology. In particular, as postulated by post-modern philosophy (Lyotard, 1979; Williams, 1998), we know that the denomination of an object constrains it in its definition and that when the definition is inappropriate, one cannot conceive of the object in a reasonable way.

DEFINITION OF EUKARYOTES AND PROKARYOTES

The word "microbe," literally meaning "small life," was introduced by the French surgeon Charles Sédillot in 1878 to define infinitely small living organisms (Vallery-Radot, 1885). One of the most important advances in our understanding of the living world was the realization by the French scientist Edouard Chatton that there are two major groups of organisms that he named the prokaryotic (bacteria) and the eukaryotic (organisms with nucleated cells) type (Chatton, 1925; Stanier and van Niel, 1962; Sapp, 2005). This classification was adopted by Stanier and van Niel (1962) and the prokaryote–eukaryote dichotomy was universally accepted as the natural order of things until the 1970s and the emergence of rRNA

phylogenetics (Sapp, 2005). At that time, Woese achieved a comprehensive understanding of bacterial phylogeny using laborious molecular sequencing methods (Woese et al., 1975). Those data revealed two separate lineages among prokaryotes: the Archaea (Archaebacteria) and the Bacteria (Eubacteria). The prokaryote/eukaryote system was replaced by the "three domain system" and the classification of Eukarya, Archaea, and Bacteria (Woese, 1994). However, bacteria had always been defined largely in negative terms: they lacked a nucleus, compartmentation, and sexual reproduction (Sapp, 2005). This negative description is somewhat invalid because it does not define what a prokaryote is but rather what it is not (Pace, 2006). Furthermore, recent observations of Planctomycetes prove that the definitions of eukaryotes and prokaryotes are erroneous. Planctomycetes is a distinctive phylum of the domain Bacteria, in which the cells possess a different structural plan than other prokaryotes; the cells of all cultured and some uncultured species are divided into compartments by one or more membranes (**Figure 1**). In addition, in one particular species, *Gemmata obscuriglobus*, the nucleoid is enveloped in two membranes to form a nuclear body that is analogous to the structure of a eukaryotic nucleus. The existence of these organisms clearly calls into question the accuracy of the actual definitions of eukaryotes and prokaryotes (Fuerst, 1995, 2005, 2010). The nucleus of these cells likely resulted from autogenous membrane development in a prokaryote lineage (Taylor, 1976; Lake and Rivera, 1994; Glansdorff et al., 2008), most likely in Planctomycetes and the closely related *Chlamydia* (Ward et al., 2000; Horn et al., 2004; **Figure 2**). This theory has been strengthened by the discovery of nuclear envelope fold topology in Planctomycetes, which is analogous to the eukaryotic cell structure (Fuerst, 2005, 2010). Moreover, the eukaryotes all harbor mitochondria, or mitochondria-related genes, inherited from *Rickettsiales* (Golding and Gupta, 1995; Lang et al., 1999). Therefore, eukaryotes are younger than *Rickettsia*, their other ancestors are unknown, and there is no evidence that these ancestors had a nucleus. As it turns out, the three domain system, as previously defined, does not exist (Lake, 1988).

ARCHAEA

When they were identified in the late 1970s based upon ribosomal sequences, Archaea were viewed as a group of archaic bacteria (Woese and Fox, 1977). Indeed, because of their capacity of methanogenesis, archaea were supposed to live in ancient organisms and received the name of "archaea." This name is misleading as it speculates that these organisms resemble ancient cells and live in specific and "archaic like" environment. They have long been considered as extremophile bacteria that can be found in the most extreme environments (temperature, salinity, and pH). This explains the fact that archaea have not been extensively studied in clinical microbiology and their place among living organisms long went unrecognized.

Because of their archaic label, Archaea have been used as models of the early evolution of cellular life forms (Romano and Conway, 1996; Embley and Martin, 2006; Poole and Penny, 2007; Cox et al., 2008). The information processing machineries of archaea are considered ancestral forms of the more complex replication, transcription, and translation machineries of the eukaryotic cell (Gribaldo et al., 2010). Other evolutionary hypotheses about the path of life reject the archaic status of archaea. They suggested that the three domains of life evolved from a pre-cellular community containing different types of genes using a process that led to the fixation of specific subsets of genes in the ancestors of these domains (Woese, 1998). Considering evidence from molecular sequences, envelope structure, and motility mechanisms, other hypothesis suggested that the archaea evolved from Gram-positive bacteria as an adaptation to hyperthermophilic or hyperacidity (Cavalier-Smith, 2002) or in response to antibiotic selection pressure (archaea are resistant to a wide variety of antibiotics that are

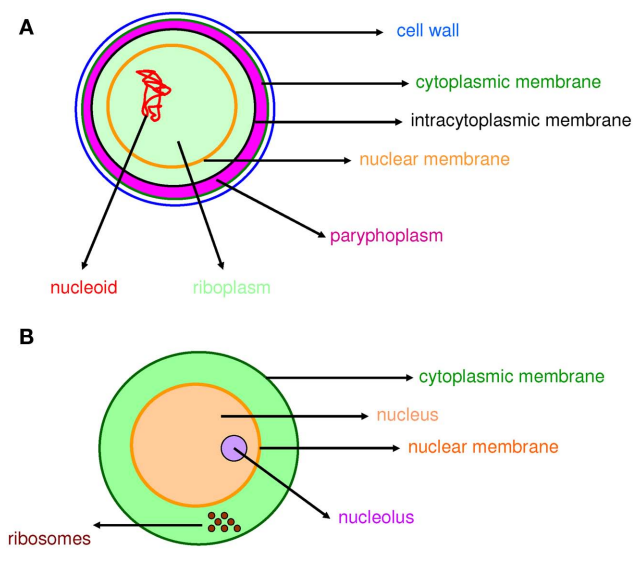


FIGURE 1 | Compartmentation in prokaryotes and eukaryotes. Compartmentation of *Gemmata obscuriglobus* **(A)** and a eukaryotic cell **(B)** is comparable.

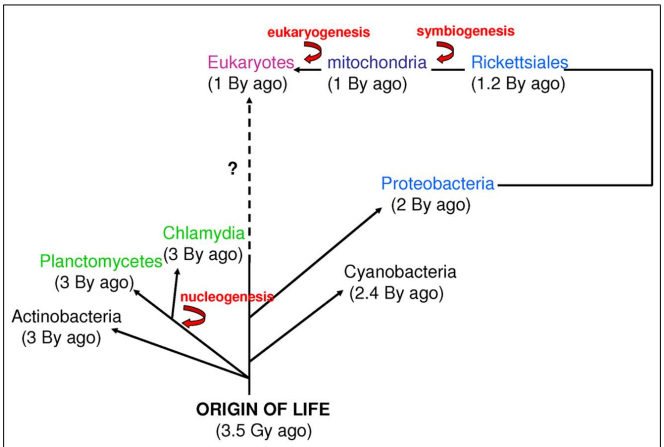


FIGURE 2 | Time scale of eukaryogenesis and nucleogenesis. Eukaryotes are not the only species with compartmentation. First eukaryotes emerged from an endosymbiotic event. The first nucleus appeared approximately 3 billion years ago in Planctomycetes and *Chlamydia*. These numbers are approximations (Bromham and Penny, 2003; Cavalier-Smith, 2004; Trevors and Abel, 2004).

primarily produced by Gram-positive bacteria; Gupta, 1998a,b, 2000).

Recent results obtained using molecular approaches and metagenomic studies have changed our perspective of the nature and the diversity of archaea. Indeed, archaea were considered predominant over bacteria in all extreme environments. This is true for high-temperature environments, as only archaea can thrive at temperatures from 95 to 113°C (Huber et al., 2000). However, in all other situations, species of Bacteria and Eukarya have been found together with those of archaea (Aravalli et al., 1998; DeLong, 1998; Rothschild and Mancinelli, 2001). Novel archaea have been isolated from a variety of temperate and cold environments (Cavicchioli, 2006), agricultural and forest soils, plankton, fresh water lake sediments, and the deep waters of oceans (Schleper et al., 2005). Archaea seem to constitute a major part of global ecosystems. They were estimated to account for approximately 34% of the total marine biomass of Antarctica (DeLong, 1998) and for nearly 20% of the total marine picoplankton biomass worldwide (Karner et al., 2001). The ubiquitous abundance of archaea and their influence on biogeochemical cycles remain largely unexplored. A recent tentative to infer the ancestral conditions of life suggests that he last common ancestor of archaea has been hyperthermophilic and mesophilic species have showed adaptations to cooler environments (Groussin and Gouy, 2011).

Methanogenic bacteria play a paramount role in digestion processes. Indeed, metagenomic analysis of the gut flora in three healthy individuals established that *Methanobrevibacter smithii* comprised up to 11.5% of the gut microorganisms (Eckburg et al., 2005). On the contrary, while many studies using 16S rDNA sequencing confirmed the presence of *M. smithii* in the human gut, the prevalence was low and *Methanosphaera stadtmanae* was not detected in most cases (Miller and Wolin, 1982; Belay et al., 1988; Dridi et al., in revision). This contrast is due to limitations in the experimental protocols that are largely designed for the study of bacteria. Recently, in our laboratory we developed an optimized protocol for the extraction and specific PCR-based detection of *M. smithii* and *M. stadtmanae* in DNA stool samples, using specific primers (Dridi et al., 2009). Using this protocol it was demonstrated that all individuals carried methanogenic archaea with a high prevalence of *M. smithii* (95.5%). The application of this specific approach allowed the isolation of *Methanomassiliicoccus luminyensis* and its description as a new species (Dridi et al., in revision), and the same protocol can be used to identify other archaeal species (Dridi et al., 2009, 2011). It is obvious that previous methods did not allow the identification of Archaea because they were not designed for Archaea identification.

Molecular experiments and genomic approaches have suggested that the different criteria used to define archaea are not completely valid. The definition currently used for Archaea merely cloaks our lack of knowledge of this domain of life. Undoubtedly, Archaea are not a form of "archaic" bacteria, they rather represent a distinct evolutionary domain.

BACTERIAL VIRULENCE FACTORS

It is not surprising that many people believe that bacteria that are dangerous to us are better armed than non-pathogenic bacteria. Toxins were identified in 1884 and defined as virulence factors; since, early genetic studies on bacterial virulence demonstrated that removing a certain number of genes from pathogenic bacteria decreases their capacity to infect hosts. Therefore, the conclusions of most studies on bacterial virulence, driven by anthropocentric intuition and perspective, suggested, and some still suggest that non-pathogenic bacteria lack supplementary virulence factors (Lawrence, 1999; Ochman et al., 2005).

An outstanding example of this way of thinking is the *Shigella* paradigm. *Shigella* spp. are human pathogens associated with bacillary dysentery, or shigellosis. *Shigella dysenteriae* causes deadly epidemics in many of the world's poorest countries. *Shigella* spp. and *E. coli* have always been considered closely related, and they have even been placed in the same species (Pupo et al., 2000). However, most *E. coli* strains are commensals of the human intestine (Maurelli et al., 1998), and *Shigella* spp. differ from *E. coli* in their lack of certain phenotypic traits, such as extracellular mobility and the ability to ferment lactose and many sugars (Karaolis et al., 1994; Pupo et al., 2000). Similar to *S. dysenteriae*, pathogenic enteroinvasive *E. coli* lack lysine decarboxylase (LDC) activity. In a study by Maurelli et al. (1998), the induction of LDC expression attenuated the virulence of a transformed strain of *S. flexneri*. It seems plausible that *Shigella* evolved from the *E. coli* complex through the acquisition of a plasmid containing critical genes. Plasmids of *Shigella* spp. have been directly associated with virulence and were even named "virulence plasmids" after their discovery (Hale et al., 1983). Furthermore, actin-based motility initiated by the *icsA* gene has also been reported to be a virulence factor (Goldberg et al., 1994). However, virulence increased after massive gene deletions (Maurelli et al., 1998). In conclusion, *S. dysenteriae* was not found to have more virulence genes than related bacteria (Georgiades and Raoult, 2011a).

Many recent comparative genomics studies have demonstrated that the specialization of bacteria for the colonization of eukaryotic hosts is associated with massive gene loss (Nierman et al., 2004; Merhej et al., 2009a) and the loss of identified "virulence factors" (Audic et al., 2007). Genomic analysis has revealed that *Borellia recurrentis*, the agent of deadly louse-borne relapsing fever, encodes fewer putative virulence factors than *Borellia duttonii* (Lescot et al., 2008). Gene loss has also accompanied the evolution of pathogenic *Bordetella* species (Cummings et al., 2004) and gene deletions in *Mycobacterium tuberculosis* have resulted in a hypervirulent phenotype (Bokum et al., 2008). Finally, in a study by Audic et al. (2007), the number of putative virulence factors was found to be higher in water-dwelling bacteria than in any other categories of bacteria, including specialized pathogens (Audic et al., 2007).

One of the best examples of genome reduction can be found in the epidemic-causing *Rickettsia prowazekii*, which is the most dangerous rickettsial species. Genome comparisons of *R. prowazekii* with the less virulent *R. conorii* have revealed that *R. prowazekii* is a subset of *R. conorii,* with only 834 open reading frames (ORFs) compared to the 1,374 ORFs of *R. conorii* (Ogata et al., 2001). Although intracellular motility has been considered a virulence factor of *Shigella* (Goldberg and Theriot, 1995) and *Listeria monocytogenes* (Tilney and Portnoy, 1989; Mounier et al., 1990), *R. prowazekii* is completely immobile in the cytoplasm (Teysseire et al., 1992). Actin-based motility in *R. conorii* and *R. rickettsii*

requires two proteins functioning together, Sca2 and RickA, suggesting that these two proteins could be virulence factors of *R. rickettsii*. *R. typhi* possesses only the Sca2 protein and is also mobile in the cytoplasm but less than *R. conorii* (Teysseire et al., 1992; **Figure 3**). However, none of these proteins are found in *R. prowazekii*, which lacks actin-based motility (Kleba et al., 2010). Consequently, motility is not a virulence factor *per se* but can be part of a virulence repertoire in some pathogens (Georgiades and Raoult, 2011a). Other studies have also demonstrated genome reduction to a lower extent in the extremely successful and fit *R. africae*, the agent of African tick-bite fever. In contrast with their possession of virulence factors, *R. prowazekii* and *R. africae* have the most and the least decayed genomes, respectively, among pathogenic *Rickettsiae* (Fournier et al., 2009). A comparison of *R. africae* with *R. rickettsii* suggested the loss of essential genes in *R. rickettsii* as a possible factor involved in the development of virulence (Fournier et al., 2009). In general, pathogenic *Rickettsia* species lack what was defined as "pathogenicity islands" and that are present in other bacterial pathogens (Hacker and Kaper, 2000). It has been suggested that plasmids contain genes encoding proteins responsible for host recognition, invasion, and pathogenicity. The presence of plasmids in *Rickettsia* species, however, did not show any correlation with virulence (Paddock et al., 2004; Ogata et al., 2005; Blanc et al., 2007a). The examples of *Rickettsiae* and *Shigella* spp. show that the terms "pathogenicity islands" and "virulence plasmids" are misleading. The genomic analysis of rickettsial species has revealed that the shift to pathogenicity does not require the acquisition of new genes, but in more cases, and not only in *Rickettsia*, gene loss seems to be implicated in the emergence of virulence (Moran, 1996; Andersson and Kurland, 1998; Andersson and Andersson, 1999; Blanc et al., 2007a; Darby et al., 2007; Merhej et al., 2009a). In a recent study in our laboratory, we demonstrated that the only features found at higher levels in extremely dangerous bacterial pathogenic species than in closely related less pathogenic species were toxins and toxin–antitoxin modules (TA; Georgiades and Raoult, 2011b).

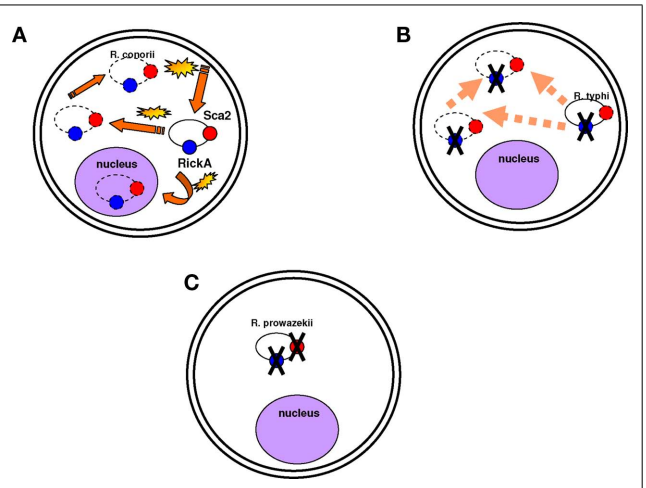


FIGURE 3 | Motility is not necessarily a virulence factor. (A) *Rickettsia conorii* is mobile in the cytoplasm and in the nucleus and moves quickly. Actin-based motility is associated with the Sca2 and RickA proteins (represented in red and blue circles respectively). Yellow stars are to demonstrate that the bacterium moves fast in the cytoplasm; dotted lines are to show that the bacterium can be found anywhere in the cytoplasm and even in the nucleus. **(B)** *Rickettsia typhi* is also mobile in the cytoplasm, but it moves less quickly than *R. conorii*. Its mobility is associated with the Sca2 protein. **(C)** *R. prowazekii* is completely immobile in the cytoplasm. The Sca2 and RickA proteins are absent.

In conclusion, except for toxins and TA modules, which have a direct effect and are indeed virulence factors, other products named "virulence factors" are, in reality, associated with fitness in a genomic context and in a specific environment, including in tested experimental models. Comparative genomics have shown than pathogenic bacteria have smaller genomes than non-specialized bacteria. Therefore, it is not possible to say that supplementary virulence factors establish pathogenicity, but rather, the overall gene repertoire is more associated with virulence than specific genes. In a recent study, the deletion of four different gene clusters in fungi attenuated their virulence in plants, while deletion of the "divergence cluster 8–12" (region encoding effector genes with mow sequence conservation) caused a hypervirulent fungal phenotype (Schirawski et al., 2010). Under these conditions, a virulent gene repertoire is composed of both present and absent genes. The term "virulence factor" seems to be invalid, and we propose that it should be abandoned.

PHYLOGENY AND TAXONOMY

Biological dogma states that phylogeny reflects taxonomy. Indeed, the 16S rRNA sequence has been widely used for the description of many newly classified bacterial species (Rosello-Mora and Amann, 2001; Drancourt et al., 2004; Roux et al., 2004). A 16S rRNA divergence of 1–2% is considered to correspond to approximately 50 million years of divergence (Ochman et al., 1999; Ogata et al., 2001), and a cut-off of 98.7% similarity in 16S rRNA reflects a new species (Stackebrandt and Ebers, 2006). However, an accurate delineation of bacterial species cannot be guaranteed by the use of ribosomal DNA sequence identity, which often leads to misleading species definitions (Fox et al., 1992; Rosello-Mora and Amann, 2001). *Bartonella henselae* has two copies of 16S rRNA in some cases, which likely emerged through recombination (Sanogo et al., 2003), and these copies present a divergence higher than 1.3% (Viezens and Arvand, 2008). For several bacterial species, the presence of multiple copies of the 16S rRNA gene has been documented (Acinas et al., 2004). Although generally, these multiple copies in an organism are either identical or nearly identical, in some cases, they are divergent enough to overestimate the number of bacterial species. This overestimation can be seen in the case of *Delisea pulchra* strains, in which 16S rRNA gene copies were used to illustrate the effects of 16S rRNA heterogeneity in the marine bacterial community (Dahllöf et al., 2000; Adékambi et al., 2008). The use of 16S rRNA for such analysis is limited due to its inherent heterogeneity (Dahllöf et al., 2000). Moreover, using the molecular clock scale based on 16S rRNA as a species definition criterion, specialized bacteria within mammalian hosts are not defined as species (Georgiades and Raoult, 2011a). Species definition cannot be based on the percent divergence of 16S rRNA because bacteria having a divergence less than 1.3% correspond to bacterial complexes rather than species (Doolittle and Papke, 2006).

There are 9,000 validated bacterial species and 1.5 million eukaryotic species, even though the biomass of bacteria is comparable to that of eukaryotes; this suggests that use of the 16S rDNA sequence as a taxonomic tool is not adapted to the definition of species. Furthermore, genomic contents are not represented by phylogeny. In a study based on the genomic content comparison of bacteria with different lifestyles, discrepancies between taxonomy and gene content were observed (Audic et al., 2007; Merhej et al., 2009b). The phylogenomic analysis yielded a tree similar to the one produced using the 16S rDNA gene sequence. However, γ-proteobacteria appeared to be divided into three groups, confirming that these species were more similar to each other in terms of gene content than to their close phylogenetic relatives (Audic et al., 2007). Similarly, rickettsial species and relatives, such as *Wolbachia* and *Ehrlichia* species, comprise an α-proteobacterial clade characterized by small genomes; this clade is distantly related to other α-proteobacterial species with larger genomes (Moran, 2002). Phylogenetic analysis of *Rickettsia* species based on 16S rRNA sequences has been frequently performed; however, significant inferences about intragenus phylogeny are not possible because the sequences are almost identical (Roux and Raoult, 1997). In fact, the official molecular criteria used for the classification of a bacterial species, DNA/DNA hybridization >70%, GC content <5%, and a 16S rRNA divergence <1, 3%, cannot be applied to *Rickettsia* species. A 16S rRNA divergence <2% alone would classify all *Rickettsiae* within the same species (Fournier and Raoult, 2009). Furthermore, based on this criterion, bacteria specialized to mammalian hosts are not defined as a species (Georgiades and Raoult, 2011a). *Homo sapiens* emerged approximately 250,000–400,000 years ago, while the first human-specialized pathogenic bacterial species, *M. tuberculosis*, emerged much later, only 20,000 years ago (Wirth et al., 2008). For organisms such as archaea, bacteria, and some unicellular eukaryotes, the species and gene trees do not show much identity with each other on an evolutionary scale (Bapteste et al., 2009). This result is due to the fact that individual gene histories can be different from the history of a species. Over the past 15 years, lateral inheritance (as opposed to vertical inheritance) has been proven to be a major evolutionary force in microorganisms (Bapteste and Boucher, 2008). Examples of extensive chimerism and LGT across prokaryotes are common, and it is absolutely plausible that every gene in prokaryotes has been affected by LGT at some point in evolutionary history (Bapteste et al., 2009). With this in mind, whole gene content and present and absent genes should be taken into consideration when searching for a reliable species classification (**Figure 4**).

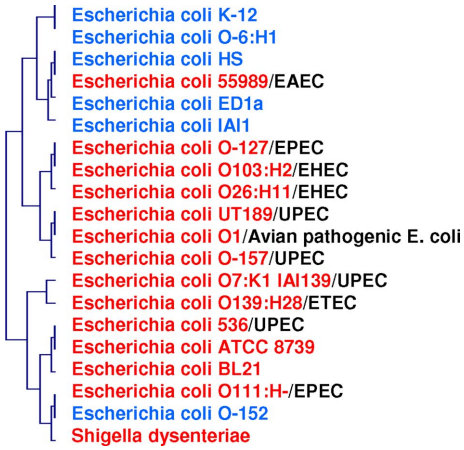


FIGURE 4 | Phylogenomic tree based on whole gene content (present/absent genes) in pathogenic and non-pathogenic *E. coli* species and *S. dysenteriae*. Two clusters are formed: one for pathogenic species (in red), and one for non-pathogenic species (in blue). Pathogenic strains are divided into five groups: EPEC, enteropathogenic; EHEC, enterohemorrhagic; UPEC, uropathogenic; ETEC, enterotoxigenic; EAEC, enteroaggregative.

DEFINITION OF A VIRUS

The discovery of giant viruses with large genomes has raised many questions about virus definitions and evolution. According to Lwoff, viruses have typically been defined as "filterable infectious agents" smaller than 200 nm that are unable to undergo binary fission and have one type of nucleic acid with few protein-encoding genes (Lwoff, 1957). Giant viruses, such as mimivirus (Raoult et al., 2004, 2007) and mamavirus (La Scola et al., 2008), challenge the size criteria used for the definition of a virus. These viruses, with an icosahedral capsid diameter of nearly 400 nm, have particle sizes comparable to that of bacteria such as *Mycoplasma* (La Scola et al., 2003; Raoult et al., 2004). Mimivirus possesses a large double-stranded DNA genome (1,181 kb). The mimivirus genome has 1,262 putative ORFs, of which 911 are predicted to be protein-coding genes, and 298 could be associated with functional attributes (Raoult et al., 2004). Mamavirus has a slightly larger genome than mimivirus (1,200 kb), and 99% of its predicted genes are orthologous to mimivirus ORFs (Colson and Raoult, 2010). The concept of the small particle (and genome) that once defined viruses is no longer valid.

The discovery of large viruses prompted a re-evaluation of the commonly used viral isolation methods and consideration of the role played by amebae as a source of giant viruses. Because amebae ingest any particle that is larger than 100 nm, these phagocytes represent a potential source of giant viruses with chimeric repertoires (Raoult and Boyer, 2010). Indeed, another virus, Marseillevirus, has recently been isolated from amebae. It has a diameter of 250 nm and a genome size of 368,454 bp (Boyer et al., 2009). Mimivirus, Mamavirus, and Marseillevirus belong to the *Mimiviridae*, a family in the group of nucleo-cytoplasmic large DNA viruses (NCLDVs; Iyer et al., 2006; Boyer et al., 2009). Genomic analysis of the giant viruses showed that only 4.6 and 11.2% of the ORFs of mimivirus and marseillevirus, respectively, had homologs in the NCLDV core gene set. Hence, the majority of the genome is lineage-specific. In addition to the core genome, the gene repertoire of these amoeba-associated viruses contains duplicated genes, ORFans and genes likely acquired through LGT. Indeed, a substantial proportion of the genome exhibits sequence similarities to gene homologs found in bacteria, archaea, eukaryotes, and viruses (Raoult et al., 2004). Using phylogenetic analyses, a bacterial or bacteriophage origin has been inferred for 49 genes and a eukaryotic origin for 85 genes of the marseillevirus genome (Boyer et al., 2009). Likewise, 60 genes from the mimivirus genome had reliable homologs

in cellular species and seemed to be acquired from eukaryotes, especially from amebae (Moreira and Brochier-Armanet, 2008). These chimeric gene contents may have resulted from acquisitions through LGT involving the eukaryotic host (ameba) and sympatric bacteria and viruses. Amebae may serve as a genetic mixing bowl from which giant viruses may have gathered a complex set of genes, leading to large chimeric genomes (Raoult and Boyer, 2010).

The genomes of giant viruses help to elucidate their origin and early evolution. The position of viruses within the tree of life (TOL) has been a subject of disagreement. Indeed, the classification of organisms into a universal TOL based on ribosomal RNA sequences (Pace, 2006) evidently excludes viruses, which lack ribosomes. As acellular organisms, viruses were intentionally not represented with other living ribosome-encoding organisms in the TOL (Moreira and Lopez-Garcia, 2009). Like other viruses, the mimivirus genome contains genes for replication. Surprisingly, the genome of mimivirus also contains genes that code for components of translation machinery never before found in viruses, including four amino-acyl transfer RNA synthetases, peptide release factor 1, the translation elongation factor EF-TU, and translation initiation factor 1 (Raoult et al., 2004). The presence of these genomic features has triggered a reappraisal of the definition of living beings (Raoult and Forterre, 2008) and the evolutionary implication of viruses. The phylogenetic analysis of mimivirus proteins that have closely related eukaryotic homologs support the appearance of mimivirus as representing a fourth domain of life together with bacteria, archaea, and eukaryotes (Raoult et al., 2004). Indeed, there are some genes that allow tracing history, including DNA processing genes, even thought a whole, complete organism cannot be represented by a classic TOL. An additional genomic study revealed the early emergence of NCLDVs whose core genome is as ancient as the three currently accepted domains of life (Boyer et al., 2010a). These findings confirm previous hypotheses stating that viruses may be at the origin of many eukaryotic genes (Villarreal and De Filippis, 2000; Forterre, 2006) and might have contributed to nucleus formation (Bell, 2001; Takemura, 2001). Thus, the study of giant virus genomes sheds light on the origin of eukaryotes and emphasizes the possible role played by capsid-containing organisms in the evolution of ribosome-encoding organisms.

METAGENOMICS AND MICROBIAL DIVERSITY

The study of many species is difficult or even impossible, mainly due to our inability to culture them in the laboratory (Zengler et al., 2005). Metagenomics, or the culture-independent genomic analysis of an assemblage of organisms, allows us to study microorganisms by deciphering their genetic information from DNA that is extracted directly from communities of environmental microorganisms. Metagenomics has revealed that the vast majority of microbial diversity has been missed using cultivation-based methods (Handelsman, 2004; Riesenfeld et al., 2004; Eckburg et al., 2005). Indeed, approximately 10 and 60% of the sequences from environmental microbial and viral metagenomes, respectively, are novel sequences; they have no significant similarity to any sequence in the GenBank non-redundant database (Tyson et al., 2004; Venter et al., 2004; Edwards and Rohwer, 2005). Thus, our knowledge has been gleaned from the relatively small number of presently culturable representatives while ignoring the "uncultured majority" (Hugenholtz et al., 1998).

Metagenomics has offered unprecedented insights into microbial diversity and sparked a revolution in the field of microbiology. Historically, microbiology has focused on single species in pure laboratory cultures; thus, the understanding of microbial communities has lagged behind the understanding of their individual members. In addition, limited information about physiology and functional roles can be gained from microbes in culture. Metagenomics is a new tool to study microbes in the complex communities in which they live and to begin to understand how these communities work. Indeed, metagenomics relies on high-throughput sequencing, which permits the isolation of large portions of genomes, providing access to protein-coding genes and biochemical pathways. Metagenomics focuses on profiling the functions encoded by a microbial community in a selected environment rather than the types of organisms producing them. Analysis of the genomic content of communities of organisms sheds light on the metabolic variability of an environment and on specific physiological functions (Eckburg et al., 2005; Dinsdale et al., 2008). Metagenomic studies of the pathogen-associated microbiome have allowed for an understanding of the role of microbial communities and their clinical implications (Gill et al., 2006; Ley et al., 2006; Turnbaugh et al., 2006; Willner et al., 2009). Information from metagenomic libraries has the ability to enrich our knowledge and has applications in many aspects of industry, therapeutics, and environmental sustainability.

Metagenomic approaches have revealed insights into environmental features with important evolutionary implications. Metagenomic functional analyses of ecosystems have revealed the correlation between geochemical conditions, metabolic capacity, and genetic diversity in microbial communities (Edwards et al., 2006; Frias-Lopez et al., 2008; Simon et al., 2009). Indeed, sequencing projects provide a means for sampling the genetic diversity of natural microbial populations by estimating the rate of recombination and have the potential to reveal much about the evolution of these populations (Johnson and Slatkin, 2009). Moreover, this gene-centric approach to environmental sequencing suggests that the functional profile predicted from environmental sequences of a community is similar to that of other communities whose environments of origin pose similar demands. These findings have provided insight into the processes of adaptation and the evolution of life on earth.

Notably, metagenomics represents a powerful tool that can be used to access the abundant biodiversity of environmental samples, but its accuracy depends on many limitations. Technical and economic constraints limit the depth of analysis necessary for obtaining a representative picture of microbial and viral communities, their metabolic profiles and their adaptation dynamics (Morgan et al., 2010). Indeed, large-scale sequencing of metagenomic DNA permits the isolation of the most predominant species in the environment, while sequences from low-abundance species may go undetected. In this way, only the most frequently represented functional genes and metabolic pathways that are relevant in a given ecosystem can be identified and assessed. However, the low-abundance species and their encoded functions could also

play a critical role in the ecology and physiology of the studied environment (Piganeau and Moreau, 2007).

In conclusion, metagenomics has shown that the uncultured microbial world far outnumbers the cultured world and has emphasized the extent of our ignorance about the microbial world. Metagenomics has helped elucidate the structure, function, and interactions of microbial communities; these advances were not possible in the culture-dependent era. Metagenomics constitutes a comprehensive approach for understanding evolution and physiology.

ORPHAN GENES

Orphan genes constitute a class of lineage-specific genes that do not show homology to sequences in other species (Fischer and Eisenberg, 1999). They typically encode small proteins and show high non-synonymous substitution rates, but their functions are unknown (Domazet-Loso and Tautz, 2003; Daubin and Ochman, 2004). Recently, a classification of ORFans has been proposed, dividing ORFans into singletons, multipletons, and lineage ORFans (Boyer et al., 2010b). Each newly sequenced genome contains significant numbers of such genes (Toll-Riera et al., 2009). For example, of 60 fully sequenced microbial genomes, 14% of genes are species-specific orphans (Siew and Fischer, 2003), while 18% of genes in *Drosophila* are restricted to the *Drosophila* group (Zhang et al., 2007; Zhou et al., 2008). However, the origin of orphan genes remains a mystery (Merkeev and Mironov, 2008). One proposed scenario is that they derived from gene duplication events in which one copy accumulated so many sequence changes that the ancestral similarity is no longer detectable (Domazet-Loso and Tautz, 2003). It was recently proposed that such ORFans could also represent genes of viral or plasmid origin (Rocha et al., 2006), and some seem to correspond to truly new genes formed *de novo* through diverse mechanisms of gene evolution (Boyer et al., 2010b). This mechanism has been proposed to have made a significant contribution to the formation of novel genes in mammals, specifically in primates, in which 5.5% of orphan genes could have originated *de novo* from non-coding genomic regions (Toll-Riera et al., 2009). The formation of novel genes has also been described in *Drosophila* (Begun et al., 2006; Levine et al., 2006; Zhou et al., 2008) and *Saccharomyces cerevisiae* (Cai et al., 2008).

In a recent study in our laboratory, we identified a small number of gene sequences in *Rickettsia* species that had no match in any database and that seem to have resulted from *de novo* creation (Georgiades et al., 2011). Indeed, 17 rickettsial gene sequences seem to have no homologs in the NR database. The Ka/Ks ratio revealed that 15 of these sequences were either non-functional or had adopted functionality. Of course, the probability of pseudogenization or even of a possible viral origin of these genes should not be excluded, but because these genes were not found in regions with traces of active or ancient integrated extra-chromosomal elements, we strongly believe that they are novel genes (Georgiades et al., 2011).

Finally, it has been reported that new genes might be essential to an organism's viability. In the case of *Drosophila*, 59 *de novo* genes were found to be as essential as the old genes in terms of viability. The observation of lethal phenotypes caused by the knockout of new genes suggested that *de novo*-created genes may integrate a vital pathway by interacting with existing genes, and this co-evolution may lead to the new gene becoming indispensable (Chen et al., 2010).

In summary, gene creation is a continuous and unsettled phenomenon, and this idea is supported by the discovery of new genes, which are permanently generated and whose identification is becoming increasingly frequent (Raoult, 2010a; Boyer et al., 2010b). *De novo*-created genes are evidence of life's permanent creativity.

THE TREE OF LIFE

The TOL was used by Darwin approximately 150 years ago, as a concept to explain the evolutionary relationships between different species (Doolittle, 1999; Lawton, 2009). It has been accepted as a biological fact since (Doolittle and Bapteste, 2007). According to Darwin's theory, namely the "descent with modification theory" (Penny, 2011), the common descent of species is demonstrated by similarities between species, while modifications driven by natural selection create differences in species that result in speciation (Doolittle and Bapteste, 2007). The TOL is therefore composed of a common ancestor, the root of the tree, species separated quickly and in a stable way, key branches, and branches containing the most recently arisen species (Raoult, 2010b,c). However, evidence acquired using comparative genomic analyses contradicts the existence of a single common ancestor for the gene repertoire of any organism. Evidence obtained through genomic analyses suggests that nearly all genes have been exchanged or recombined at some point and that there are no two genes with a similar history on the phylogenetic tree (Raoult, 2010c).

Since the late 1990s, LGT and gene loss in bacterial genomes have been recognized as much more frequent than previously proposed (Ochman et al., 2000; Lawrence, 2005; Dagan and Martin, 2007). Up to 30% of the genome-to-genome variation within a species is the result of LGT and gene loss, and homologous recombination is now thought to be the first cause of sequence divergence in many bacteria (Doolittle and Bapteste, 2007). Thus, LGT had been considered a rare phenomenon in intracellular bacteria (Audic et al., 2007) until the discovery of the mobilome in *Rickettsiae*, suggesting that such events were possible (Merhej and Raoult, 2010). Consequently, several further studies identified candidates for LGT in *Rickettsia* species (Wolf et al., 1999; Ogata et al., 2006; Blanc et al., 2007a,b; Georgiades et al., 2011). Moreover, genetic elements invade and proliferate in rickettsial genomes and eventually integrate genes into their host's chromosomes (Merhej and Raoult, 2010). Analysis of the *R. felis* genome has provided evidence for gene transfers between the chromosome and the *R. felis* plasmid, while the plasmids themselves seem to have been acquired through conjugation (Ogata et al., 2005). The first evidence for LGT in *R. bellii* also indicated the role of amebae in gene exchanges; amebae constitute a melting pot in which species can exchange genetic material (Ogata et al., 2006; Moliner et al., 2010). Indeed, the genome of *R. bellii* contains many genes highly similar to those of intracellular bacteria of amebae, such as *Legionella pneumophila* and *Protochlamydia amoebophila* (Ogata et al., 2006). *L. pneumophila* has developed the ability to infect different species of amebae (Rowbotham, 1980; Fields et al., 2002). A recent study on *L. pneumophila* provided evidence for non-vertical

inheritance: 34–57% of the genome has been involved in recombination events. In this study, LGT events between *Legionella* and all bacterial groups known to be present in amebae were detected (Coscolla et al., 2011). In parallel, other studies have identified eukaryotic-like genes in *Legionella* that are most likely of amebal origin (Lurie-Weinberger et al., 2010; Moliner et al., 2010; Schmitz-Esser et al., 2010). The most plausible scenario for the multiple phylogenetic origins of an important fraction of *Legionella* genes is the exchange of genetic material in the common ameba host.

These lateral transfer events do not always involve whole genes or certain gene functions. The *R. felis* paradigm is the first rickettsial genomic analysis in which random transfers of DNA sequences were found to occur independently of gene functions or sequence lengths (Merhej et al., 2011). The functional vision of genes and sequences often influences scientists' analytical strategies and interpretations. Some bacterial genomes contain up to 40% of genes with no apparent function aside from duplication (selfish genes; Raoult, 2010b). Likewise, random sequences could have hybridized between species because of their sympatric lifestyle (Mayr, 1957).

In light of these post-genomic data, a post-Darwinist concept should be introduced, one that assimilates the chimerism and mosaic structure (**Figure 5**) of all living organisms through both non-vertical inheritance and *de novo* creation (Raoult, 2010c). The TOL is a biblical phrase (Penny, 2011) that matches well the desire to have classification reflecting the "natural order" that is inclusively hierarchical and goes back to a single ancestor (Doolittle, 1999). Our current genomic knowledge no longer matches with Darwin's representation of the TOL. Species evolution looks much more like a rhizome (Deleuze and Guattari, 1976; Raoult, 2010c), reflecting all of the various origins of genomic sequences in each species (Raoult, 2010c). Every living organism has a variety of ancestors; exchanges between species are intense, and the creation of new genes is frequent and constant in all organisms. For example, the human genome is a chimera and viruses and bacterial species are also our ancestors. Retroviruses left relics in our genomes, in the same way that both HHV-6A and B viruses can integrate into human chromosomes and may be vertically transmitted in the germ line (Arbuckle et al., 2010). *Trypanosoma cruzi* sequences were also integrated and identified into the genomes of

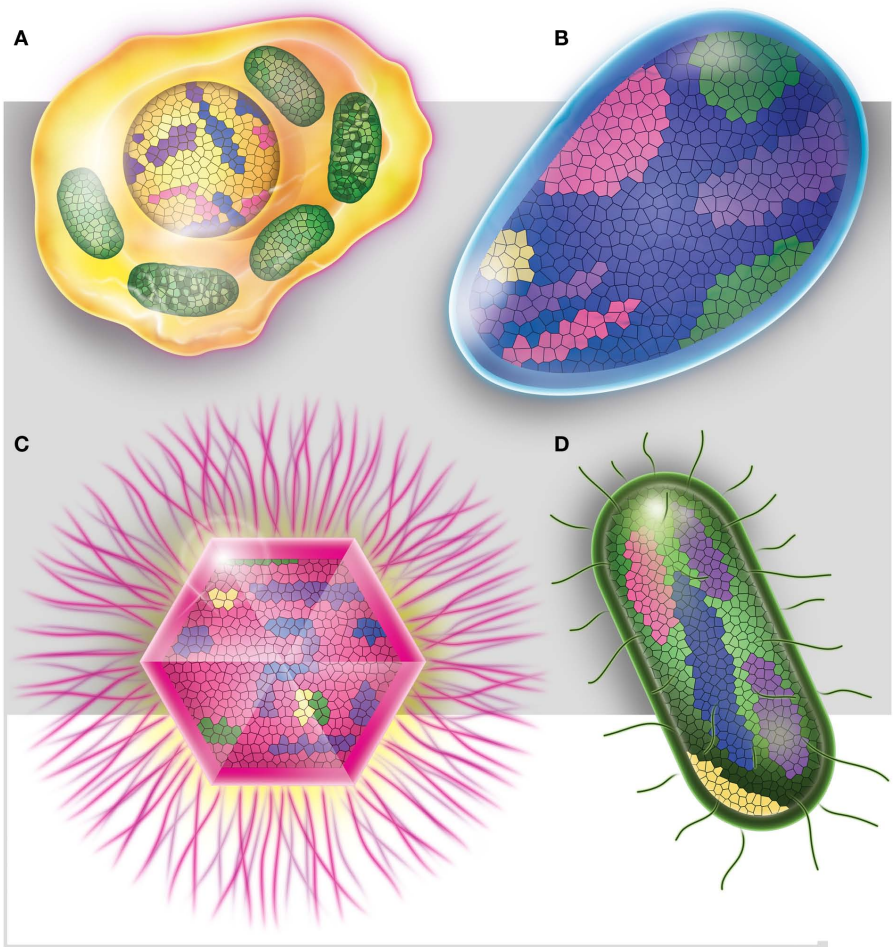


FIGURE 5 | Each one of the four domains of life, (A) Eukaryotes (in yellow), (B) Archaea (in blue), (C) Viruses (in pink), and (D) Bacteria (in green), is represented as mosaics containing genes from all four domains. Purple squares represent ORFan genes.

patients (Hecht et al., 2010). Therefore, the definition of a common ancestor should be revised and instead of referring to a single ancestor, refer to viral ancestors, bacterial ancestors, eukaryotic ancestors, and archaeal ancestors.

CONCLUSION

We think that the radical approach developed by the post-modern French philosophers is useful at this time, as technology has allowed for important discoveries. From this perspective, rickettsiologists, virologists, and bacteriologists, all of whom have different points of view, can make a real contribution to their fields and to the study of the evolution of living organisms. Without the adoption of a non-traditional vision, a large proportion of living organisms, which are now within reach, will remain invisible because we will be trapped by the theories of the twentieth century. Objects are constrained by their definitions. For example, giant viruses were missed by scientists and were not identified earlier because of the misleading definitions of viruses that wanted them to be filterable and smaller than 200 nm (Lwoff, 1957) If the definitions are false, like we demonstrated for the great denominations of microbiology, objects cannot be conceived in a reasonable way and the conclusions derived from the observations of the microorganisms will be biased by misleading believes and theories.

ACKNOWLEDGMENTS

We would like to thank Christelle Forzales for her help with **Figure 5**.

REFERENCES

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., and Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J. Bacteriol.* 186, 2629–2635.

Adékambi, T., Drancourt, M., and Raoult, D. (2008). The rpoB gene as a tool for clinical microbiologists. *Trends Microbiol.* 17, 37–45.

Andersson, J. O., and Andersson, S. G. (1999). Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* 16, 1178–1191.

Andersson, J. O., and Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends Microbiol.* 6, 263–268.

Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C., Podowski, R. M., Näslund, A. K., Eriksson, A. S., Winkler, H. H., and Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140.

Aravalli, R. N., She, Q., and Garrett, R. A. (1998). Archaea and the new age of microorganisms. *Trends Ecol. Evol. (Amst.)* 13, 190–194.

Arbuckle, J. H., Medveczkya, M. M., Lukab, J., Hadleya, S. H., Luegmayrc, A., Ablashid, D., Lunde, T. C., Tolare, J., Meirleirf, K. D., Montoyag, J. G., Komaroffh, A. L., Ambrosc, P. F., and Medveczkya, P. G. (2010). The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5563–5568.

Audic, S., Robert, C., Campagna, B., Parinello, H., Claverie, J. M., Raoult, D., and Drancourt, M. (2007). Genome analysis of *Minibacterium massiliensis* highlights the convergent evolution of water-living bacteria. *PLoS Genet.* 3, e138. doi: 10.1371/journal.pgen.0030138

Bapteste, E., and Boucher, Y. (2008). Lateral gene transfer challenges principles of microbial systematics. *Trends Microbiol.* 16, 200–207.

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupré, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4, 34.

Bechah, Y., El Karkouri, K., Mediannikov, O., Leroy, Q., Pelletier, N., Robert, C., Médigue, C., Mege, J. L., and Raoult, D. (2010). Genomic, proteomic and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. *Genome Res.* 20, 655–663.

Begun, D. J., Lindfors, H. A., Thompson, M. E., and Holloway, A. K. (2006). Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags. *Genetics* 172, 1675–1681.

Belay, N., Johnson, R., Rajagopal, B. S., Conway de, M. E., and Daniels, L. (1988). Methanogenic bacteria from human dental plaque. *Appl. Environ. Microbiol.* 54, 600–603.

Bell, P. J. (2001). Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* 53, 251–256.

Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestris, G., Claverie, J. M., and Raoult, D. (2007a). Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.* 3, e14. doi: 10.1371/journal.pgen.0030014

Blanc, G., Ogata, H., Robert, C., Audic, S., Claverie, J. M., and Raoult, D. (2007b). Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res.* 17, 1657–1664.

Bokum, A. M. C., Movahedzadeh, F., Frita, R., Bancroft, G. J., and Stoker, N. G. (2008). The case for hypervirulence through gene deletion in *Mycobacterium tuberculosis*. *Cell* 16, 436–441.

Boyer, M., Madoui, M. A., Gimenez, G., La Scola, B., and Raoult, D. (2010a), Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLoS ONE* 5, e15530. doi: 10.1371/journal.pone.0015530

Boyer, M., Gimenez, G., Suzan-Monti, M., and Raoult, D. (2010b). Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. *Intervirology* 53, 310–320.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M. G., Suzan-Monti, M., La Scola, B., Koonin, E. V., and Raoult, D. (2009). Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21848–21853.

Bromham, L., and Penny, D. (2003). The modern molecular clock. *Nat. Rev. Genet.* 4, 216–224.

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179, 487–496.

Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* 52, 7–76.

Cavalier-Smith, T. (2004). Only six kingdoms of life. *Proc. Biol. Sci.* 271, 1251–1262.

Cavicchioli, R. (2006). Cold-adapted archaea. *Nat. Rev. Microbiol.* 4, 331–343.

Chatton, E. (1925). *Pansporella perplexa*. Réflexions sur la biologie et la phylogènie des protozoaires. 10e serie. *Ann. Sci. Nat. Zool.* 7, 1–84.

Chen, S., Zhang, Y. E., and Long, M. (2010). New genes in *Drosophila* quickly become essential. *Science* 330, 1682–1685.

Colson, P., and Raoult, D. (2010). Gene repertoire of amoeba-associated giant viruses. *Intervirology* 53, 330–343.

Coscolla, M., Comas, I., and Gonzales-Candelas, F. (2011). Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol. Biol. Evol.* 28, 985–1001.

Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R., and Embley, T. M. (2008). The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20356–20361.

Cummings, C. A., Brinig, M. M., Lepp, P. W., van de Pas, S., and Relman, D. A. (2004). *Bordetella* species are distinguished by patterns of substantial gene loss and gene adaptation. *J. Bacteriol.* 186, 1484–1492.

Dagan, T., and Martin, W. (2007). Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 870–875.

Dahllöf, I., Baillie, H., and Kjelleberg, S. (2000). rpoB-based microbial community analysis avoids limitations inherent in 16s rRNA genes intraspecies heterogeneity. *Appl. Environ. Microbiol.* 66, 3376–3380.

Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J., and Andersson, S. G. (2007). Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* 23, 511–520.

Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 14, 1036–1042.

Deleuze, G., and Guattari, F. (1976). *Rhizome*. Paris: Les éditions de minuit.

DeLong, F. F. (1998). Everything in moderation: archea as "non-extremophiles." *Curr. Opin. Genet. Dev.* 8, 649–654.

Dinsdale, E. A., Edwards, E. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M. A., Nelson, K. E., Nilsson, C., Olson, R., Paul, J., Brito, B. R., Ruan, Y., Swan, B. K., Stevens, R., Valentine, D. L., Thurber, R. V., Wegley, L., White, B. A., and Rohwer, F. (2008). Functional metagenomic profiling of nine biomes. *Nature* 452, 629–632.

Domazet-Loso, T., and Tautz, D. (2003). An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* 13, 2213–2219.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* 284, 2124–2128.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the tree of life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Doolittle, W. F., and Papke, R. T. (2006). Genomics and the bacterial species problem. *Genome Biol.* 7, 116.

Drancourt, M., Berger, P., and Raoult, D. (2004). Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. *J. Clin. Microbiol.* 42, 2197–2202.

Dridi, B., Raoult, D., and Drancourt, M. (2011). Archaea as emerging organisms in complex human microbiomes. *Anaerobe* 17, 56–63.

Dridi, B., Henry, M., El Karkouri, K., Raoult, D., and Drancourt, M. (2009). High prevalence of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* detected in the human gut using an improved DNA detection protocol. *PLoS ONE* 4, e7063. doi: 10.1371/journal.pone.0007063

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.

Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., Saar, M. O., Alexander, S., Alexander, E. C. Jr., and Rohwer, F. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7, 57. doi: 10.1186/1471-2164-7-57

Edwards, R. A., and Rohwer, F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3, 504–510.

Embley, T. M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630.

Fields, B. S., Benson, R. F., and Besser, R. E. (2002). *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin. Microbiol. Rev.* 15, 506–526.

Fischer, D., and Eisenberg, D. (1999). Finding families for genomic ORFans. *Bioinformatics* 15, 759–762.

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5–16.

Fournier, P. E., Elkarkouri, K., Leroy, Q., Robert, C., Guimelli, B., Renesto, P., Socolovschi, C., Parola, P., Audic, S., and Raoult, D. (2009). Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. *BMC Genomics* 10, 166. doi: 10.1186/1471-2164-10-166

Fournier, P. E., and Raoult, D. (2009). Current knowledge on phylogeny and taxonomy of *Rickettsia* spp. *Ann. N. Y. Acad. Sci.* 1166, 1–11.

Fox, G. E., Wisotzkey, J. D., and Jurtshuk, P. (1992). How close is close – 16S ribosomal –RNA sequence identity may no be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., and DeLong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3805–3810.

Fuerst, J. A. (1995). Planctomycetes-a phylum of emerging interest for microbial evolution and ecology. *Microbiology* 141, 1493–1506.

Fuerst, J. A. (2005). Intracellular compartmentation in planctomycetes. *Annu. Rev. Microbiol.* 59, 299–328.

Fuerst, J. A. (2010). Beyond prokaryotes and eukaryotes: planctomycetes and cell organization. *Nat. Educ.* 3, 44.

Georgiades, K., Merhej, V., El Karkouri, K., Raoult, D., and Pontarotti, P. (2011). Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol. Direct* 6, 6.

Georgiades, K., and Raoult, D. (2011a). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151

Georgiades, K., and Raoult, D. (2011b). Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS ONE* 6, e17962. doi: 10.1371/journal.pone.0017962

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355–1359.

Glansdorff, N., Xu, Y., and Labedan, B. (2008). The last common universal ancestor: emergence, constitution and genetic legacy of an elucive forerunner. *Biol. Direct* 3. doi: 10.1186/1745-6150-3-29

Goldberg, M. B., and Theriot, J. A. (1995). *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6572–6576.

Goldberg, M. B., Theriot, J. A., and Sansonetti, P. J. (1994). Regulation of surface presentation of IcsA, a *Shigella* protein essential to intracellular movement and spread, is growth phase dependent. *Infect. Immun.* 62, 5664–5668.

Golding, G. B., and Gupta, R. S. (1995). Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* 12, 1–6.

Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P., and Brochier-Armanet, C. (2010). The origin of eukaryotes and their relationship with the archea: are we at a phylogenomic impasse? *Nat. Rev. Microbiol.* 8, 743–752.

Groussin, M., and Gouy, M. (2011). Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol. Biol. Evol.* 28, 2661–2674.

Gupta, R. S. (1998a). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among alpha proteobacteria, eubacteria and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491.

Gupta, R. S. (1998b). What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol. Microbiol.* 29, 695–707.

Gupta, R. S. (2000). The natural evolutionary relationships among prokaryotes. *Crit. Rev. Microbiol.* 26, 111–131.

Hacker, J., and Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679.

Hale, T. L., Sansonetti, P. J., Schad, P. A., Austin, S., and Formal, S. B. (1983). Characterization of virulence plasmids and plasmid-associated outer membrane proteins in *Shigella flexneri*. *Infect. Immun.* 40, 340–350.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685.

Hecht, M. M., Nitz, N., Araujo, P. F., Sousa, A. O., de Cássia Rosa, A., Gomes, D. A., Leonardecz, D., and Teixeira, A. R. L. (2010). Inheritance of DNA transferred from American trypanosomes to human hosts. *PLoS ONE* 5, e9181. doi: 10.1371/journal.pone.0009181

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C. L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G. J., Droege, M., Frishman, D., Rattei, T., Mewes, H. W., and Wagner, M. (2004). Illuminating the evolutionary history of chlamydiae. *Science* 304, 728–730.

Huber, R., Huber, H., and Setter, K. O. (2000). Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol. Rev.* 24, 615–623.

Hugenholtz, P., Goebel, B. M., and Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 4765–4774.

Iyer, L. M., Blaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplamsic large DNA viruses. *Virus Res.* 117, 156–184.

Johnson, P. L., and Slatkin, M. (2009). Inference of microbial recombination rates from metagenomic data. *PLoS Genet.* 5, e1000674. doi: 10.1371/journal.pgen.1000674

Karaolis, D. K. R., Lan, R. T., and Reeves, P. R. (1994). Sequence variation in *Shigella sonnei* (*Sonnei*), a pathogenic clone of *Escherichia coli*, over 4 continents and 41 years. *J. Clin. Microbiol.* 32, 796–802.

Karner, M. B., DeLong, E. F., and Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific ocean. *Nature* 409, 507–510.

Kleba, B., Clark, T. R., Lutter, E. L., Ellson, D. W., and Hackstadt, T. (2010). Disruption of the *Rickettsia rickettsii* Sca2 autotransporter inhibits actin-based motility. *Infect. Immun.* 78, 2240–2247.

Koonin, E. V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11, 209.

Kuhn, S. T. (1962). *The Structure of Scientific Revolutions.* Chicago, IL: University of Chicago Press.

La Scola, B., Audic, S., Robert, C., Jungang, L., De, L., Drancourt, M., Birtles, R., Claverie, J. M., and Raoult, D. (2003). A giant virus in amoebae. *Science* 299, 2033.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E. V., and Raoult, D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104.

Lake, J. A. (1988). Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* 331, 184–186.

Lake, J. A., and Rivera, M. C. (1994). Was the nucleus the first endosymbiont? *Proc. Natl. Acad. Sci. U.S.A.* 91, 2880–2881.

Lang, F. B., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397.

Lawrence, J. G. (1999). Gene transfer speciation and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2, 519–523.

Lawrence, J. G. (2005). Common themes in the genome strategies of pathogens. *Curr. Opin. Genet. Dev.* 15, 584–588.

Lawton, G. (2009). Why Darwin was wrong about the tree of life. *New Sci.* 2692, 34–39.

Lescot, M., Audic, S., Robert, C., Nguyen, T. T., Blanc, G., Cutler, S. J., Wincker, P., Couloux, A., Claverie, J. M., Raoult, D., and Drancourt, M. (2008). The genome of *Borellia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet.* 4, e1000185. doi: 10.1371/journal.pgen.1000185

Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A., and Begun, D. J. (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9935–9939.

Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444, 1022–1023.

Lurie-Weinberger, M. N., Gomez-Valero, L., Merault, N., Glockner, G., Buchrieser, C., and Gophna, U. (2010). The origins of eukaryotic-like proteins in *Legionella pneumophila*. *Int. J. Med. Microbiol.* 7, 470–481.

Lwoff, A. (1957). The concept of virus. *J. Gen. Microbiol.* 17, 239–253.

Lyotard, J. F. (1979). *La condition postmoderne-Rapport sur le savoir*. Paris: Les éditions de minuit.

Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K., and Fasano, A. (1998). Black holes and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3943–3948.

Mayr, E. (1957). *The Species Problem*. Washington, DC: American Association for the Advancement of Science.

Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric *Rickettsia felis* paradigm demonstrates random transfer of sequences. *Mol. Biol. Evol.* (in press).

Merhej, V., and Raoult, D. (2010). Rickettsial evolution in the light of comparative genomics. *Biol. Rev. Camb. Philos. Soc.* 86, 379–405.

Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009a). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4, 13.

Merhej, V., El Karkouri, K., and Raoult, D. (2009b). Whole genome-based phylogenetic analysis of *Rickettsiae*. *Clin. Micrbiol. Infect.* 2, 336–337.

Merkeev, I. V., and Mironov, A. A. (2008). Orphan genes: function, evolution and composition. *Mol. Biol. (Mosk.)* 42, 127–132.

Miller, T. L., and Wolin, M. J. (1982). Enumeration of *Methanobrevibacter smithii* in human feces. *Arch. Microbiol.* 131, 14–18.

Moliner, C., Fournier, P. E., and Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot for evolution. *FEMS Microbiol. Rev.* 34, 281–294.

Moran, N. A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2873–2878.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586.

Moreira, D., and Brochier-Armanet, C. (2008). Giant viruses, giant chimeras: the multiple evolutionary histories of *Mimivirus* genes. *BMC Evol. Biol.* 8, 12. doi: 10.1186/1471-2148-8-12

Moreira, D., and Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7, 306–311.

Morgan, J. L., Darling, A. E., and Eisen, J. A. (2010). Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS ONE* 5, e10209. doi: 10.1371/journal.pone.0010209

Mounier, J., Ryter, A., Coquis-Rondon, M., and Sansonetti, P. J. (1990). Intracellular and cell-to-cell spread of *Listeria monocytogenes* involves interaction with F-actin in the enterocyte like cell line Caco-2. *Infect. Immun.* 58, 1048–1058.

Nierman, W. C., DeShazer, D., Kim, H. S., Tettelin, H., Nelson, K. E., Feldblyum, T., Ulrich, R. L., Ronning, C. M., Brinkac, L. M., Daugherty, S. C., Davidsen, T. D., Deboy, R. T., Dimitrov, G., Dodson, R. J., Durkin, A. S., Gwinn, M. L., Haft, D. H., Khouri, H., Kolonay, J. F., Madupu, R., Mohammoud, Y., Nelson, W. C., Radune, D., Romero, C. M., Sarria, S., Selengut, J., Shamblin, C., Sullivan, S. A., White, O., Yu, Y., Zafar, N., Zhou, L., and Fraser, C. M. (2004). Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14146–14251.

Ochman, H., Elwyn, S., and Moran, N. A. (1999). Calibrating bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12638–12643.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Ochman, H., Lerat, E., and Daubin, V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6595–6599.

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P. E., Claverie, J. M., and Raoult, D. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* 2, e76. doi: 10.1371/journal.pgen.0020076

Ogata, H., Renesto-Audiffren, P., Audic, S., Robert, C., Blanc, G., Fournier, P. E., Parinello, H., Claverie, J. M., and Raoult, D. (2005). The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* 3, e248. doi: 10.1371/journal.pbio.0030248

Ogata, H., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M., and Raoult, D. (2001). Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* 293, 2093–2098.

Pace, N. R. (2006). Concept time for a change. *Nature* 441, 289.

Paddock, C. D., Sumner, J. W., Comer, J. A., Zari, S. R., Goldsmith, C. S., Goddard, J., McLellan, S. L., Tamminga, C. L., and Ohl, C. A. (2004). *Rickettsia parkeri*: a newly recognized cause of spotted fever rickettsiosis in the United States. *Clin. Infect. Dis.* 38, 805–811.

Penny, D. (2011). Darwin's theory of descent with modification, versus the biblical tree of life. *PLoS Biol.* 9, e1001096. doi: 10.1371/journal.pbio.1001096

Piganeau, G., and Moreau, H. (2007). Screening the Sargasso Sea metagenome for data to investigate genome evolution in *Ostreococcus* (Prasinophyceae, Chlorophyta). *Gene* 406, 184–190.

Poole, A., and Penny, D. (2007). Eukaryote evolution: engulfed by speculation. *Nature* 447, 913.

Popper, K. (1959). *The Logic of Scientific Discovery*, New Edn. London: Taylor & Francis Group.

Pupo, G. M., Lan, R. T., and Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10567–10572.

Raoult, D. (2010a). Technology-driven research will dominate hypothesis-driven research: the future of microbiology. *Future Microbiol.* 5, 135–137.

Raoult, D. (2010b). "L'homme, cette chimère: l'évolution selon Darwin relue et corrigée," in *Dépasser Darwin*, ed. Plos (Paris: Plon V. France), 15–55.

Raoult, D. (2010c). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of *Mimivirus*. *Science* 306, 1344–1350.

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329.

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from *Mimivirus*. *Nat. Rev. Microbiol.* 6, 315–317.

Raoult, D., La Scola, B., and Birtles, R. (2007). The discovery and characterization of *Mimivirus*, the largest known virus and putative pneumonia agent. *Clin. Infect. Dis.* 45, 95–102.

Renvoisé, A., Merhej, V., Georgiades, K., and Raoult, D. (2011). Intracellular Rickettsiales: insights into manipulators of eukaryotic cells. *Trends Mol. Med.* 10, 573–583.

Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38, 525–552.

Rocha, M. T. G., Cooper, J. E., Smith, N. H., and Feil, E. J. (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* 239, 226–235.

Romano, A. H., and Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* 147, 448–455.

Rosello-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67.

Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. *Nature* 409, 1092–1101.

Roux, V., Drancourt, M., Stein, A., Riegel, P., Raoult, D., and La Scola, B. (2004). *Corynebacterium* species isolated from bone and joint infection identified by 16 rRNA gene sequence analysis. *J. Clin. Microbiol.* 42, 2231–2233.

Roux, V., and Raoult, D. (1997). Rickettsioses as paradigms of new emerging infectious diseases. *Clin. Microbiol. Rev.* 10, 694–719.

Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J. Clin. Pathol.* 33, 1179–1183.

Sanogo, Y. O., Zeaiter, Z., Caruso, G., Merola, F., Shpynov, S., Brouqui, P., and Raoult, D. (2003). *Bartonella henselae* in *Ixodes ricinus* ticks (Acari: Ixodida) removed from humans, Bellulo province, Italy. *Emerg. Infect. Dis.* 9, 329–332.

Sapp, J. (2005). The prokaryotic-eukaryotic dichotomy: meanings and mythology. *Microbiol. Mol. Biol. Rev.* 69, 292–305.

Schirawski, J., Mannhaupt, G., and Münch, K. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330, 1546–1548.

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005). Genomic studies of uncultivated Archaea. *Nat. Rev. Microbiol.* 3, 479–488.

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and Horn, M. (2010). The genome of the amoeba symbionts "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J. Bacteriol.* 192, 1045–1057.

Siew, N., and Fischer, D. (2003). Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* 53, 241–251.

Simon, C., Wiezer, A., Strittmatter, A. W., and Daniel, R. (2009). Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* 75, 7519–7526.

Stackebrandt, E., and Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* 33, 152–155.

Stanier, R. Y., and van Niel, C. B. (1962). The concept of a bacterium. *Arch. Microbiol.* 42, 17–35.

Takemura, M. (2001). Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* 52, 419–425.

Taylor, F. J. R. (1976). Autogenous theories for the origin of eukaryotes. *Taxon* 25, 377–390.

Teysseire, N., Chiche-Portiche, C., and Raoult, D. (1992). Intracellular movements in *Rickettsia conorii* and *R. typhi* based on actin polymerization. *Res. Microbiol.* 143, 821–829.

Tilney, L. G., and Portnoy, D. A. (1989). Actin filaments and the growth, movement and spread of the intracellular bacterial parasite *Listeria monocytogenes*. *J. Cell Biol.* 109, 1597–1608.

Toll-Riera, M., Bosch, N., Bellora, N., Castelo, R., Armengol, L., Estivill, X., and Alba, M. M. (2009). Origin of primate orphan genes: a comparative genomics approach. *Mol. Biol. Evol.* 26, 603–612.

Trevors, J. T., and Abel, D. L. (2004). Chance and necessity do not explain the origin of life. *Cell Biol. Int.* 28, 729–739.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43.

Vallery-Radot, P. (1885). *Lois Pasteur: Life and Labours, By His Son in-Law. Tr, From the French By Lady Claud Hamilton.* London: Longmans, Green and Co.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004). Environmental genomes shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.

Viezens, J., and Arvand, M. (2008). Simultaneous presence of two different copies of the 16S rRNA gene in *Bartonella henselae*. *Microbiology* 154, 2881–2886.

Villarreal, L. P., and De Filippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* 74, 7079–7084.

Ward, N., Rainey, F., Hedlund, B., Staley, J., Ludwig, W., and Stackebrandt, E. (2000). Comparative phylogenetic analyses of members of the order Planctomycetales and the division Verrucomicrobia: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. *Int. J. Syst. Evol. Microbiol.* 50, 1965–1972.

Wicks, R. (2003). *Modern French Philosophy: From Existentialism to Postmodernism.* Oxford: Oneworld Publications.

Williams, J. (1998). *Lyotard: Towards a Postmodern Philosophy*. Cambridge: Polity Press.

Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F. E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., and Rohwer, F. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* 4, e7370. doi: 10.1371/journal.pone.0007370

Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., van Soolingen, D., Rüsch-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., and Niemann, S. (2008). Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog.* 4, e1000160. doi: 10.1371/journal.ppat.1000160

Woese, C. R. (1994). There must be a prokaryote somewhere-microbiology search for itself. *Microbiol. Rev.* 58, 1–9.

Woese, C. R. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6854–6859.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090.

Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, K. B. J, and Stahl, D. (1975). Conservation of primary structure in 16 ribosomal-RNA. *Nature* 254, 83–86.

Wolf, Y. I., Aravind, L., and Koonin, E. V. (1999). *Rickettsiae* and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* 15, 173–175.

Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E. J., Woodnutt, G., Short, J. M., and Keller, M. (2005). High-throughput cultivation of microorganisms using microcapsules. *Meth. Enzymol.* 397, 124–130.

Zhang, G., Wang, H., Shi, J., Wang, X., Zheng, H., Wong, G. K., Clark, T., Wang, W., Wang, J., and Kang, L. (2007). Identification and characterization of insect-specific proteins by genome protein data analysis. *BMC Genomics* 8, 93. doi: 10.1186/1471-2164-8-93

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in *Drosophila*. *Genome Res.* 18, 1446–1455.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 July 2011; accepted: 10 October 2011; published online: 31 October 2011.

Citation: Georgiades K, Merhej V and Raoult D (2011) The influence of rickettsiologists on post-modern microbiology. Front. Cell. Inf. Microbio. ***1****:8. doi: 10.3389/fcimb.2011.00008*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

PERSPECTIVE ARTICLE
published: 06 September 2012
doi: 10.3389/fcimb.2012.00116

The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories

Bo Segerman*

National Veterinary Institute, Uppsala, Sweden

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Kalai Mathee, Florida International University, USA
Nick S. Jakubovics, Newcastle University, UK
Jean-Marc Rolain, URMITE CNRS-IRD UMR 6236, France

***Correspondence:**
Bo Segerman, National Veterinary Institute SVA, SE-751 89 Uppsala, Sweden.
e-mail: bo.segerman@sva.se

Strains within a bacterial species typically have a set of conserved core genes and a variable set of accessory genes. The accessory genes often appear to move laterally between strains, thereby forming new trait combinations. Sometimes, genetic material also moves laterally between species, thereby resulting in diffuse borders between them. The growing number of genome sequences offers new possibilities to study these processes. Ten species for which abundant genomic data exists were here selected for analysis of the species border integrity. The average core genome similarities and relative core genome sizes (RCGSs) were determined for strain pairs within the species and for strain pairs crossing the species border. The variability within the species as well as the border integrity varies for different bacterial species. Some have very distinct borders while others are more or less indefinable. From the growing amount of genomic data, it becomes even clearer that the concept of bacterial species is, in many cases, far from absolute.

Keywords: accessory genome, bacterial species, core genome, lateral gene transfer, species border

INTRODUCTION

In higher eukaryotes, a species is often defined as a group of organisms that are so reproductively isolated that interbreeding with other species cannot occur or does not result in a fertile offspring. This is believed to maintain the genetic integrity of the species over time. A genetic pool that is much larger than that present in each individual is maintained within the species and sexual reproduction accounts for the formation of new allele combinations. Thus, the genetic material is inherited vertically from the combined genetic pool of the parents and the apparent universality of this inheritance mode led Darwin to propose that all organisms could be organized into a "tree of life."

In prokaryotes, the situation is somewhat different. Their genetic material is asexually inherited from the ancestral cell. Accumulation of mutations during this clonal expansion can give rise to sub-populations with selective advantages. If prokaryotes were to rely on only this mechanism for adaptation, new trait combinations would require "reinventing the wheel" over and over. Thus, it is not surprising that lateral gene transfer (LGT) mechanisms exist in prokaryotes (Ochman et al., 2000). LGT allows advantageous genes to sweep through populations (Shapiro et al., 2012). As a consequence of the lateral movements of genetic material, the organisms becomes chimerical and a strict tree model cannot adequately represent their phylogenetic relationships (Bapteste et al., 2009). To reflect this, the phrase "rhizome of life" is sometimes used as an alternative to "tree of life" (Raoult, 2010). It is clear that different models to represent this complex evolutionary history are necessary, depending on what scientific questions we are addressing, i.e., we need to use "pattern pluralism" in our way of thinking (Doolittle and Bapteste, 2007). However, the relative extent to which the tree model should, or must, be discarded in favor of alternative models is still debated and varies depending on the species in question. A further consequence of LGT is that the actual concept of the bacterial species becomes partially undermined (Doolittle and Papke, 2006). A broader viewpoint for describing bacterial population structures constitutes the presence of sympatric species complexes with high plasticity and lateral gene exchange from which specialized allopatric species, such as pathogens, can escape (Georgiades and Raoult, 2010). Reductive evolution accompanying the pathogenic lifestyle will, in many cases, confine the species (Merhej et al., 2009).

Recently, a clearer understanding of bacterial genome evolution has emerged. This is mainly because of the intense technological development of high-throughput sequencing (Metzker, 2010). Sequencing is now done in enormous amounts of randomly primed parallel sequencing reactions from the same sample. The technologies are often collected under the name Next Generation Sequencing (NGS). NGS has unquestionably had an enormous impact on the growth rate of the bacterial genome database (**Figure 1A**). As parallel sequencing machines are moving from the core facilities into regular laboratories, it is likely that the number of sequences will continue to increase exponentially. In the early days of bacterial genome sequencing, most projects typically aimed to produce a complete genome sequence as the final product. Making a complete genome sequence still requires a lot of resources. In contrast, draft sequences have become easy and affordable to produce and are often sufficient for answering many biological questions. Consequently, the draft genome database is growing more rapidly than the complete one. At the time of the analysis presented in this paper, there were over 2000 completed genomes and almost 3000 draft genomes in the form of

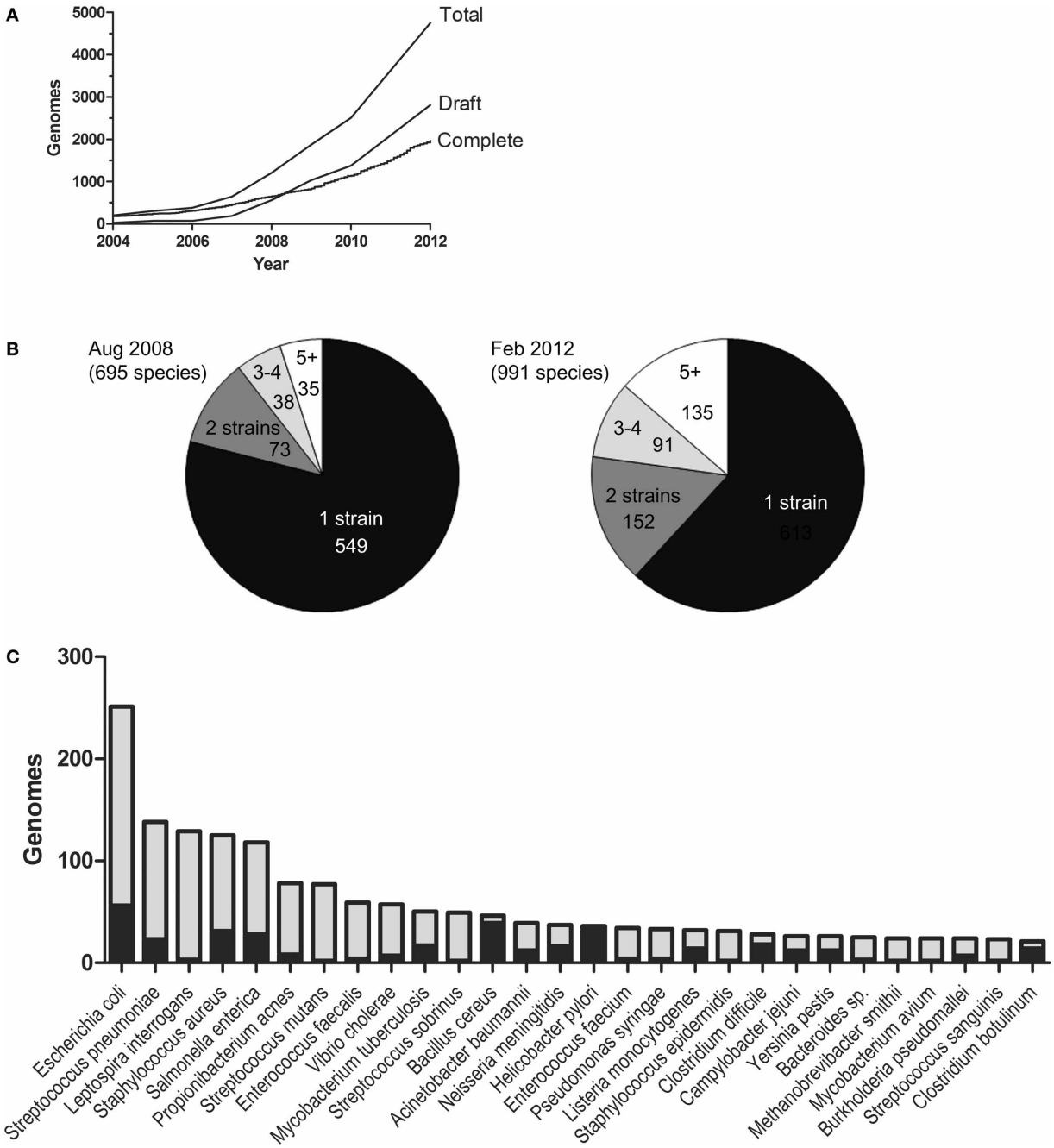


FIGURE 1 | The genome database. The figure is based on the NCBI GenomeReport dated 21 Feb, 2012 and only includes genomes projects where completed sequences or draft assemblies were available for download. **(A)** The growth of the databases for draft and complete bacterial genomes. **(B)** The number of species with 1, 2, 3, 4, and 5 (or more) strains sequenced Aug 2008 and Feb. 2012, respectively. **(C)** The species with most genome sequences. Black represents completed genomes and grey draft genomes.

contigs/scaffolds. There are also a large number of draft genome assemblies not yet submitted to the database.

How are the current sequencing efforts being directed? **Figure 1B** shows the number of species in the genome database in August 2008 and February 2012. The number of species has grown from approximately 700 to almost 1000 (a 1.4-fold increase) and the number of sequences from 1255 to ∼4900 (a 3.9-fold increase). This illustrates that sequencing activities aimed at producing more sequences from strains belonging to already represented species are far more intense than sequencing projects directed towards new species. The ten most sequenced species account for 22% of the genome database. The bacterial species for which the highest number of genome sequences are available (at the time of this writing) are shown in **Figure 1C**.

The large number of available genomic sequences has given us the opportunity to compare genomic variation within species as well as between them. However, sequencing efforts are undoubtedly biased towards strains of medical or economical importance and this may very well bias the conclusions we make.

The genomic sequencing efforts have made us realize that, similar to eukaryotes, bacterial species maintain a "genetic pool" much larger than the one present in each strain. Each strain has a conserved set of core genes and additionally, a number of accessory genes. The dynamics of the accessory genes can give rise to strains with "customized genomic repertoires" (Mathee et al., 2008). Thus, different strains may have different sets of accessory genes and the superset of all different genes present in a species is often referred to as the pan-genome (Tettelin et al., 2005). It is likely that the accessory genes constitute a reservoir for functionality that can be transferred laterally to create new trait combinations. However, a large fraction of the accessory genes have often no functional annotation; our knowledge about many of these genes is poor. The accessory genome has become an important field of study (Sim et al., 2008; Bennett et al., 2010; Kung et al., 2010).

In this study, a limited dataset of species representing different life strategies and having a high number of both draft and complete genome sequences was selected for a more detailed analysis (**Table 1**). The genomic information was here used to study the genetic integrity of these bacterial species in terms of core genome sequence variability and variations in the relative size of the core genome/accessory genome.

COMPARATIVE ANALYSIS

The degree of sequence similarity within the core genome is considered to be one of the best phylogenomic measures for comparing microbial genomes (Rokas et al., 2003). By averaging comparisons of a large number of genes, the risk for disturbances in the result caused by laterally moved genetic material is minimized. In this study, pairwise average core genome similarity (ACGS) values and relative core genome size (RCGS) values were calculated using the Gegenees fragmented alignment method (Ågren et al., 2012). In brief, the genomes were fragmented into overlapping 200-basepair pieces and for each fragment a BLASTN score was calculated. The core genome was defined as the regions constituted by fragments with scores of at least 25% of the score value of a perfect match. The average similarity was normalized towards the value obtained when the genome was compared to itself. The RCGS value represents the core genome size relative to the whole genome size (core genome + accessory genome).

If genetic material were only inherited vertically, the differences in size of the accessory genetic material between two strains would depend on gene loss events and duplication events followed by acquisition of new functionality (neo-functionalization or sub-functionalization). It would then be expected that RCGS would gradually decrease as ACGS decrease. However, if lateral movement of accessory genetic material occur, a much greater variation in RCGS values would be expected because lateral movements would be more or less uncoupled with the vertical inheritance.

In this study, the relationship between ACGS and RCGS values and the integrity of the species border in terms of these values were examined by analysis of a large number of pairwise comparisons within a selected set of bacterial genera. Ten species, with good representation in the sequence database (**Table 1**), were selected. The ACGS and RCGS values were then calculated, pairwise, for every possible genome combination in the genus. Thus, there were both intra- and inter-species comparisons. The data were used to create a diagram with the highest ACGS (shortest vertical inheritance distance) first, and then the pairwise comparisons plotted in descending order. In the same diagram, the corresponding RCGS values were plotted and the part of the diagram that represented intra-species comparisons was indicated. Hereafter, this diagram type is referred to as a "species integrity diagram." The software Gegenees (Ågren et al., 2012) can, from version 1.1.5, generate this type of diagram and also gives interactive annotations of individual data points.

Table 1 | Species included in the analysis.

Genus	Species analyzed	Number of species represented in the genome sequence database
Staphylococcus	*S. aureus*	12
Streptococcus	*S. pneumoniae*	37
Escherichia	*E. coli*	3
Salmonella	*S. enterica*	2
Mycobacterium	*M. tuberculosis*	17
Neisseria	*N. meningitidis*	15
Helicobacter	*H. pylori*	12
Clostridium	*C. botulinum*	14
Bacillus	*B. anthracis*	20
Burkholderia	*B. pseudomallei*	17

SPECIES INTEGRITY OF TEN SELECTED SPECIES

In the intra-species pairwise comparison of *Staphylococcus aureus* strains, ACGS values varied between 90% and 100% (**Figure 2A**). The RCGS values fluctuated between 90–99% (i.e., up to 10% accessory genetic material) with modest correlation with the ACGS values. This indicates that the accessory genetic material is mobile between strains within the species border. Lateral transfer of genetic material in *S. aureus* has been described previously in relation to mobile genetic elements (Deurenberg and Stobberingh, 2008; Lindsay, 2010). The closest strain in the inter-species comparison had distinctly lower ACGS and RCGS values. Thus, on the basis of the genomic data available today, there seems to be a distinct border between the species *S. aureus* and its closest neighbors.

In *Streptococcus pneumoniae*, the intra-species comparison results were very similar to those found in *S. aureus* (**Figure 2B**). Lateral transfer within *S. pneumoniae* has been described previously (Coffey et al., 1991). Inter-species comparisons show that the distance between *S. pneumoniae* and its closest related strains outside the species border (belonging to *S. mitis*) is quite small and there is a continuous decline in similarity with not fully distinct plateaus. This indicates that *S. pneumoniae* is exchanging

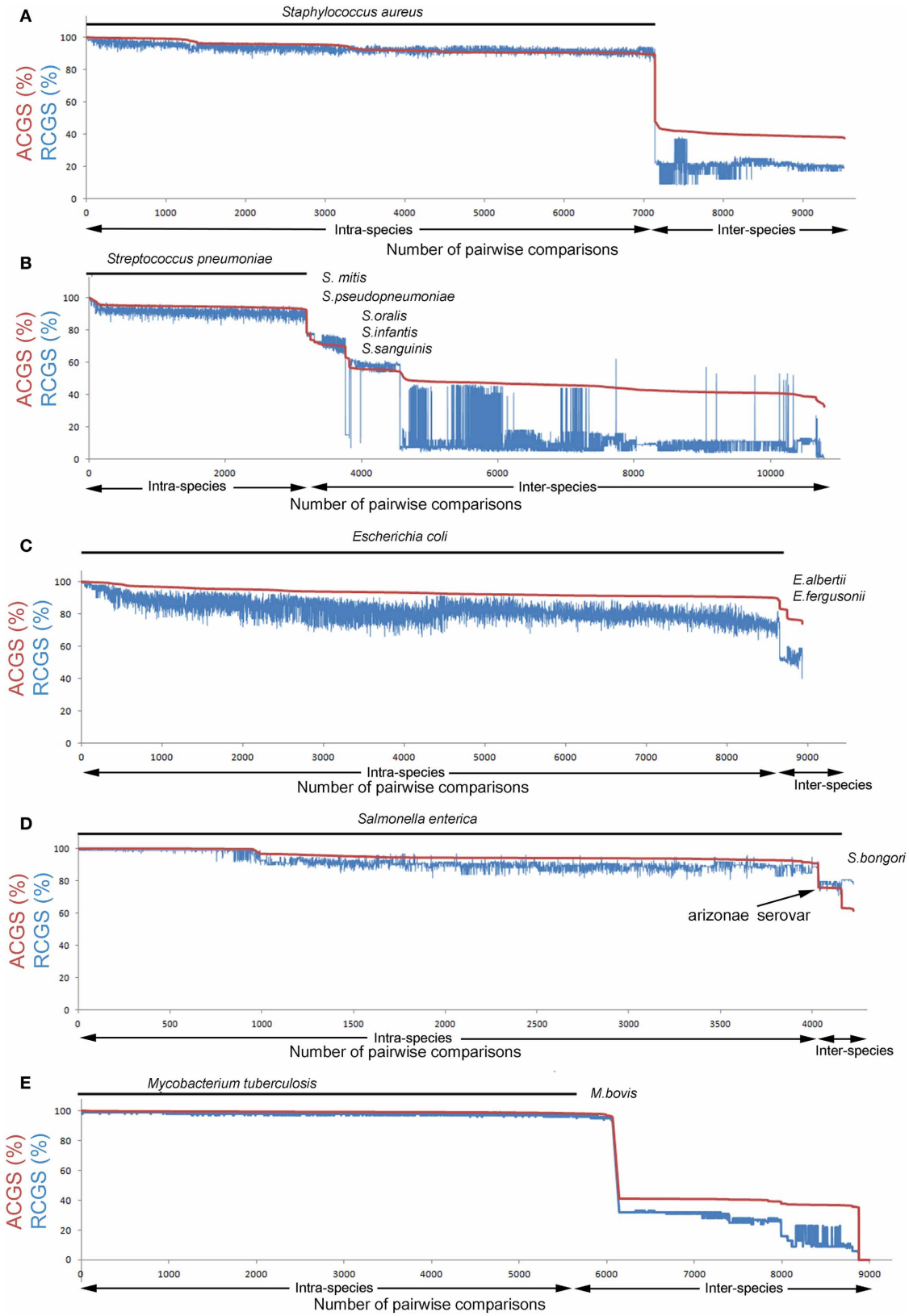


FIGURE 2 | Continued

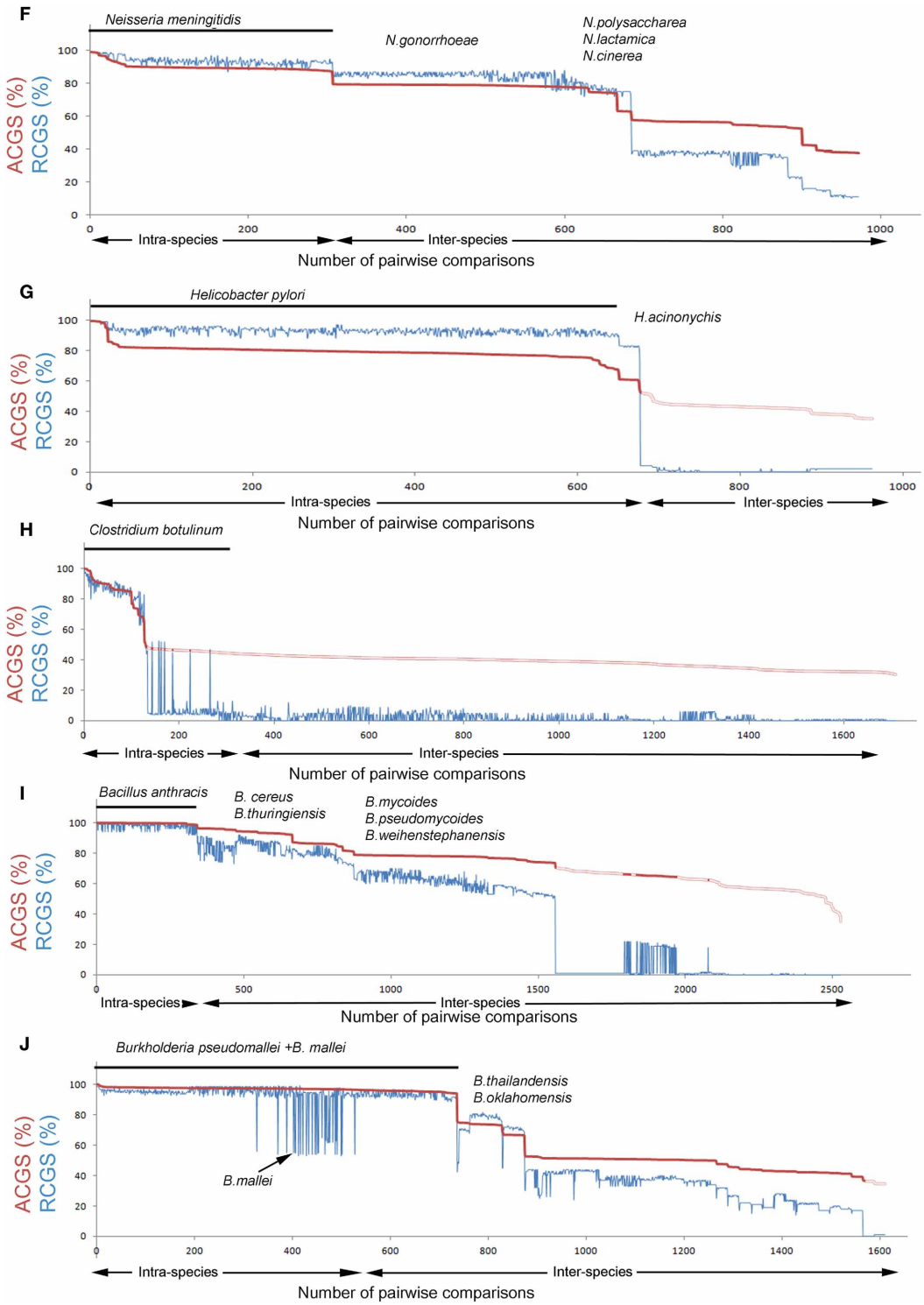


FIGURE 2 | Species integrity diagrams for 10 selected species. Pairwise genome comparisons defining average core genome similarity (ACGS) are shown in red and relative core genome size (RCGS) in blue. ACGS measures phylogenomic distance while RCGS measures the size of the core genome relative to the total genome size (core + accessory genome). Intra-species comparisons are plotted first (indicated with the black line) and inter-species comparisons thereafter. Data are sorted by ASCG. **(A–J)** Intra-species comparisons of strains from the selected species [*Staphylococcus aureus, Streptococcus pneumoniae, Escherichia coli, Salmonella enterica, Mycobacterium tuberculosis, Neisseria meningitidis, Helicobacter pylori, Clostridium botulinum, Bacillus anthracis* (a part of the *Bacillus cereus* group), and *Burkholderia pseudomallei*] and inter-species comparison with strains from other species in the same genus.

genes with other *Streptococcus* spp. The population structure and dynamics of *S. pneumoniae* in terms of the pan genome has been studied previously (Donati et al., 2010).

In the genus *Escherichia*, almost all sequences came from *E. coli*. The high amount of accessory genetic material and the large RCGS fluctuations between strain pairs with similar ACGS values (**Figure 2C**) suggest that there is an extraordinary mobility of accessory genetic material in *E. coli*. The border with other *Escherichia*. spp is not distinct although the sequencing database is underrepresented for non *E. coli* species. Like *Escherichia*, the related *Salmonella* genome database is dominated by a single species, *S. enterica*. *S. enterica* strains typically have an ACGS value of ∼55% compared to *E. coli* strains. The diagram shows that *S. enterica strains* have a lower variability in accessory genome size compared to *E. coli* (**Figure 2D**).

In contrast to the examples above, *Mycobacterium tuberculosis* is very distinct (**Figure 2E**). Low variability was seen for both for RCGS and ACGS values as was a clear difference to related, non-*M. tuberculosis* strains (except for *M. bovis* as discussed below). This indicates lateral gene movements are less frequent in this species. However, another type of problem in species designation becomes apparent here. *M. bovis* is, on a genomic level, an indistinguishable part of *M. tuberculosis*. Detailed genome analysis of a *M. bovis* strain has also shown that there is a very high genomic similarity to *M. tuberculosis* (Garnier et al., 2003).

Neisseria meningitidis is distinct but closely related to strains outside the species (**Figure 2F**). The closest strains belonged to the species *N. gonorrhoeae*. Many *N. meningitidis* strains showed only 10% lower RCGS values when compared to *N. gonorrhoeae* strains than when compared to the other strains within the species. This suggests crossover of genetic material can occur over the species border; nevertheless the ACGS values show clear distinct plateaus. The accessory genome has been observed to evolve differently from the core genome in *Neisseria* (Bennett et al., 2010).

In *Helicobacter pylori*, the ACGS values were, in most intra-species comparisons, relatively low (**Figure 2G**). This suggests that *Helicobacter* has a comparatively high mutation rate. This has also been discussed previously (Wang et al., 1999). The RCGS values fluctuate, as described in many species above, indicating the lateral movement of genes.

Clostridium botulinum is a classical example of how long distance lateral gene movements can affect the integrity of what we call a bacterial species. The Botulinum Neuroxin (BoNT) gene, *bont*, has during several occasions jumped between quite distantly related *Clostridium* strains and this can be seen in the species integrity diagram (**Figure 2H**). *C. botulinum* can actually be seen as four distinct species that all are able to produce BoNT (Hill et al., 2007; Skarin and Segerman, 2011; Skarin et al., 2011). Some strains without a functional *bont* gene can, from a genomic point of view, be considered to be the same species as *C. botulinum* but they go under other names (e.g., *C. sporogenes, C. novyi)*. Furthermore, the *bont* gene can also be found in some *C. baratii* strains and some *C. butyricum* strains.

Bacillus anthracis is an example of a species that represents a monophyletic clade within a larger group of related strains (**Figure 2I**). *B. anthracis* strains all have two virulence plasmids. Apart from the plasmids, they are extremely similar to other related strains (Kolsto et al., 2009). There is a very diffuse boarder between the species in a large group of *Bacillus* strains containing *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides, B. weihenstephanensis*, and *B. pseudomycoides*. This group is commonly called the *B. cereus* group (Rasko et al., 2005). Many *B. cereus* strains are much more closely related to *B. thuringiensis* strains than to other *B. cereus* strains.

Finally *Burkholderia pseudomallei* was analyzed (**Figure 2J**). The most striking observation is its relation to *B. mallei. B. mallei* is a lineage of *B. pseudomallei*, that has undergone massive reductive evolution (Losada et al., 2010). Hence, the pronounced drop in the RCGS values without a corresponding drop in ACGS values.

CONCLUSIONS

When comparing different bacterial species from a genomic perspective, large differences can be found in what we call a species. A large part of the sequencing efforts today are focused on important human pathogens. The niches these species colonize are quite different from those for most environmental species. During their speciation, pathogens have generally shifted from a sympatric to an allopatric lifestyle and this is usually accompanied by reductive evolution and reduced pan genome size. Most of these pathogenic species are more or less distinct, based on core genome conservation, although lateral gene movement channels probably are quite common. In environmental species, however, we see a much larger variability and more poorly defined borders between the species. However, more data are needed on environmental species.

One extreme species is *B. anthracis*. It resides as a spore, without growth. When it does grow, it does so mainly without contact with other bacterial species before returning to the dormant spore state. Thus, it has little opportunity for contact with other bacterial species and hence low genetic variability. On the other extreme is *E coli*. This species has an extraordinarily large and variable set of accessory genes. It also lives in an environment where it is surrounded by a community of different, competing bacteria. Recently, we saw an example of an *E. coli* strain that took up new genetic material by lateral transfer and thereby transformed in to a new, highly virulent variant (Mellmann et al., 2011; Rasko et al., 2011; Rohde et al., 2011).

Other species are located in between these extremes. They may be very close to neighboring species, but are typically still more or less distinguishable based on core sequence similarity. They have specialized in separate, environmental niches. As more strains are sequenced, we will probably find more intermediate forms of these species, showing that the bacterial species concept does not have an absolute definition. However, some draft genomes are of poor quality or are incorrectly annotated and must be treated with caution in any analysis of them.

How should we then relate to bacterial species? For most purposes, species of importance for humans are definable based on the core genome similarity. This is basically what we measure with the classical 16S analysis. However, we must be aware that the species borders are often (perhaps almost always) not absolute, and there will be intermediate strains occurring every now and then. Finally, it is definitely going to be equally important to classify strains according to their accessory gene content. This type of analysis will become much more efficient with the new upcoming sequencing technologies. We are standing at the beginning of a period with an enormous genome database growth and the possibility to greatly increase our understanding of what defines a bacterial species as such.

REFERENCES

Ågren, J., Sundström, A., Håfström, T., and Segerman, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS ONE* 7:e39107. doi: 10.1371/journal.pone.0039107

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupre, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4, 34.

Bennett, J. S., Bentley, S. D., Vernikos, G. S., Quail, M. A., Cherevach, I., White, B., Parkhill, J., and Maiden, M. C. (2010). Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020–006. *BMC Genomics* 11, 652.

Coffey, T. J., Dowson, C. G., Daniels, M., Zhou, J., Martin, C., Spratt, B. G., and Musser, J. M. (1991). Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Mol. Microbiol.* 5, 2255–2260.

Deurenberg, R. H., and Stobberingh, E. E. (2008). The evolution of *Staphylococcus aureus*. *Infect. Genet. Evol.* 8, 747–763.

Donati, C., Hiller, N. L., Tettelin, H., Muzzi, A., Croucher, N. J., Angiuoli, S. V., Oggioni, M., Dunning Hotopp, J. C., Hu, F. Z., Riley, D. R., Covacci, A., Mitchell, T. J., Bentley, S. D., Kilian, M., Ehrlich, G. D., Rappuoli, R., Moxon, E. R., and Masignani, V. (2010). Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol.* 11, R107.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Doolittle, W. F., and Papke, R. T. (2006). Genomics and the bacterial species problem. *Genome Biol.* 7, 116.

Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V., and Hewinson, R. G. (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7877–7882.

Georgiades, K., and Raoult, D. (2010). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151

Hill, K. K., Smith, T. J., Helma, C. H., Ticknor, L. O., Foley, B. T., Svensson, R. T., Brown, J. L., Johnson, E. A., Smith, L. A., Okinaka, R. T., Jackson, P. J., and Marks, J. D. (2007). Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J. Bacteriol.* 189, 818–832.

Kolsto, A. B., Tourasse, N. J., and Okstad, O. A. (2009). What sets *Bacillus anthracis* apart from other Bacillus species? *Annu. Rev. Microbiol.* 63, 451–476.

Kung, V. L., Ozer, E. A., and Hauser, A. R. (2010). The accessory genome of *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* 74, 621–641.

Lindsay, J. A. (2010). Genomic variation and evolution of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 300, 98–103.

Losada, L., Ronning, C. M., Deshazer, D., Woods, D., Fedorova, N., Kim, H. S., Shabalina, S. A., Pearson, T. R., Brinkac, L., Tan, P., Nandi, T., Crabtree, J., Badger, J., Beckstrom-Sternberg, S., Saqib, M., Schutzer, S. E., Keim, P., and Nierman, W. C. (2010). Continuing evolution of *Burkholderia mallei* through genome reduction and large-scale rearrangements. *Genome Biol. Evol.* 2, 102–116.

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J. M., Koehrsen, M., Rokas, A., Yandava, C. N., Engels, R., Zeng, E., Olavarietta, R., Doud, M., Smith, R. S., Montgomery, P., White, J. R., Godfrey, P. A., Kodira, C., Birren, B., Galagan, J. E., and Lory, S. (2008). Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3100–3105.

Mellmann, A., Harmsen, D., Cummings, C. A., Zentz, E. B., Leopold, S. R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S. F., Henkhaus, J. K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P. M., Moore, R. L., Guenther, S., Rothberg, J. M., and Karch, H. (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104, H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* 6:e22751. doi: 10.1371/journal.pone.0022751

Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4, 13.

Metzker, M. L. (2010). Sequencing technologies–the next generation. *Nat. Rev. Genet.* 11, 31–46.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Rasko, D. A., Altherr, M. R., Han, C. S., and Ravel, J. (2005). Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev.* 29, 303–329.

Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E. E., Sebra, R., Chin, C. S., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A. O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J. C., Steyert, S. R., Frimodt-Moller, J., Struve, C., Petersen, A. M., Krogfelt, K. A., Nataro, J. P., Schadt, E. E., and Waldor, M. K. (2011). Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J. Med.* 365, 709–717.

Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., Xi, F., Li, S., Li, Y., Zhang, Z., Yang, X., Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., Kobbe, R., Loos, S., Oh, J., Yang, L., Danchin, A., Gao, G. F., Song, Y., Yang, H., Wang, J., Xu, J., Pallen, M. J., Aepfelbacher, M., and Yang, R. (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104, H4. *N. Engl. J. Med.* 365, 718–724.

Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabo, G., Polz, M. F., and Alm, E. J. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51.

Sim, S. H., Yu, Y., Lin, C. H., Karuturi, R. K., Wuthiekanun, V., Tuanyok, A., Chua, H. H., Ong, C., Paramalingam, S. S., Tan, G., Tang, L., Lau, G., Ooi, E. E., Woods, D., Feil, E., Peacock, S. J., and Tan, P. (2008). The core and accessory genomes of *Burkholderia pseudomallei:* implications for human melioidosis. *PLoS Pathog.* 4:e1000178. doi: 10.1371/journal.ppat.1000178

Skarin, H., Hafstrom, T., Westerberg, J., and Segerman, B. (2011). *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. *BMC Genomics* 12, 185.

Skarin, H., and Segerman, B. (2011). Horizontal gene transfer of toxin genes in *Clostridium botulinum*: involvement of mobile elements and plasmids. *Mob. Genet. Elements* 1, 213–215.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M.,

Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., Madupu, R., Brinkac, L. M., Dodson, R. J., Rosovitz, M. J., Sullivan, S. A., Daugherty, S. C., Haft, D. H., Selengut, J., Gwinn, M. L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'connor, K. J., Smith, S., Utterback, T. R., White, O., Rubens, C. E., Grandi, G., Madoff, L. C., Kasper, D. L., Telford, J. L., Wessels, M. R., Rappuoli, R., and Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955.

Wang, G., Humayun, M. Z., and Taylor, D. E. (1999). Mutation as an origin of genetic variability in *Helicobacter pylori*. *Trends Microbiol.* 7, 488–493.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2012; paper pending published: 04 May 2012; accepted: 20 August 2012; published online: 06 September 2012.

Citation: Segerman B (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. Front. Cell. Inf. Microbio. 2:116. doi: 10.3389/fcimb.2012.00116

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 July 2012
doi: 10.3389/fcimb.2012.00091

Lamarckian evolution of the giant Mimivirus in allopatric laboratory culture on amoebae

Philippe Colson*[1,2] *and Didier Raoult*[1,2]

[1] *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Centre National de la Recherche Scientifique Unité Mixte de Recherche (UMR) 7278, Institut de Recherche pour le Développement (IRD) 3R198, INSERM U1095, IHU Méditerranée Infection, Facultés de Médecine et de Pharmacie, Aix-Marseille University, Marseille, France*

[2] *Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Marseille, France*

Edited by:
Eugene V. Koonin, National Center for Biotechnology Information, USA

Reviewed by:
Jon Audia, University of South Alabama School of Medicine, USA
Hua Xie, Meharry Medical College, USA

***Correspondence:**
Didier Raoult, Unité des Rickettsies, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Faculté de Médecine, Aix-Marseille Université, 27 boulevard Jean Moulin, 13385 Marseille CEDEX 05, France.
e-mail: didier.raoult@gmail.com

Acanthamoeba polyphaga Mimivirus has been subcultured 150 times on germ-free amoebae. This allopatric niche is very different from that found in the natural environment, where the virus is in competition with many other organisms. In this experiment, substantial gene variability and loss occurred concurrently with the emergence of phenotypically different viruses. We sought to quantify the respective roles of Lamarckian and Darwinian evolution during this experiment. We postulated that the Mimivirus genes that were down-regulated at the beginning of the allopatric laboratory culture and inactivated after 150 passages experienced Lamarckian evolution because phenotypic modifications preceded genotypic modifications, whereas we considered that genes that were highly transcribed in the new niche but were later inactivated obeyed Darwinian rules. We used the total transcript abundances and sequences described for the genes of Mimivirus at the beginning of its laboratory life and after 150 passages in allopatric culture on *Acanthamoeba* spp. We found a statistically significant positive correlation between the level of gene expression at the beginning of the culture and gene inactivation during the 150 passages. In particular, the mean transcript abundance at baseline was significantly lower for inactivated genes than for unchanged genes (165 ± 589 vs. 470 ± 1,625; $p <$ 1e–3), and the mean transcript levels during the replication cycle of Mimivirus M1 were up to 8.5-fold lower for inactivated genes than for unchanged genes. In addition, proteins tended to be less frequently identified from purified virions in their early life in allopatric laboratory culture if they were encoded by variable genes than if they were encoded by conserved genes (9 vs. 15%; $p = 0.062$). Finally, Lamarckian evolution represented the evolutionary process encountered by 63% of the inactivated genes. Such observations may be explained by the lower level of DNA repair of useless genes.

Keywords: Mimivirus, Lamarckian evolution, Darwinian evolution, gene expression profile, transcription profile, genome reduction, allopatry

INTRODUCTION

Two primary mechanisms of evolution, Lamarckian and Darwinian, have been generally recognized (Koonin, 2009; Koonin and Wolf, 2009). Among the elements that differentiate the theory of evolution of Lamarck (1809) from that of Darwin (1859) is the central Lamarckian concept that phenotypic changes result from adaptation to the environment and can be transmitted vertically. According to this view, phenotypic changes precede genotypic changes (**Figure 1**). In contrast, in the current vision of evolutionary biology and in accordance with the post-Darwinian modern synthesis, genetic modifications produce phenotypic changes and precede selection of the fittest in a given environment (Koonin, 2009). In this scenario, genotypic changes precede phenotypic changes.

Our strain of *Acanthamoeba polyphaga* Mimivirus was recently subcultured 150 times on germ-free amoebae. This allopatric niche is very different from that found in the natural environment of Mimivirus, where the virus is in competition with many other organisms (Raoult and Boyer, 2010; Boyer et al., 2011; **Figure 2**). An interesting feature of this process of experimental evolution was the occurrence of Mimivirus gene variability and loss concurrently with the emergence of phenotypically different viruses (Boyer et al., 2011). The observed phenotypic changes included a lack of surface fibers and morphologically different viral factories compared with the first generation of wild-type Mimivirus at the beginning of the allopatric laboratory culture. In addition, another team had analyzed the transcriptome of this first generation of Mimivirus over its entire replication cycle (Legendre et al., 2010). These studies provided an opportunity to assess the respective roles of Lamarckian and Darwinian evolution in Mimivirus (**Figures 1–3**). Thus, we analyzed the nucleotide and amino acid variability of Mimivirus genes during the maintenance of the allopatric laboratory culture on amoebae by comparing their sequences in Mimivirus M4, recovered after 150 passages, with those in M1, recovered at the beginning of the *in vitro* culture (**Figures 2** and **3**). To quantify the apparent Lamarckian and

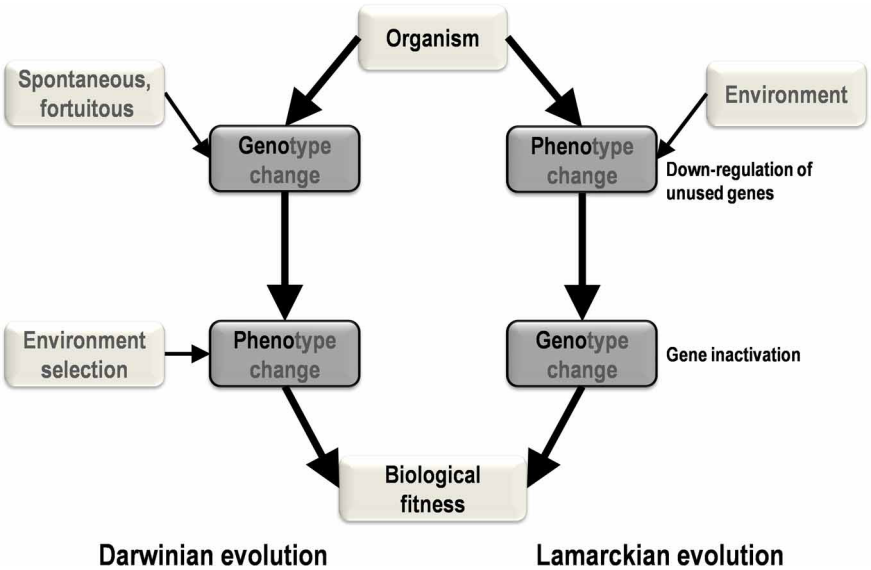


FIGURE 1 | A schematic diagram of the major steps and causes and effects in Darwinian and Lamarckian evolution. In Darwinian evolution (left), genotypic change precedes phenotypic change, whereas these changes occur in the opposite order in Lamarckian evolution (right).

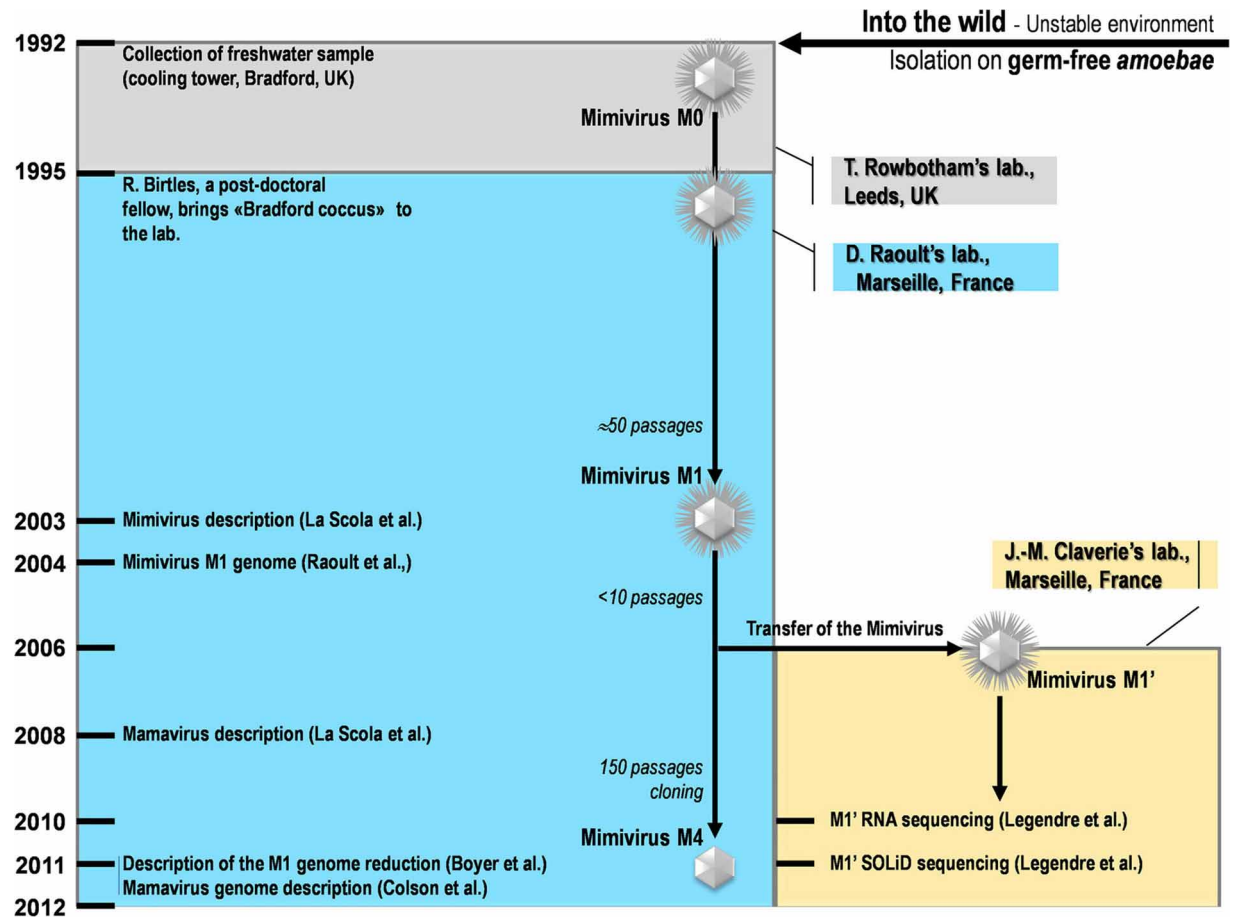


FIGURE 2 | A schematic diagram of the collection, isolation, and experiments conducted for Mimivirus and Mamavirus.

Darwinian evolutionary processes in this experiment, we postulated that the inactivated genes of Mimivirus that had been down-regulated at an early phase in the new niche of the virus experienced Lamarckian evolution, as phenotypic modifications preceded genotypic modifications (**Figure 1**). Indeed, the regulation of gene transcription has been increasingly described in association with phenotypic changes that occur in life forms when they are introduced to a new biological and ecological niche (Revel et al., 2002; La et al., 2008; Smith and Kruglyak, 2008). An example of the dramatic alteration of gene expression was recently reported in a plant bacterium following host switching (Oshima et al., 2011). In contrast, inactivated Mimivirus genes that were normally transcribed during the early phase of life in the new niche were considered to obey Darwinian rules (**Figure 1**).

MATERIALS AND METHODS

MIMIVIRUS M1 AND M4 GENES STUDIED

The sets of genes for Mimivirus at the beginning of its laboratory life (Mimivirus M1) and after 150 passages in allopatric culture in *Acanthamoeba polyphaga* (Mimivirus M4) corresponded to those reported previously (Raoult et al., 2004; Boyer et al., 2011; **Figure 2**). BLASTn searches were performed with the Mimivirus M1 set of open reading frames (ORFs) against the Mimivirus M4 genome.

GROUPS OF GENES DEFINED BASED ON THEIR VARIABILITY DURING ALLOPATRIC LABORATORY CULTURE

The genes present in Mimivirus M1 were classified into two major groups: group A is composed of genes that remained unchanged (i.e., showed 100% nucleotide identity) in the genome of Mimivirus M4, whereas group B includes Mimivirus genes showing variability at the nucleotide level during the allopatric laboratory culture (nucleotide identity $<100\%$). The variable genes with a frameshift associated with a $>30\%$ size reduction of their product were considered to be inactivated. Group C is composed of Mimivirus M1 genes lost during the transition to the Mimivirus M4 genome, as indicated by the absence of significant BLASTn hits; these genes are located within two large fragments of the genome of Mimivirus M1 that have been deleted in Mimivirus M4. These two large deletions described by Boyer et al. have been considered to be "catastrophic" events that are neither Lamarckian nor Darwinian evolutionary processes, and they were, therefore, not included in our analysis (Boyer et al., 2011).

TRANSCRIPTION AND EXPRESSION PROFILES OF MIMIVIRUS M1 GENES

The transcription profile of the first generation of Mimivirus corresponded to the transcript abundances determined by Legendre et al. at $T = 0$, 1.5, 3, 6, 9, and 12 h of the viral replication cycle (Legendre et al., 2010). These results were those for Mimivirus M1' and were considered to correspond to an early stage of the laboratory culture after Mimivirus moved from a sympatric niche to an allopatric niche (**Figure 2**). For each gene, the total number of normalized reads counts encompassing the different time points of the replication cycle was used (Legendre et al., 2010). The mean [±standard deviation (SD)] of gene total transcript abundance was 372 ± 1,342. The median value for all of the genes was 69. Gene expression was considered to be high at the beginning of the laboratory culture if the transcript abundance was equal to or greater than the median transcript abundance for all Mimivirus genes. Conversely, genes were considered to be weakly expressed at the beginning of the laboratory culture if the transcript abundance was lower than the median value for all genes. Centiles were calculated for the values of transcript abundance for all genes, and the proportions of unchanged and inactivated genes were calculated per groups of 10 centiles. The proportion of Lamarckian evolution was inferred from the proportion of Mimivirus genes weakly expressed at the beginning of the culture relative to the proportion inactivated after 150 passages in allopatric laboratory culture. Finally, the proteins cited as showing an association with Mimivirus M1 are those previously identified from purified virions by capillary LC-MS/MS, 2D gel electrophoresis, and MALDI-TOF mass spectrometry (Renesto et al., 2006).

COMPARISON OF THE TRANSCRIPTION PROFILES, EXPRESSION, AND VARIABILITY OF MIMIVIRUS M1 GENES

We tested whether the transcription levels obtained for Mimivirus M1' genes at the baseline state of the laboratory culture predict the inactivation of these genes after 150 passages on germ-free amoebae (**Figures 2** and **3**). The proportions of highly and weakly transcribed genes, of the genes among the 25 most highly expressed or with transcript abundances above the 70th centile were compared between unchanged and variable or inactivated genes. In addition, the mean transcript abundance was compared among groups of genes defined on the basis of their variability during the allopatric laboratory culture. The correlations between the initial transcript abundance for each gene and the nucleotide variability or the number of nucleotide and amino acid positions that varied during the 150 passages were also studied. Moreover, both the transcript abundance and the number of differences between corresponding genes in Mimivirus M1 and M4 were plotted according to the location of the genes within the genome. For improved clarity, the mean values calculated for a sliding window of 10 genes and a step of 1 gene were presented with Microsoft Excel software. The number of A- or T-homopolymers with a stretch of ≥ 4 was determined for each Mimivirus gene. The occurrence of nucleotide differences between Mimivirus M4 and M1 flanking such a homopolymer was also assessed. In the statistical analysis of the data, proportions were compared with a corrected chi-square test or a Fisher exact test, and comparisons of means were performed with OpenEpi Epidemiologic Calculators v. 2.3.1 (www.OpenEpi.com). Linear regression was performed with MedCalc v. 11.6.1.0 (http://www.medcalc.org). P values <0.05 were considered to be statistically significant.

RESULTS

Of the 960 genes present in Mimivirus M1, excluding those lost during the allopatric laboratory culture and representing large deletions, 606 (77%) were unchanged in the Mimivirus M4 genome (group A), and 185 (23%) were variable at the nucleotide level (group B) in the genome of Mimivirus M4. A total of 83 genes (10%) were considered to have been inactivated

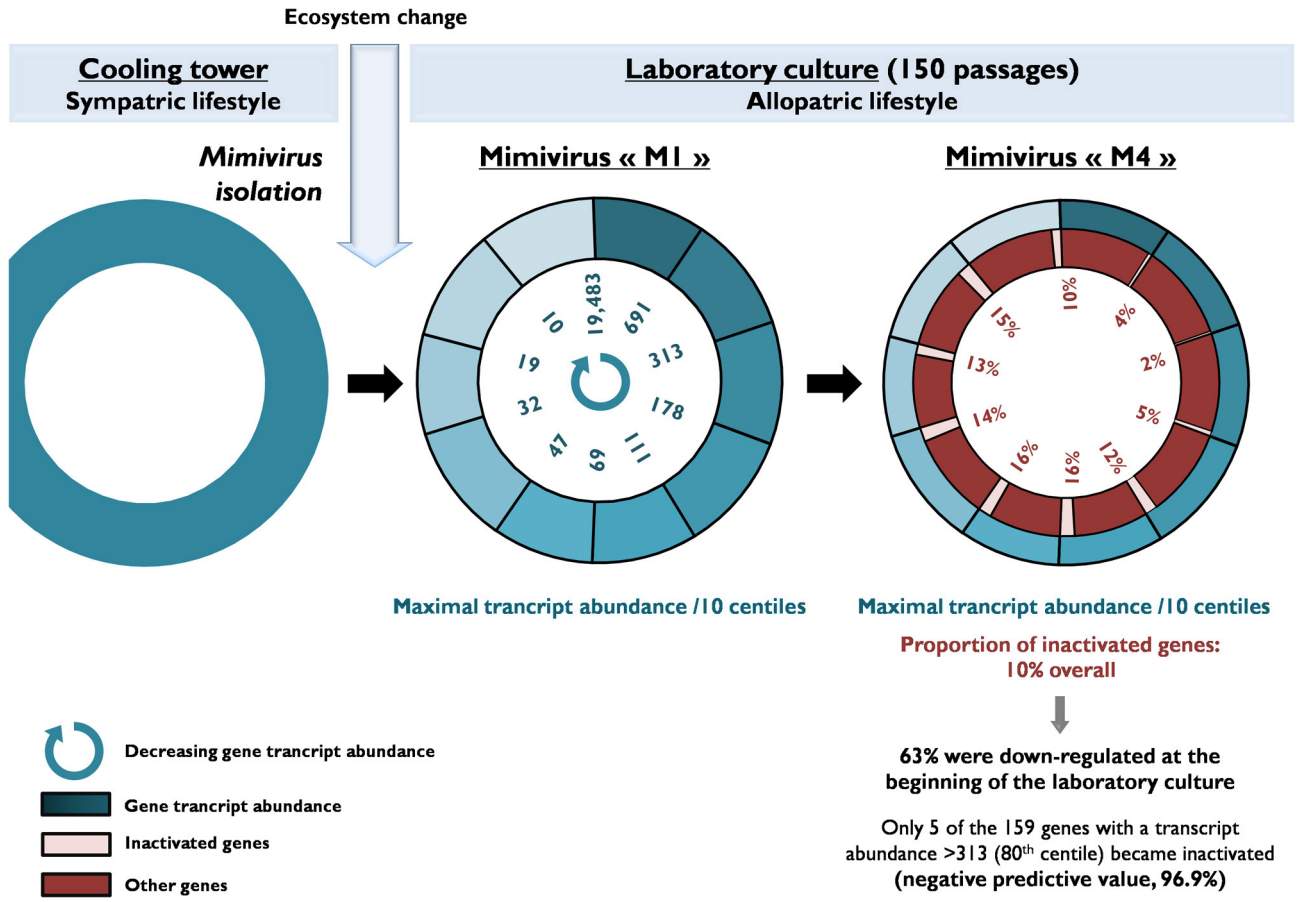


FIGURE 3 | A schematic diagram of the evolution of the gene content of Mimivirus during allopatric laboratory culture on amoebae. The diagram shows the transcript abundance for Mimivirus M1 genes and the proportions of inactivated genes after 150 passages.

during the allopatric laboratory culture. The number of variable nucleotide positions between the same genes in Mimivirus M1 and M4 ranged between 0 and 7 (mean ± SD, 1.8 ± 1.3; **Figure 4**). The mean transcript abundance at baseline was significantly lower for inactivated genes than unchanged genes (165 ± 589 vs. 470 ± 1,625, respectively; $p < 1e{-}3$) and for variable genes than unchanged genes (141 ± 415 vs. 470 ± 1,625, respectively; $p < 1e{-}3$; **Table 1**; **Figures A1** and **5**). In addition, the mean transcript levels at different time points of the replication cycle of Mimivirus M1 were up to 8.5-fold lower for inactivated genes than unchanged genes (**Figure 6**). Moreover, the proportion of inactivated genes was significantly lower among the genes that were highly expressed at baseline than among those that were weakly expressed [7.7% (31/405) vs. 13.5% (52/386); $p = 0.0077$; relative risk (RR), 0.57 (95% confidence limits for RR (CI95), 0.37–0.87)] (**Figure 3**), and the proportion of variable genes was significantly lower among the genes that were highly expressed at baseline than among those that were weakly expressed [18.8% (76/405) vs. 28.2% (109/386); $p = 0.0017$; RR, 0.66 (CI95, 0.51–0.86)]. Otherwise, the proportion of genes with a transcript abundance greater than the 70th centile of the values for all genes (178) was significantly lower among the inactivated genes than among the unchanged genes [13.3 vs. 35.0%; $p < 1e{-}3$; RR, 0.38 (CI95, 0.22–0.66)]. This proportion was also significantly lower among the variable genes than among the unchanged genes [18.9 vs. 35.0%; $p < 1e{-}3$; RR, 0.29 (CI95, 0.19–0.46)]. The proportion of variable genes in the 25 genes most expressed at baseline tended to be significantly lower than the corresponding proportion of unchanged genes [0.5 vs. 3.5%; $p = 0.034$; RR, 6.4 (CI95, 0.87–47.34)]. The negative predictive value (NPV) of being inactivated was 92.3% for highly expressed genes. In addition, this NPV was 95.5 and 96.9% for genes with a transcript abundance greater than the 70th and 80th centile, respectively, of the values calculated for all genes. Furthermore, the proportion of genes encoding proteins identified from purified virions in their early life in allopatric laboratory culture tended to be lower in the variable than in the unchanged genes [9% (17/185) vs. 15% (88/606); $p = 0.062$; RR, 0.63 (CI95, 0.39–1.04)]; nine of the genes encoding proteins present in the virions were inactivated. Also, among the 23 class I–III core genes of the nucleocytoplasmic large DNA viruses (NCLDVs) that were not located within the large deletions observed in the Mimivirus M4 genome, 17 (74%) remained unchanged during the 150 passages on germ-free amoebae. Eleven (65%) of these 17 genes were highly expressed. In contrast, only one (4.3%) NCLDV core gene of classes I–III was inactivated; this

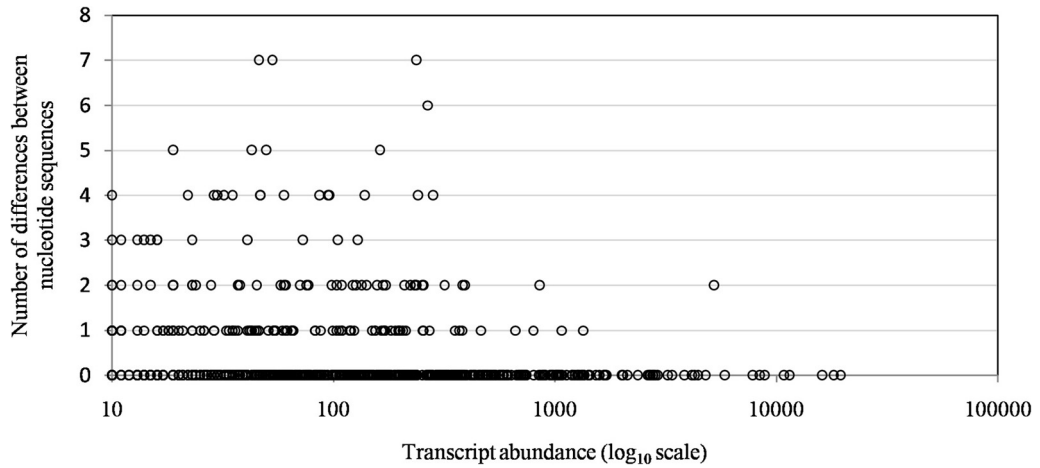


FIGURE 4 | The distribution of transcript abundance for Mimivirus M1 genes according to the number of differences between the nucleotide sequences of the Mimivirus M4 and M1 genes. A difference corresponds to the presence of different nucleotides at the same position within a pairwise alignment provided by BLAST for the nucleotide sequences of corresponding genes of Mimivirus M1 and Mimivirus M4.

Table 1 | Comparative features for unchanged, variable, and inactivated Mimivirus genes after 150 passages in allopatric culture on amoebae.

Features of Mimivirus M1 genes	Group A: unchanged genes ($n = 606$)	Group B: variable genes	
		All ($n = 185$)	Inactivated genes ($n = 83$)
Mean transcript abundance[a]	470 ± 1,625	141 ± 415*	165 ± 589*
Number of genes among the 25 most transcribed (%)	21 (3.5)	1 (0.5)*	1 (1.2)*
Number of genes with transcript abundance > the 70th centile for all genes (%)	212 (35)	35 (19)*	11 (13)*
Number of genes with transcript abundance ≤ the 10th percentile for all genes (%)	65 (10.7)	17 (9.2)	8 (9.6)
Proteins identified by proteomics (%)	88 (14.5)	17 (9.1)†	9 (11)

** Proportion is significantly different from that of the genes of group A ($p < 0.05$; chi-square corrected test or Fisher's exact test).*

† Proportion tends to differ statistically from that of the genes of group A ($0.05 < p < 0.1$; chi-square corrected test).

[a] From Legendre et al. (2010).

latter gene was weakly expressed at baseline in the laboratory culture.

The proportion of inactivated genes was the lowest [2.4% (2/82)] for those with a transcript abundance between the 80th and 90th centiles (313–691). This proportion was < 5% above the 80th centile and >10% below the 80th centile (**Figure 3**). Moreover, the proportion of variable genes ranged from 6.3% (5/79) for those with a transcript abundance above the 90th centile to 38.0% (30/79) for those with a transcript abundance between the 10th and 20th centiles (**Figure 3**). Among the 79 Mimivirus M1 genes with a transcript abundance above the 90th centile, 5 (6%) were variable and 3 (4%) were inactivated. The three genes that were inactivated were two hypothetical proteins (R401, R750b) and an S/T protein kinase (R400). Conversely, among the 82 Mimivirus M1 genes with a transcript abundance below the 10th centile (<10), 17 (21%) were variable and 8 (10%) were inactivated. The distribution of genes according to their levels of transcript abundance showed very low proportions above the 80th centile for inactivated and variable genes, whereas the distribution of unchanged genes was homogeneous (**Figure 7**). Finally, the proportion of Lamarckian evolution, as defined by the proportion of genes weakly expressed at the beginning of the allopatric laboratory culture among those inactivated, was 63%.

We sought to visually assess the relationship between gene transcription and nucleotide variability between the same genes in Mimivirus M1 and M4. For this purpose, we plotted the mean values for these two parameters along the genome according to a sliding window of 10 genes (step = 1) because representing the values for all genes did not allow sufficient legibility. As shown in **Figure A2**, this representation clearly indicates a strong inverse correlation between the initial transcript abundance and further gene variability. Thus, the regions of the genome composed of genes that were initially weakly expressed are those in which the genes showed the greatest variability. It might be hypothesized that mutations observed in the genome of Mimivirus M4 relative to Mimivirus M1 are the result of sequencing errors. Thus, the mean ± SD number of A- or T-homopolymers with a stretch of ≥4 per gene was 11.2 ± 9.1 (range, 0–79), and

the mean ± SD number of nucleotide differences that flanked such homopolymers was 1.2 ± 1.0 (range, 0–6). Nevertheless, the mean number of homopolymers per 100 nucleotides was similar for the unchanged and variable genes, 1.0 ± 0.5 (range, 0–3) and 1.2 ± 0.3 (0–2), respectively.

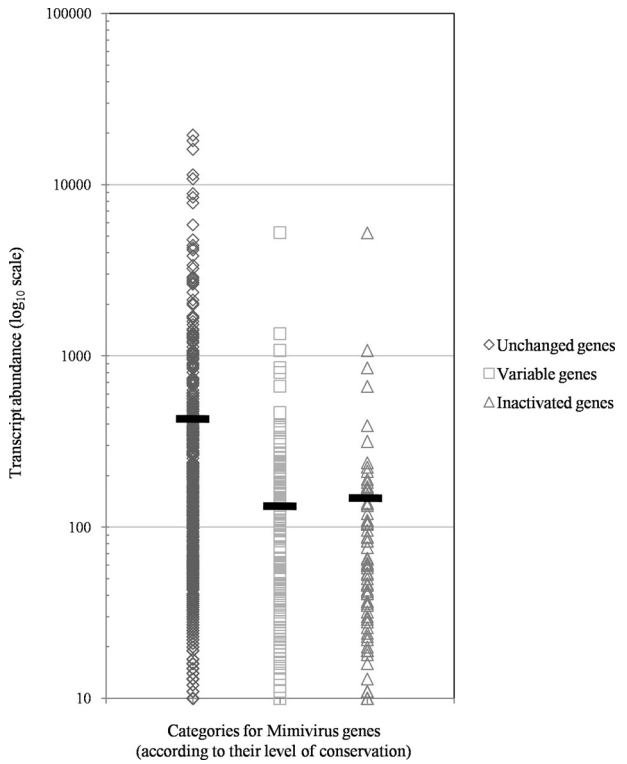


FIGURE 5 | The distribution of transcript abundance in Mimivirus M1 for different groups/subgroups of genes defined based on their variability and evolution during 150 passages in allopatric laboratory culture on amoebae. The horizontal bars indicate the mean values for each group/subgroup.

This work is based on data obtained for Mimivirus at different stages of its early life under laboratory conditions (**Figure 2**). However, it is very unlikely that the genome sequences have been significantly affected by the initial number of subcultures on *Acanthamoeba* spp. Indeed, the genomes of Mimivirus M1, Mimivirus M1' and *Acanthamoeba castellanii* Mamavirus, another strain of Mimivirus, are highly similar although the viruses experienced different numbers of passages (from less than 10 to approximately 50) on *Acanthamoeba* spp. The comparison of the Mimivirus M1 genome with that recently recovered by ultra-deep sequencing of genomic DNA and total RNA on a SOLiD platform (Legendre et al., 2011), which we called Mimivirus M1', showed that the two sequences differ by only 196 substitutions, 29 deletions and 174 insertions. The comparison of the Mamavirus genome with the Mimivirus genome showed ≈99% nucleotide identity in their alignable regions, which represent nearly the entire length of these genomes. Moreover, the Mamavirus and Mimivirus pairs of genes with bidirectional best hits show a mean nucleotide identity of 98.8%, and the majority of the pairs have identity levels greater than 99%. Interestingly, among the 19 genes present in the Mimivirus genome and absent in Mamavirus that were considered in the present work (not located within the large deletions), 16 (84%) have been classified as weakly expressed, and 4 (25%) of these genes are among the inactivated genes. Moreover, among the 22 genes in which frameshifts were identified in a comparison of the Mimivirus and Mamavirus genomes, 15 (68%) were weakly expressed at baseline, and three (20%) were inactivated.

DISCUSSION

Our work shows that the majority (63%) of the genes inactivated during Mimivirus evolution in the allopatric laboratory culture was initially weakly expressed and that low gene expression at the beginning of the culture is significantly positively correlated with gene inactivation and variability during the 150 passages. A possible bias exists and is related to the initial stages of culture. Thus, Mimivirus M1 might have been selected from the

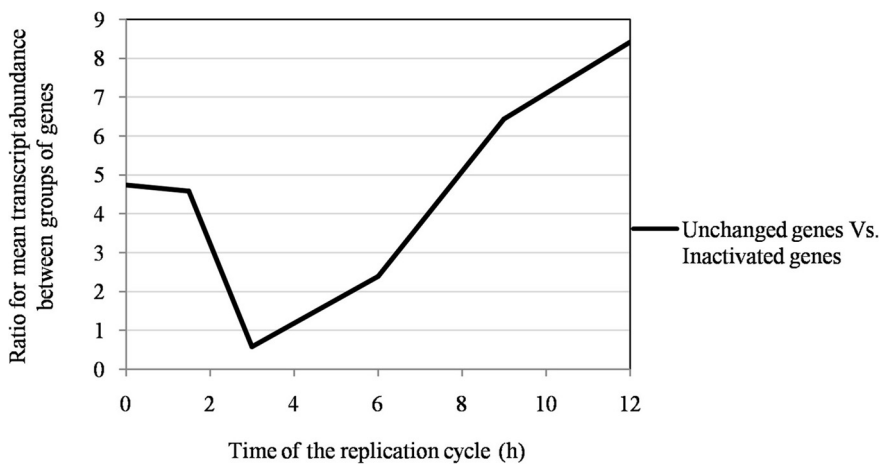


FIGURE 6 | The ratio of the mean transcript abundance in Mimivirus M1 between unchanged and inactivated genes during allopatric laboratory culture on amoebae.

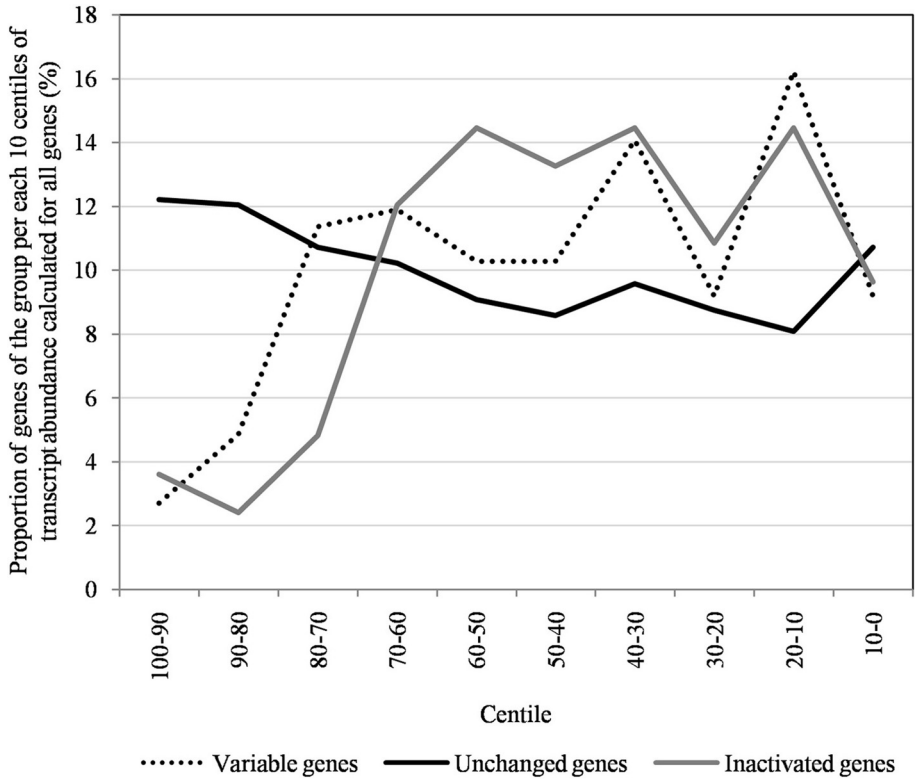


FIGURE 7 | The distribution of unchanged, variable, and inactivated genes after 150 passages in allopatric laboratory culture on amoebae per each 10 centiles of transcript abundance, calculated for all Mimivirus M1 genes.

original pool of viruses recovered from their natural environment due to a growth advantage when it was first moved to the laboratory culture environment, as is the case for every artificial system of culture. This issue has not been studied in our laboratory or by other investigators. However, the comparison of the genomes of Mamavirus (considered to be another strain of Mimivirus), Mimivirus M1 and Mimivirus M1' revealed very few differences despite differences in the number of passages on amoebae. In addition, the aim of the present work was to assess the capacity to predict genes that will be inactivated under stable laboratory conditions based on their initial transcription profile. In this context, we observed that a high level of expression for a gene strongly predicted its absence of inactivation. This finding suggests that the adaptation of Mimivirus to the modification of its environment after infection of germ-free amoebae *in vitro* is associated with a down regulation of certain genes that tended to be degraded and not repaired because they had become useless. A mechanism of this type, in which adaptation to a new ecosystem determines a new phenotype and this new phenotype promotes genotype changes transmitted to future generations, is the form of evolution described by Lamarck. One could consider that genotypic changes in the Mimivirus M1 genome were actually transmitted to new generations, as gene losses are usually considered irreversible (Krylov et al., 2003).

It was previously emphasized that the evolutionary rate of a gene sequence was negatively correlated with its level of expression or the abundance of its product (Pal et al., 2001; Koonin, 2011). This relationship was previously assessed by estimating the number of substitutions per nucleotide site between orthologous sequences in several lineages or organisms, but it has not been assessed for the same organism during experimental evolution, as is the case in the present work. In the same context, a positive correlation was recently described between the propensity for gene loss and a sequence's evolutionary rate and gene dispensability (Krylov et al., 2003). Moreover, it was found that highly expressed proteins evolve slowly (Drummond et al., 2005). In addition, it is known that DNA repair and damage processing particularly targets actively transcribed genes, as in the case of transcription-coupled repair (Hanawalt and Spivak, 2008). A similar process may explain the finding that down-regulated Mimivirus genes are more variable; they are less likely to be repaired than highly transcribed genes.

Lamarckian evolution may be involved in bacterial speciation events associated with a reduction of the genome size (Merhej et al., 2009), a finding opposed to the dominant model that considers that speciation and fitness gain are associated with an increase in gene repertoires. Thus, the major route of speciation (through adaptation to a given ecological niche) is typically through allopatry (Georgiades and Raoult, 2010) and is associated with genome size reduction through the loss of useless genes according to the mode described by Moran: "use it or lose it" (Moran, 2002). It is probable that such modifications of the gene repertoire are associated with the radical impossibility of returning to a previous ecosystem. Thus,

in a noncompetitive environment during the 150 passages in allopatric culture, Mimivirus experienced a rapid and dramatic modification of its gene content that may substantially compromise its biological fitness in more complex environments. This observation is consistent with Ernst Mayr's vision of cause and effect in biology, in which the effects of changes differ in the short term and in the long-term (Mayr, 1961). In this case, we found that Mimivirus can be selected for rapid growth in an environment without competition. However, these changes prevent the virus from attaining fitness in competition with other intraamoebal organisms (Boyer et al., 2011). Therefore, the conservation of unused genes in allopatry is only important as a long-term strategy. Together with previous data, our results suggest that the transcriptome of Mimivirus may predict the evolution of its genome in a stable laboratory culture system and that Lamarckian evolution may contribute to the evolution of the Mimivirus genome in this environment. These findings offer an incentive to study the correlation between transcription profiles and the evolution of gene sequences and repertoires in particular organisms.

REFERENCES

Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., Fournous, G., Borg, A., Robert, C., Zhang, X., Desnues, C., Henrissat, B., Rossmann, M. G., La S. B., and Raoult, D. (2011). Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10296–10301.

Colson, P. Yutin, N., Shabalina, S. A., Robert, C., Fournous, G., La Scola, B., Raoult, D., and Koonin, E. V. (2011). Viruses with more than 1, 000 genes: Mamavirus, a new *Acanthamoeba polyphaga* Mimivirus strain, and reannotation of Mimivirus genes. *Genome Biol. Evol.* 3, 737–742.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 1st edn. London: John Murray.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14338–14343.

Georgiades, K., and Raoult, D. (2010). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151

Hanawalt, P. C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* 9, 958–970.

Koonin, E. V. (2009). Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37, 1011–1034.

Koonin, E. V. (2011). Are there laws of genome evolution? *PLoS Comput. Biol.* 7:e1002173. doi: 10.1371/journal.pcbi.1002173

Koonin, E. V., and Wolf, Y. I. (2009). Is evolution Darwinian or/and Lamarckian? *Biol. Dir.* 4, 42.

Krylov, D. M., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235.

La, M. V., Raoult D., and Renesto, P. (2008). Regulation of whole bacterial pathogen transcription within infected hosts. *FEMS Microbiol. Rev.* 32, 440–460.

Lamarck, J. B. (1809). *Philosophie Zoologique, Ou Exposition Des Considérations Relatives À L'histoire Naturelle Des Animaux*. Paris: Dentu.

Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., Lartigue, A., Lescot, M., Bernadac, A., Poulain, J., Abergel, C., and Claverie, J. M. (2010). mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res.* 20, 664–674.

Legendre, M., Santini, S., Rico, A., Abergel, C., and Claverie, J. M. (2011). Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. *Virol. J.* 8, 99.

Mayr, E. (1961). Cause and effect in biology. *Science* 134, 1501–1506.

Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Dir.* 4, 13.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586.

Oshima, K., Ishii, Y., Kakizawa, S., Sugawara, K., Neriya, Y., Himeno, M., Minato, N., Miura, C., Shiraishi, T., Yamaji, Y., and Namba, S. (2011). Dramatic transcriptional changes in an intracellular parasite enable host switching between plant and insect. *PLoS ONE* 6:e23242. doi: 10.1371/journal.pone.0023242

Pal, C., Papp, B., and Hurst, L. D. (2001). Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La, S. B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329.

Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., Fourquet, P., Gorvel, J. P., and Claverie, J. M. (2006). Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J. Virol.* 80, 11678–11685.

Revel, A. T., Talaat, A. M., and Norgard, M. V. (2002). DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1562–1567.

Smith, E. N., and Kruglyak, L. (2008). Gene-environment interaction in yeast gene expression. *PLoS Biol.* 6:e83. doi: 10.1371/journal.pbio.0060083

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 February 2012; accepted: 18 June 2012; published online: 05 July 2012.

Citation: Colson P and Raoult D (2012) Lamarckian evolution of the giant Mimivirus in allopatric laboratory culture on amoebae. Front. Cell. Inf. Microbio. 2:91. doi: 10.3389/fcimb.2012.00091

APPENDIX

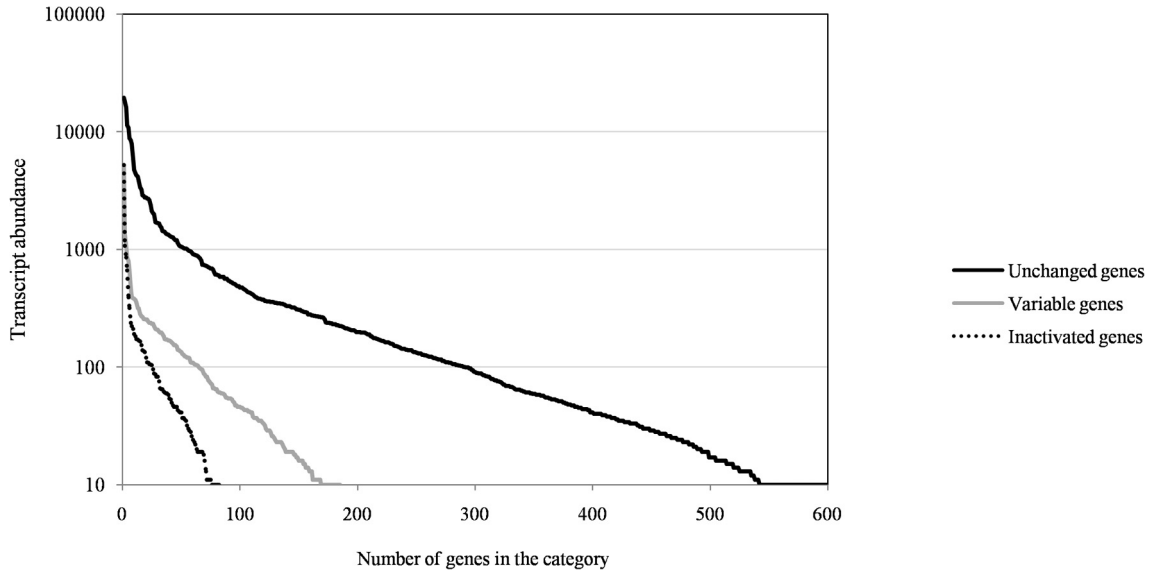


FIGURE A1 | The distribution of transcript abundance of Mimivirus M1 genes for unchanged, variable, and inactivated genes after 150 passages in allopatric laboratory culture on amoebae.

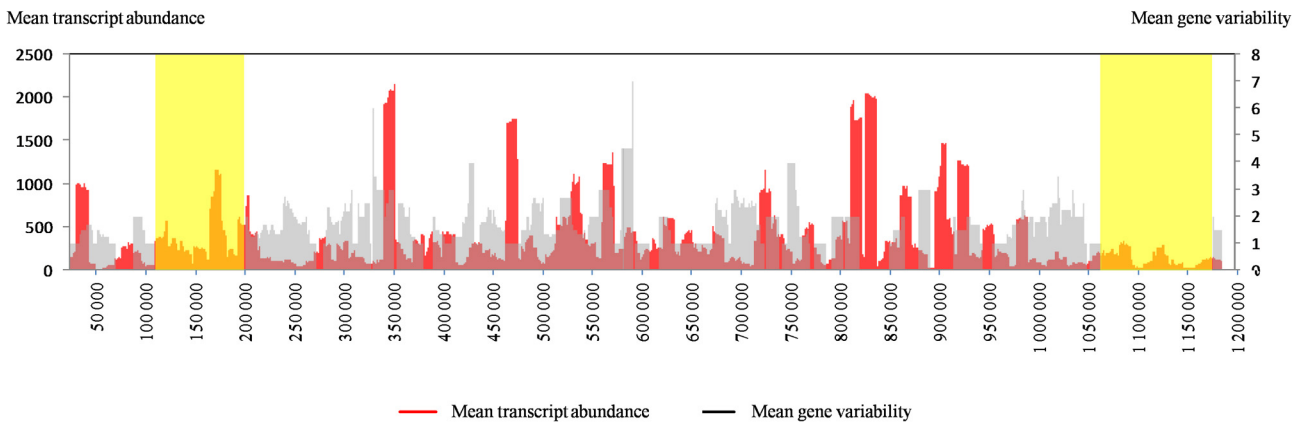


FIGURE A2 | The distribution along the Mimivirus M1 genome of the transcript abundance of Mimivirus M1 genes and the nucleotide variability for the same genes in Mimivirus M1 and M4. For both parameters, the mean values for a sliding window of 10 genes and a step of 1 gene are represented. The variability was defined as the number of variable positions within genes. The yellow boxes toward the tips of the genome indicate large deletions encountered by the genome of Mimivirus during allopatric laboratory culture on amoebae, as previously reported (Boyer et al., 2011).

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 11 April 2012
doi: 10.3389/fcimb.2012.00050

The rhizome of life: what about metazoa?

Hemalatha G. Ramulu[1,2], Didier Raoult[2] and Pierre Pontarotti[1]*

[1] *LATP UMR-CNRS 7353, Evolution Biologique et Modélisation, Aix-Marseille Université, Marseille, France*
[2] *URMITE CNRS-IRD UMR6236-198, Marseille, France*

Edited by:
Eugene V. Koonin, National Institutes of Health, USA

Reviewed by:
Jose A. Bengoechea, Fundacion Caubet-CIMERA Illes Balears, Spain
Andrey P. Anisimov, State Research Center for Applied Microbiology and Biotechnology, Russia

***Correspondence:**
Pierre Pontarotti, Evolution Biologique et Modélisation, Aix Marseille Université, UMR-CNRS 7353, case 19, 3, Place Victor Hugo, Marseille cedex 3 13331 France.
e-mail: pierre.pontarotti@univ-provence.fr

The increase in huge number of genomic sequences in recent years has contributed to various genetic events such as horizontal gene transfer (HGT), gene duplication and hybridization of species. Among them HGT has played an important role in the genome evolution and was believed to occur only in Bacterial and Archaeal genomes. As a result, genomes were found to be chimeric and the evolution of life was represented in different forms such as forests, networks and species evolution was described more like a rhizome, rather than a tree. However, in the last few years, HGT has also been evidenced in other group such as metazoa (for example in root-knot nematodes, bdelloid rotifers and mammals). In addition to HGT, other genetic events such as transfer by retrotransposons and hybridization between more closely related lineages are also well established. Therefore, in the light of such genetic events, whether the evolution of metazoa exists in the form of a tree, network or rhizome is highly questionable and needs to be determined. In the current review, we will focus on the role of HGT, retrotransposons and hybridization in the metazoan evolution.

Keywords: tree of life, horizontal gene transfer, retrotransposons, hybridization, metazoa

PROPOSED CONCEPTS FOR TREE OF LIFE AND POSITION OF METAZOA

Understanding the relationships among all living organisms by phylogenetic tree reconstruction is one of the fundamental challenges in biology. For almost 200 years, Tree of Life (TOL) has been the most powerful metaphors for biologists in depicting the evolutionary history of organisms. One of the first and most explicit form of TOL was presented by German zoologist Ernst Haeckel (1866), but its exact shape has remained elusive. Indeed, several studies to deduce TOL using various methods were carried out (Fox et al., 1980; Doolittle, 1981; Fitz-Gibbon and House, 1999; Snel et al., 1999; Tekaia et al., 1999; Lin and Gerstein, 2000; Brown et al., 2001; Clarke et al., 2002; Korbel et al., 2002; Rokas et al., 2003; Kunin et al., 2005), but its principal existence is heavily debated.

The rapid increase in molecular and genomic data in recent years have contributed to genetic events such as horizontal gene transfer (HGT), that is often considered as a major constraint in the reconstruction of phylogenetic trees. HGT "the non-genealogical transmission of genetic material from one organism to another" (Goldenfeld and Woese, 2007) is an important driving force in genomic evolution. HGT has contributed to early evolution of life to a larger extent than is presently occurring in modern biota (Zillig et al., 1992; Kandler, 1994, 1998; Woese, 2002). Indeed, number of studies have been reported about the genes acquired by HGT in three domains of life, such as Bacteria (Saunders et al., 1999; Ochman et al., 2000), Archaea (Doolittle and Logsdon, 1998; Faguy and Doolittle, 1999) and Eukaryotes (Andersson, 2005) and also between domains, i.e., from Bacteria to Archaea (Gophna et al., 2004), from Archaea to Eukarya (Andersson et al., 2003), from Bacteria to Eukarya (Watkins and Gray, 2006), from Eukarya to Bacteria (Guljamow et al., 2007) and even within Eukarya (Nedelcu et al., 2008).

Conversely, it has been assumed that the role of HGT is not prevalent in other multicellular eukaryotic organisms like in kingdom Animalia or metazoa. However, the possibility of gene transfers among them have increased in recent years and were reported in various groups such as Porifera (Rot et al., 2006), Cnidaria (Chapman et al., 2010), Nematoda (Danchin et al., 2010), Arthropoda (Fenn et al., 2006; Hotopp et al., 2007), Rotifera (Gladyshev et al., 2008), and Craniata (Graham et al., 2008; Pace et al., 2008). In addition to HGT, other genetic events such as transfer by retrotransposons and hybridization (Seehausen, 2004) were also reported, which are proposed to play an important role in the evolution of metazoa.

The studies discussed above support that HGT has played a significant role in modulating the metazoan evolution including prokaryotes and eukaryotes. Consequently, the increase in the prevalence of HGT events in Bacteria, Archaea, and Eukaryotes have resulted in the chimeric nature of genomes, where different parts of the genome can have different evolutionary histories and its difficult to identify a single common ancestor for the gene repertoire of any organism. All these results have conferred to undermine the TOL concept, thereby giving rise to a new paradigm. As a result, many proposals have emerged for the tree-like pattern replacing it with more complex models such as the "reticulate evolution" (Sneath, 1975), "synthesis of life" (Bapteste et al., 2004), "web of life" (Doolittle, 1999), "ring of life" (Rivera and Lake, 2004), "network of life" (Ragan et al., 2009), the "forest" of evolutionary trees (Puigbo et al., 2009; Schliep et al., 2010), the genetic network (Puigbo et al., 2010; Popa et al., 2011). Moreover, the evolution of species was described more like

a rhizome (Deleuze and Guattari, 1976; Raoult, 2010), reflecting various origins of genomic sequences in each species (Raoult, 2010).

The phylogenetic relationships among the major groups of animals was represented in the form of a tree (Bergstrom, 1985; Field et al., 1988; Valentine, 1997; Halanych, 2004; Giribet et al., 2007; Dunn et al., 2008). However, the increase in the extent of HGT, retrotransposons, and hybridization events in the metazoan lineage precludes the reconstruction of animal TOL. In addition to these, recent studies on whole genomic sequences from various metazoan phyla suggest a chimeric origin for its major groups due to the presence of HGT and hybridization (Syvanen and Ducore, 2010).

The results from these studies pose new questions about the evolution of metazoa, as whether it exists in the form of network or rhizome of life? Therefore, there is a need for a new pattern to be determined. In the current review, we will discuss the role of HGT, transfer by retrotransposons, hybridization and the challenges that it proposes for the current metazoan evolutionary paradigm.

METAZOA AND ITS CLASSIFICATION

Metazoa (animals) represent a group of multicellular eukaryotes (Haeckel, 1874) and constitute a monophyletic clade with heterotrophic organisms. They are believed to have diversified around the beginning of the Cambrian period (~543 million years ago). According to the recent reports, the metazoan group includes approximately 1.3 million described living species distributed in 35–40 phyla (subjected to the classification) (Edgecombe et al., 2011). This exponential growth in molecular sequence data in the metazoan group has increased the amount of phylogenetic information to study the animal relationships.

According to the Linnaeus (1758) system of classification, kingdom Animalia was classified into six classes: Amphibia, Aves, Mammalia, Pisces, Reptilia, Insecta, and Vermes, the latter divided into Intestina, Mollusca, Testacea, Lithophyta, and Zoophyta. Later on this classification has been revised constantly by biologists to study their evolutionary relationships. The early metazoan tree on animal phylogeny was based on 18S rRNA (Field et al., 1988) and Cnidaria, Ctenophora, Placozoa, Porifera, and Bilateria constituted the basal metazoan groups (Medina et al., 2001; Collins et al., 2005). However, the increase in molecular data has given rise to so-called "new animal phylogeny" (Adoutte et al., 2000; Halanych, 2004; Giribet et al., 2007) supporting the monophyly of Bilateria, which is again divided into two major lineages, Protostomia, and Deuterostomia. These two clades are well resolved by broad taxon sampling (Dunn et al., 2008; Hejnol et al., 2009; Philippe et al., 2009). The diversity within Protostomia led to division of two clades, Ecdysozoa (Aguinaldo et al., 1997; Schmidt-Rhaesa et al., 1998; Giribet, 2003; Telford et al., 2008), and Spiralia (Spiralia or Lophotrochozoa; Halanych et al., 1995; Giribet et al., 2000, 2009; Halanych, 2004; Giribet, 2008). Deuterostomia encompass two clades, Ambulacraria (Winchell et al., 2002; Brown et al., 2008), and Chordata (Delsuc et al., 2006; Mallatt and Winchell, 2007). The present consensus on metazoan phylogeny based on various hypotheses is represented in **Figure 1**.

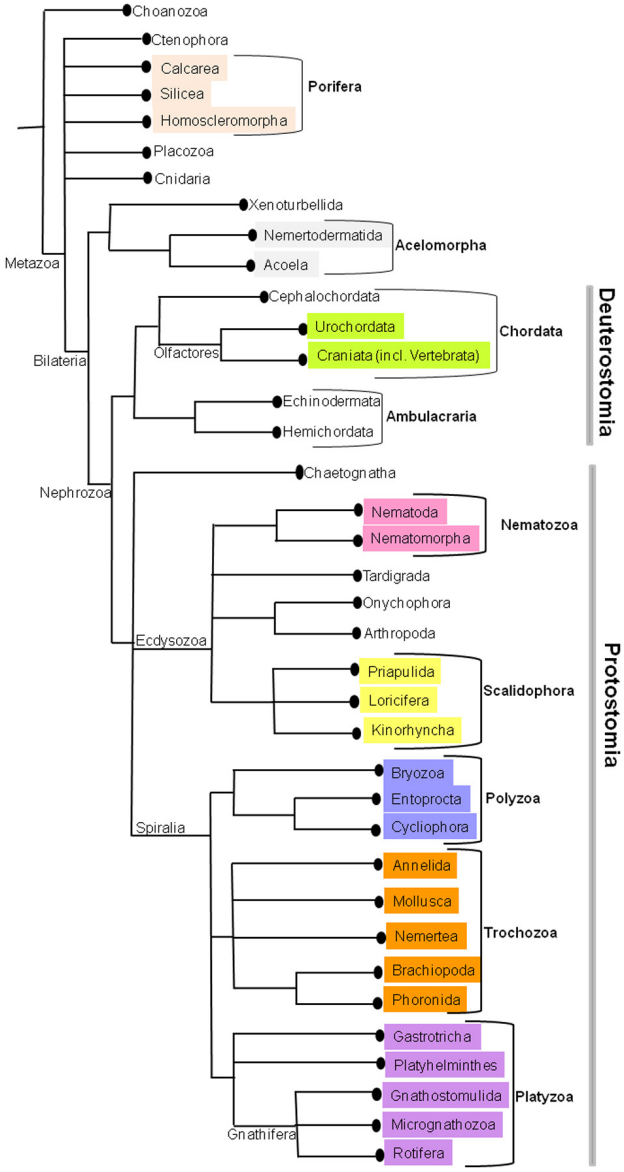


FIGURE 1 | The current phylogenetic relationships in Metazoa. This tree was generated by consensus based on various phylogenetic studies.

(**Figure 1,** Edgecombe et al., 2011). The relationships on deep metazoan groups have been extensively reviewed elsewhere (Edgecombe et al., 2011).

HORIZONTAL GENE TRANSFER IN METAZOA

HGT in animals has long been neglected and considered to be rare. However, the increase in molecular data in recent years has contributed to the possibility of gene transfers in various metazoan phyla such as Porifera, Cnidaria, Nematoda, Arthropoda, Rotifera, and Craniata, thereby creating the need to analyze more transfer events in other unidentified metazoan groups. Therefore, we will discuss the transfer events in each of them in the following sections respectively.

TRANSFER IN PHYLUM PORIFERA (SPONGE)

Poriferans represent the earliest diverging metazoans due to the presence of distinct cell types called choanocytes, which are similar to choanoflagellates, the closest unicellular relatives of metazoans (Medina et al., 2001; Nielsen, 2001; Muller, 2003). Other unique features include the lack of intestinal epithelium and digestive parenchyma (Ereskovsky and Dondua, 2006). The molecular analyses resolved sponges at the base of metazoa (Peterson and Butterfield, 2005).

The mitochondrial genome of Metazoa lacks introns, except in case of corals and sea anemones (Cnidaria), in which group I introns have been discovered in the *cox1* and nad5 genes (Beagley et al., 1996; van Oppen et al., 2002). Infact, a recent cross-kingdom HGT of group I intron of *cox 1* gene in sponge *Tetilla* sp. (Spirophorida) mitochondrial genome from fungal origin was reported, indicating the unexpected plasticity of the mitochondrial genomes of basal Metazoa (Rot et al., 2006).

A unique and first known case of HGT event of octocoral *mtMuTS* gene into animal mitochondrial genome has been reported (Bilewitch and Degnan, 2011), suggesting the need to reconsider the evolution of mitochondrial genome in metazoa.

TRANSFER IN PHYLUM CNIDARIA

The phylum Cnidaria constitutes a diverse monophyletic group (Collins, 2002). Cnidarians have many different cell types, including gametes and nematocytes, which originate in the adult form from an interstitial cell lineage. Many of them are characterized by a complex, metagenetic life cycle including a sexually produced planula larva that metamorphoses into a sessile polyp stage, which may in turn asexually produce morphologically distinct, free-swimming, sexual medusae (Hyman, 1940).

Few instances of HGT were also seen in Cnidarians. For example, a subunit of bacterial poly-γ-glutamate (PGA) synthase was transferred to metazoan ancestor, suggesting its significant role on the evolution of stinging cells (nematocytes) in cnidarians (sea anemones, jellyfish, corals, etc.) (Denker et al., 2008). Hydra is simple freshwater animal which reproduce asexually by budding. The genome of *Hydra magnipapillata* contains 71 candidates for HGT, that show closer relationship to bacterial genes than metazoan genes and 70% of these are supported by ESTs (Chapman et al., 2010).

PROTOSTOMIA

Protostomes are defined as a group of animals in which blastopore typically becomes the future mouth in most of the groups (Nielsen, 2001). It consists of two clades, Ecdysozoa and Spiralia (Spiralia or Lophotrochozoa). Ecdysozoa includes the following phyla: Nematoda, Nematomorpha, Tardigrada, Onychophora, Arthropoda, Priapulida, Loricifera, and Kinorhyncha (**Figure 1**) and Spiralia include two clades Platyzoa (Cavalier Smith, 1998) and Trochozoa (Roule, 1891).

Ecdysozoa are called as moulting protostomes and two phyla that are included under Ecdysozoa with reported cases of HGT include Nematoda and Arthropoda (Aguinaldo et al., 1997; Schmidt-Rhaesa et al., 1998; Edgecombe et al., 2000; Garey, 2001; Peterson and Eernisse, 2001; Zrzavý, 2003, **Figure 1**).

TRANSFER IN PHYLUM NEMATODA

Nematodes represent the largest animal phylum, with an estimated number in the range of one to ten million species (Lambshead, 1993) and are found in virtually all habitats on earth. Many of them are parasites of plants and animals, including humans. The recent increase in nematode genomes has made attributions in comparative genomics to study the impact of HGT on their adaptation to new ecological niches. Although inter-kingdom HGT was initially controversial, it has been established with evidence of such recent events (Richards et al., 2011). Danchin et al. (2010) have studied the whole-genome sequences of root-knot nematodes and cyst-nematodes for the genes encoding proteins involved in the plant cell wall degradation and showed the incorporation of at least six distinct types of bacterial genes encoding proteins that can modify the plant cell wall into their genomes. These have subsequently undergone extensive gene duplication in the nematode lineages.

Furthermore, the cases of HGT from a diverse set of microorganisms into various nematode genomes have also been identified. The genome of *Bursaphelenchus xylophilus* (the pine wilt nematode) has incorporated six glycoside hydrolase family 16 (GH16) proteins from gammaproteobacteria, two hydrolases from *Firmicutes* and four aspartic-type endopeptidases and 11 GH45 cellulases from the Ascomycota of fungal origin (Kikuchi et al., 2011). The genome of *Pristionchus pacificus* (a necronemic nematode) suggests that it contains substantial amount of genes of insect origin (Rödelsperger and Sommer, 2011). The *Meloidogyne incognita* (the root-knot nematode) contains genes similar to those of actinobacteria, proteobacteria, and fungi (Abad et al., 2008). The plant-parasitic nematode *Heterodera glycines* contains a biosynthetic pathway for vitamin B6 of bacterial origin (Craig et al., 2008).

TRANSFER IN PHYLUM ARTHROPODA

The members of the phylum Arthropoda are characterized by exoskeleton, segmented bodies and jointed appendages (Valentine, 2004). The appendages form part of an exoskeleton, which is mainly made of α-chitin, a derivative of glucose (Cutler, 1980). They are important members of marine, freshwater, land, and air ecosystems and are one of only two major animal groups that have adapted to life in dry environments (Ruppert et al., 2004). They include insects, arachnids, crustaceans, others and account for over 80% of all known living animal species (Anna, 2008).

The reported cases of HGT in insects include the acquisition of P elements by *Drosophila melanogaster* from *Drosophila willistoni* (Daniels et al., 1990; Engels, 1997), transfer of entire genes for carotenoid biosynthetic pathway from fungi (Moran and Jarvik, 2010) and 12 genes from bacteria (IAGC, 2010; Nikoh et al., 2010) to *Acyrthosphion pisum*. The largest HGT transfer such as the transfer of entire *Wolbachia* bacterial genome (∼1.4 Mb) into *Drosophila ananassae* Hawaii 2L chromosome (Hotopp et al., 2007) and also of *Wolbachia* to a wider range of insects such as pea aphid (Hemiptera), mosquitoes (Diptera), beetle (Coleoptera), fruit flies (Diptera), and parasitoid wasps (Hymenoptera) (Fenn et al., 2006; Hotopp et al., 2007) has been described. Recently, transfers have also been reported in *Bombyx*

mori that has acquired 10 genes from plant and bacteria (Zhu et al., 2011) and 14 types of 22 transferred genes from entomopathogenic bacteria, of which 13 types shared homology with sequences of Lepidopteran insects (Li et al., 2011), thus providing novel insights on the biological significance of HGT in the evolution of metazoan recipients.

TRANSFER IN PHYLUM ROTIFERA

The Rotifera (also called Rotatoria) is a group of aquatic micrometazoans that usually occur in freshwater and marine environments (Wallace, 1998; Wallace et al., 2006; Segers, 2007) and are classified under Gnathifera of Platyzoa (**Figure 1**). The word "rotifer" is derived from a Latin word meaning "wheel-bearer" (Pechenik, 2005) and commonly called wheel animals due to the presence of corona in the cephalic region that resembles a wheel, which is used for locomotion and food gathering. They form the largest metazoan asexual group where no sexual reproduction has ever been reported and represent an ancient origin of asexuality with great evolutionary success in the diversification of the species (Mark Welch and Meselson, 2000; Fontaneto et al., 2007).

The first evidence of HGT has been reported in bdelloid rotifers to contain many foreign genes from diverse bacterial and eukaryotic origin, which are strongly concentrated in the telomeric regions along with diverse transposable elements (TEs) (Gladyshev et al., 2008, **Figure 2**). The evidence of extensive HGT in these asexual animals might be due to repeated cycles of desiccation-induced membrane disruption and DNA breakage and repair that occur as part of their life style.

The HGT in the asexual organisms (i.e., Rotifera and Hydra), suggests that they might have accumulated mutations in an irreversible manner in the absence of recombination through sexual reproduction (Muller, 1932, 1964; Felsenstein, 1974).

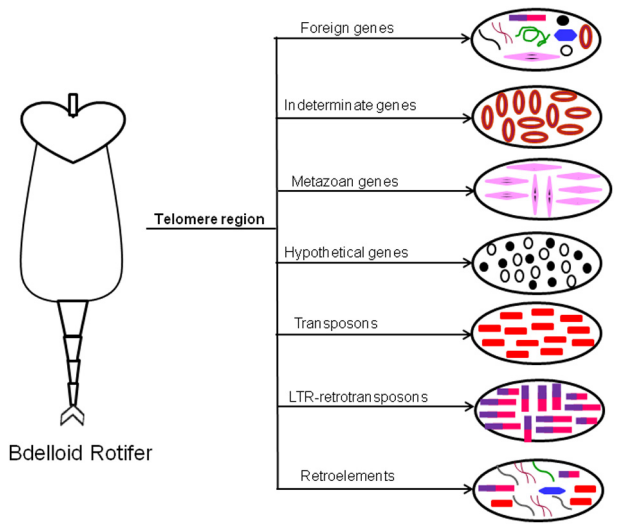


FIGURE 2 | Horizontal gene transfer in Bdelloid Rotifer. Bdelloid rotifer contain diverse array of foreign genes acquired by HGT from various origin (like bacteria, metazoa, retrovirus, etc.). They are strongly concentrated in the telomeric region. The names of the transferred genes are indicated above the arrows.

DEUTEROSTOMIA

Deuterostomes are a group of animals in which the blastopore becomes the anus in the adult, while the mouth develops as a new opening from the end of the archenteron (Nielsen, 2001). It includes two main clades, Ambulacraria and Chordata. Phylum Chordata is again classified into Cephalochordata, Urochordata, and Craniata (incl. Vertebrata, see **Figure 1**).

CRANIATA

The term "Craniata" was coupled with "Vertebrata" by Linnaeus (1758) to include lampreys, jawed fishes, and terrestrial vertebrates (tetrapods). The Craniata or craniates, are characterized by a skull (or cranium, hence their name). They comprise of all fishes including jawless fishes as hagfishes and lampreys, amphibians, reptiles, birds, and mammals, including Man. Now, the majority of the craniate species are represented by one group of fish, the actinopterygians, and the tetrapods (four-legged vertebrates) (Philippe, 1997).

TRANSFER IN CLASS TELEOSTEI (FISHES)

Teleostei (the ray-finned fishes) represent one of the three classes of actinopterygii and includes most of the living fishes (Miller and Harley, 2007).

Cases of HGT in fishes are very sparse which include the transfer of lectin-like antifreeze proteins between them (Graham et al., 2008). Although the transfer of retroposons from *Schistosoma japonicum* (blood fluke) to salmonoid fishes was identified (Matveev and Okada, 2009), it was refuted with no evidence of such transfer between the two clades based on cross-species and vector contamination and was declared as erroneous report of HGT (Grunau and Boissier, 2010).

TRANSFER IN CLASS MAMMALIA

With the burgeoning database of eukaryotic genomic sequences, it is not surprising to see the increasing cases of HGT in mammals. For example, transfer of DNA SPIN trasposons in mammals and other tetrapods (Pace et al., 2008) and transfer in human germ cells (Hecht et al., 2010) were reported. Recent studies have shown that the human body contains more of bacterial cells than human cells (Gill et al., 2006; Lester et al., 2006; Hehemann et al., 2010; Robinson et al., 2010), and many of them are dominated by members of *Bacteroidetes* and *Firmicutes* (Eckburg et al., 2005; Xu et al., 2007). Most of these recent transfers were driven by ecology rather than geography or phylogeny (Smillie et al., 2011).

In summary, the currently reported cases of HGT in animal kingdom are relatively low. However, with the availability of many new whole genome sequences, we can expect more incidences of HGT that will enable to gain further knowledge in the metazoan evolution.

FATE OF TRANSFERRED GENES IN METAZOA

The identification of several HGT cases in animals has given rise to many questions such as the function of transferred genes, their evolutionary pathways and the forces governing the transferred genes. Indeed, efforts have been made recently to answer these questions. For example, in nematodes, recent work using 454-sequencing on cellulase functioning genes has shown that they

have integrated into receptor genome by providing special functions, which indicates that genes continue to evolve with several gene duplications or deletions and DNA substitution rates after the HGT event (Mayer et al., 2011). Most of them encode enzymes for the cell wall degradation in plants and fungi and play vital role in the biology of the nematodes (Abad et al., 2008; Kikuchi et al., 2011; Mayer et al., 2011). The most recent progress has been made about the functional role of cell wall degrading enzymes of plant-parasitic nematode (Haegeman et al., 2012).

The acquisition of two enzymes for vitamin B6 biosynthesis in plant pathogenic nematode *Heterodera glycines* suggests host-parasite interactions (Craig et al., 2008). Incorporation of bacterial genes by specialized animal rumen parasites, i.e., *Giardia lamblia* (Morrison et al., 2007), *Trichomonas vaginalis* (Carlton et al., 2007) and *Entamoeba histolytica* (Loftus et al., 2005), that exist in anaerobic environments suggests that adaptation to parasitism might also favor the acquisition of new genes by HGT. Transfer of lectin-like antifreeze proteins in arctic fish might have favored them to survive in cold-conditions (Graham et al., 2008). The presence of PGA synthase genes in cnidarians might have contributed to the evolution of nematocytes, that help in prey capture (Denker et al., 2008). Recent studies on transfer in bdelloid rotifers, which has acquired genes from bacterial and eukaryotic origin, some of them are expressed suggests that they may possibly provide novel metabolic functions to these asexual animals (Gladyshev et al., 2008). The transfer of bacteria to human (Robinson et al., 2010) is interesting because it may be important to human health, have the potential to provide novel functions, there by affecting the evolution.

In summary, these results support that the transferred genes having a functional role are retained, while useless genes are eliminated. The studies outlined on the function of transferred genes in animals are still rudimentary. Hence, there is need to understand the function of transferred genes in new genomes by involving robust phylogenetic investigations and biological disciplines.

Besides HGT, other genetic events such as transfer by retrotransposons and hybridization (Seehausen, 2004) were also reported in metazoa, which may form additional limitations in the reconstruction of animal TOL.

RETROTRANSPOSONS AND THEIR CLASSIFICATION

Retrotransposons (retroelements) belong to group of TEs. TEs (also known as "jumping genes") include a diverse array of DNA sequences and possess the inherent capacity to self-reproduce and move within and between genomes. Ever since their discovery in maize DNA (McClintock, 1956), TEs have been found in genomes of almost all organisms. They constitute more than 50% of maize (*Zea mays*) genome (Kidwell and Lisch, 1997; Wessler, 1998), 22% of *Drosophila* genome (Kapitonov and Jurka, 2003) and half of our human genome with just 1.5% coding for protein region (Lander et al., 2001).

TEs are divided into two groups based on their transposition mechanism and sequence organization (Finnegan, 1989; Capy, 1998): 1. DNA transposons (move predominantly via a DNA-mediated mechanism of excision and insertion and constitute appx. 3% of human genome (Craig et al., 2002) and 2. Retroelements (move by reverse transcription of an RNA intermediate (Rogers, 1985) and include the retrotransposons, eukaryotic TEs, group II mitochondrial introns, bacterial retrointrons and retroviruses). The reverse transcriptase of the retroelements is usually encoded by the element itself. They are subdivided into two major groups based on the presence or the absence of long terminal repeats, which flank the body of the element: long terminal repeat (LTR)-containing elements (LTRs) and non-LTR retrotransposons (non-LTR). Again the non-LTRs are subdivided into two classes: LINEs or L1-element (long interspersed elements) and SINEs (short interspersed elements) (Weiner et al., 1986). LINEs, the autonomous elements are widely distributed in eukaryotes. For example, they occur in >500, 000 copies (∼17%) in human genome (Lander et al., 2001), out of which only ∼80–100 were found to be active (Brouha et al., 2003). They are also found in mouse genome (∼3000, Goodier et al., 2001) and *Drosophila* genome (Priimagi et al., 1988; Levis et al., 1993; Udomkit et al., 1995). Unlike LINEs, SINEs are non-autonomous and occupy about 12% of the human genome, out of which majority of them belong to Alu elements (Lander et al., 2001). The classification of TEs is shown in **Figure 3**.

TRANSFER BY RETROTRANSPOSONS

With the widespread distribution of intergenomic TEs in eukaryotic genomes, it is not surprising to envisage the intriguing feature of HGT among them, a process by which they cross-species boundaries to enter new genomes. In the past decade, substantial evidence of HGT has been reported for all types of TEs in invertebrates as well as vertebrates. Some of the reported cases of HGT involving TEs are summarized in **Table 1** as shown below.

In addition to these, the vertebrate genomes also contain numerous copies of retroviral sequences that were acquired over the course of evolution. The majority of them belong to endogenous viral elements, which integrate into the nuclear genome of the host germ line (Tristem, 2000; Lander et al., 2001). As it is exhaustive to provide all the reported cases of retroviral elements, we present few of them. They include the human endogenous retrovirus element HERV-L, that is related distantly by homology to foamy viruses (Cordonnier et al., 1995), and recently reported cases such as the presence of endogenous viral elements in animal genomes (Katzourakis and Gifford, 2010), integration of ancient bornavirus and ebolavirus/marburgvirus sequences in

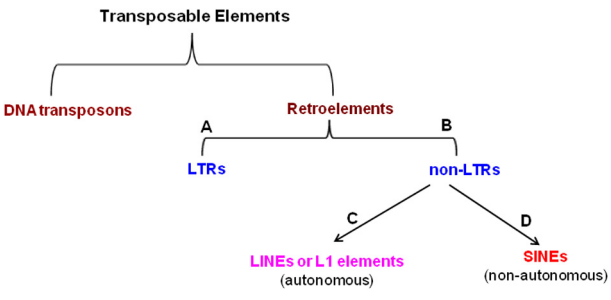


FIGURE 3 | Classification of transposable elements. (A) Long terminal repeats, **(B)** Non-long terminal repeats, **(C)** Long interspersed elements, **(D)** Short interspersed elements.

Table 1 | Table showing some of the cases of horizontal gene transfer involving Transposable Elements (TEs) among Metazoa.

Transposable elements		
Type	**Metazoan group/genome**	**Reference**
LINE/*jockey*	Arthropoda/*Drosophila*	Mizrokhi and Mazo, 1990
LINE/Bov-B	Mammals	Kordis and Gubensek, 1995
SINE/*Sma*I-cor	Teleostei (coregonid fish)	Hamada et al., 1997
LTR/*copia*	Arthropoda/*Drosophila*	Jordan et al., 1999
SURL elements	Echinodermata	Gonzalez and Lessios, 1999.
LTR/*gypsy*	Arthropoda/*Drosophila*	Vazquez-Manrique et al., 2000
LTR/*gypsy*	Arthropoda/*Drosophila*	Terzian et al., 2000
LINE/*Rex1*	Teleostei (fish genomes)	Volff et al., 2000
LINE/Bov-B	Reptiles and Mammals	Zupunski et al., 2001
P elements	Arthropoda/*Drosophila*	Daniels et al., 1990
P elements	Arthropoda/*Drosophila*	Hagemann et al., 1992
mariner elements	Arthropoda/ *Drosophila mauritiana* and *Zaprionus tuberculatus*	Maruyama and Hartl, 1991
mariner elements	Arthropoda	Robertson and MacLeod, 1993
mariner elements	Arthropoda/*Drosophila*	Brunet et al., 1994
mariner elements	Arthropoda/*Drosophila*	Lohe et al., 1995
mariner elements	Arthropoda	Robertson and Lampe, 1995
mariner elements	Mammals/*Homo sapiens*	Smit and Riggs, 1996
mariner elements	Amphibians	Lam et al., 1996
mariner elements	Platyzoa and Cnidaria	Robertson, 1997
mariner elements	Mammals	Robertson et al., 2002
DNA SPIN transposons	Mammals and other tetrapods	Pace et al., 2008
Helitrons (rolling circle DNA transposons)	Arthropoda, Reptiles, Teleostei (fish) and Mammals	Thomas et al., 2010

vertebrate genomes (Belyi et al., 2010), endogenous lentivirus in basal primates (Gifford et al., 2008; Gilbert et al., 2009) and endogenous foamy viruses in the sloth genome (Katzourakis et al., 2009).

All these results suggest that retrotransposons continue to play an active role in shaping the dynamics of metazoan evolution by forming new genes and thus contributing to the chimeric evolution of genomes. The representation of metazoan chimeric evolution of genomes is shown in **Figure 4**.

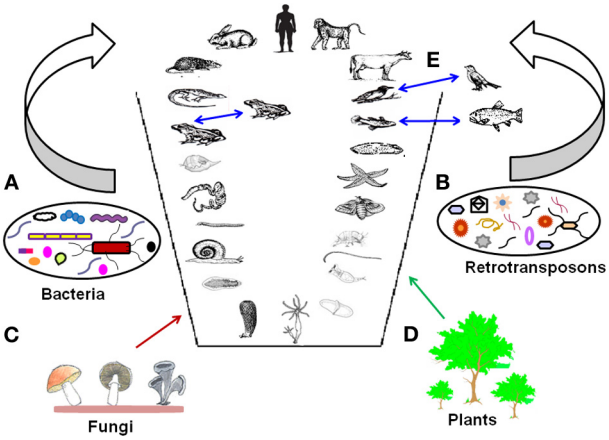


FIGURE 4 | Metazoan chimeric evolution. The figure represents chimeric nature of metazoan genomes with horizontal gene transfer, retrotransposons, and hybridization events. Horizontal gene transfer from **(A)** Bacteria, **(B)** Retrotransposons, **(C)** Fungi, **(D)** Plants and **(E)** Hybridization between closely related species (indicated by blue arrows).

HYBRIDIZATION

Hybridization, the exchange of genes between closely related species by sexual reproduction is a natural evolutionary process. The frequency of hybridization among species is common although it is rare between species on a per-individual basis. About 10–30% of multicellular animal and plant species hybridize regularly (Mallet, 2005). Closely related species tend to hybridize more. Therefore, species-rich groups in rapidly diversifying adaptive radiations tend to be those that hybridize most (Price and Bouvier, 2002; Seehausen, 2004; Gourbière and Mallet, 2010). However, the success of large number of hybridizations that occur among closely related species depends on genetic and environmental factors (Arnold et al., 2012).

Although recent studies have highlighted the events of HGT, information about hybridization is limited and highly unexplored, especially in the case of animals. However, studies have reported the cases of hybridization between more closely related lineages such as fungi, plants, and even vertebrate lineages such as amphibians, fish, and birds (Arnold, 2006). Recently, studies on primates have detected hybridization in not only between species and subspecies but also between genera, including human lineage (Zinner et al., 2011).

Hybridization among species can thus act as a catalyst for the formation of new lineages.

CONCLUDING REMARKS

The relative amount of HGT events detected in metazoa merely represent just tip of an iceberg. Currently, the reported rate of HGT in animal kingdom is relatively low and still in the state of infancy. There are still many outstanding questions that need to be addressed like: What is the rate of gene transfer within the group? How important is gene transfer in other animal genomes, especially vertebrates? There are many lineages that need to be considered for HGT, such as Rotifers, Aves (birds), Reptiles and Mammals. Why are eukaryotic genes rare in prokaryotic genomes? What is the function of transferred genes in the new genome?

Although, there is substantial evidence on transfer of TEs (especially group I introns) in the mitochondrial genomes of metazoa (ex.sponges), we are unaware on the extent of these transfers in other groups (especially fishes, rotifers, and other vertebrates), which might serve as useful indicators, raising intriguing biological questions related to HGT. The recent evidence of octocoral *mtMuTS* gene into animal mitochondrial genome also support these findings, thus challenging the evolution of mitochondrial genome in metazoa.

Therefore, there is an emergent need for future HGT studies in Metazoa. Hopefully, with the recent improvements in new sequencing technologies and increased number of diverse vertebrate genomes, we anticipate to gain novel insights into the role played by HGT that will shed light in understanding the metazoan evolution.

Given the extent of gene transfer from different origins (bacteria, plants, fungi, and eukaryotes), transfer by retrotransposons and the fusion of lineages to from new lines of descent (interspecies hybridization) in animals, we believe that majority of the animal genomes exhibit mosaic structure and chimerism, thus replacing the concept of animal TOL with a new paradigm. Therefore, the question still remains unanswered about the evolution of metazoa, as whether it can be represented in the form of a network or rhizome and needs to be determined.

ACKNOWLEDGMENTS

We would like to thank Infectiopole Sud for funding.

REFERENCES

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Ségurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., and Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nat. Biotechnol.* 26, 909–915.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., and de Rosa, R. (2000). The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4453–4456.

Aguinaldo, A. M. A., Turbeville, J. M., Lindford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., and Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387, 489–493.

Andersson, J. O. (2005). Lateral gene transfer in eukaryotes. *Cell. Mol. Life Sci.* 62, 1182–1197.

Andersson, J. O., Sjogren, A. M., Davis, L. A. M., Embley, T. M., and Roger, A. J. (2003). Phylogenetic analysis of Diplomonad genes reveals frequent lateral gene transfer affecting eukaryotes. *Curr. Biol.* 13, 94–104.

Anna, T. (2008). *The Arthopod Story*. Berkeley:University of California. http://evolution.berkeley.edu/evolibrary/article/arthropodstory

Arnold, M. (2006). *Evolution Through Genetic Exchange*. Oxford: Oxford University Press.

Arnold, M. L., Ballerini, E. S., and Brothers, A. N. (2012). Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity (Edinb.)* 108, 159–166.

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W. F. (2004). Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* 12, 406–411.

Beagley, C. T., Okada, N. A., and Wolstenholme, D. R. (1996). Two mitochondrial group I introns in a metazoan, the sea anemone Metridium senile: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5619–5623.

Belyi, V. A., Levine, A. J., and Skalka, A. M. (2010). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* 6:e1001030. doi: 10.1371/journal.ppat.1001030

Bergstrom, J. (1985). Metazoan evolution—a new model. *Zool. Scr.* 15, 189–200.

Bilewitch, J. P., and Degnan, S. M. (2011). A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evol. Biol.* 11, 228.

Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., and Kazazian, H. H. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285.

Brunet, F., Godin, F., David, J. R., and Capy, P. (1994). The mariner transposable element in the Drosophilidae family. *Heredity (Edinb.)* 73, 377–385.

Brown, F. D., Prendergast, A., and Swalla, B. J. (2008). Man is but a worm: chordate origins. *Genesis* 46, 605–613.

Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E., and Stanhope, M. J. (2001). Universal trees based on large combined protein sequence datasets. *Nat. Genet.* 28, 281–285.

Capy, P. (1998). "Classification of transposable elements," in *Dynamics And Evolution of Transposable Elements*, eds P. Capy, C. Bazin, D. Higuet, and T. Langin (Austin: Landes Bioscience), 37–52.

Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M. Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., and Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science* 315, 207–212.

Cavalier Smith, T. (1998). A revised six-kingdom system of life. *Biol. Rev.* 73, 203–266.

Chapman, J. A., Kirkness, E. F., Simakov, O., Hampson, S. E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P. G., Borman, J., Busam, D., Disbennett, K., Pfannkoch, C., Sumin, N., Sutton, G. G., Viswanathan, L. D., Walenz, B., Goodstein, D. M., Hellsten, U., Kawashima, T., Prochnik, S. E., Putnam, N. H., Shu, S., Blumberg, B., Dana, C. E., Gee, L., Kibler, D. F., Law, L., Lindgens, D., Martinez, D. E., Peng, J., Wigge, P. A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune,

S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J. S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P. R., Zhang, X., Aufschnaiter, R., Eder, M. K., Gorny, A. K., Salvenmoser, W., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K. A., Strausberg, R. L., Venter, J. C., Technau, U., Hobmayer, B., Bosch, T. C., Holstein, T. W., Fujisawa, T., Bode, H. R., David, C. N., Rokhsar, D. S., and Steele, R. E. (2010). The dynamic genome of Hydra. *Nature* 464, 592–596.

Clarke, G. D., Beiko, R. G., Ragan, M. A., and Charlebois, R. L. (2002). Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J. Bacteriol.* 184, 2072–2080.

Collins, A. G. (2002). Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *J. Evol. Biol.* 15, 418–432.

Collins, A. G., Cartwright, P., McFadden, C. S., and Schierwater, B. (2005). Phylogenetic context and Basal metazoan model systems. *Integr. Comp. Biol.* 45, 585–594.

Cordonnier, A., Casella, J. F., and Heidmann, T. (1995). Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. *J. Virol.* 69, 5890–5897.

Craig, J. P., Bekal, S., Hudson, M., Domier, L., Niblack, T., and Lambert, K. N. (2008). Analysis of a horizontally transferred pathway involved in vitamin B-6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. *Mol. Biol. Evol.* 25, 2085–2098.

Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M. (2002). *Mobile DNA II.* Washington: American Society for Microbiology.

Cutler, B. (1980). Arthropod cuticle features and arthropod monophyly. *Cell. Mol. Life Sci.* 36, 953.

Danchin, E. G., Rosso, M. N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., and Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17651–17656.

Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G., and Chovnick, A. (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124, 339–355.

Deleuze, G., and Guattari, F. (1976). *Rhizome*. Paris: Les éditions de minuit.

Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439, 965–668.

Denker, E., Bapteste, E., Le Guyader, H., Manuel, M., and Rabet, N. (2008). Horizontal gene transfer and the evolution of cnidarian stinging cells. *Curr. Biol.* 18, R858–R859.

Doolittle, R. F. (1981). Similar amino acid sequences: chance or common ancestry? *Science* 214, 149–159.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* 284, 2124–2128.

Doolittle, W. F., and Logsdon, J. M. Jr. (1998). Do Archaea have a mixed heritage? *Curr. Biol.* 8, R209–R211.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., and Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.

Edgecombe, G. D., Wilson, G. D. F., Colgan, D. J., Gray, M. R., and Cassis, G. (2000). Arthropod cladistics: combined analysis of histone H3 and U2 snRNA sequences and morphology. *Cladistics* 16, 155–203.

Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M., Neves, R. C., Rouse, G. W., Worsaae, K., and Sørensen, M. V. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Org. Divers. Evol.* 11, 151–172.

Engels, W. R. (1997). Invasions of P elements. *Genetics* 145, 11–15.

Ereskovsky, A. V., and Dondua, A. K. (2006). The problem of germ layers in sponges (Porifera) and some issues concerning early metazoan evolution. *Zool. Anz.* 245, 65–76.

Faguy, D. M., and Doolittle, W. F. (1999). Lessons from the Aeropyrum pernix genome. *Curr. Biol.* 9, R883–R886.

Fenn, K., Conlon, C., Jones, M., Quail, M. A., Holroyd, N. E., Parkhill, J., and Blaxter, M. (2006). Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. *PLoS Pathog.* 2:e94. doi: 10.1371/journal.ppat.0020094

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* 78, 737–756.

Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R., and Raff, R. A. (1988). Molecular phylogeny of the animal kingdom. *Science* 239, 748–753.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–107.

Fitz-Gibbon, S. T., and House, C. H. (1999). Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* 27, 4218–4222.

Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., and Barraclough, T. G. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biol.* 5:e87. doi: 10.1371/journal.pbio.0050087

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer,T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N., and Woese, C. R. (1980). The phylogeny of prokaryotes. *Science* 209, 457–463.

Garey, J. R. (2001). Ecdysozoa: the relationship between Cycloneuralia and Panarthropoda. *Zool. Anz.* 240, 321–330.

Gifford, R. J., Katzourakis, A., Tristem, M., Pybus, O. G., Winter, M., and Shafer, R. W. (2008). A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20362–20367.

Gilbert, C., Maxfield, D. G., Goodman, S. M., and Feschotte, C. (2009). Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genet.* 5:e1000425. doi: 10.1371/journal.pgen.1000425

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355–1359.

Giribet, G. (2003). Molecules, development and fossils in the study of metazoan evolution; Articulata versus Ecdysozoa revisited. *Zoology* 106, 303–326.

Giribet, G. (2008). Assembling the lophotrochozoan (=spiralian) tree of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1513–1522.

Giribet, G., Distel, D. L., Polz, M., Sterrer, W., and Wheeler, W. C. (2000). Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach using 18S rDNA sequences and morphology. *Syst. Biol.* 49, 539–562.

Giribet, G., Dunn, C. W., Edgecombe, G. D., Hejnol, A., Martindale, M. Q., and Rouse, G. W. (2009). "Assembling the spiralian tree of life," in *Animal Evolution: Genes, Genomes, Fossils and Trees,* eds M. J. Telford and D. T. J. Littlewood (Oxford: Oxford University Press), 53–64.

Giribet, G., Dunn, C. W., Edgecombe, G. D., and Rouse, G. W. (2007). A modern look at the Animal Tree of Life. *Zootaxa* 1668, 61–79.

Gladyshev, E. A., Meselson, M., and Arkhipova, I. R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science* 320, 1210–1213.

Goldenfeld, N., and Woese, C. (2007). Biology's next revolution. *Nature* 445, 369.

Gonzalez, P., and Lessios, H. A. (1999). Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 echinoid species. *Mol. Biol. Evol.* 16, 938–952.

Goodier, J. L., Ostertag, E. M., Du, K., and Kazazian, H. H. Jr. (2001). A novel active L1 etrotransposon subfamily in the mouse. *Genome Res.* 11, 1677–1685.

Gophna, U., Charlebois, R. L., and Doolittle, W. F. (2004). Have archaeal genes contributed to bacterial virulence? *Trends Microbiol.* 12, 213 –219.

Gourbière, S., and Mallet, J. (2010). Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution.* 64, 1–24.

Graham, L. A., Lougheed, S. C., Ewart, K. V., and Davies, P. L. (2008). Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PLoS One* 3:e2616. doi: 10.1371/journal.pone.0002616

Grunau, C., and Boissier, J. (2010). No evidence for lateral gene transfer between salmonids and schistosomes. *Nat. Genet.* 42, 918–919.

Guljamow, A., Jenke-Kodama, H., Saumweber, H., Quillardet, P., Frangeul, L., Castets, A. M.,

Bouchier, A. M., Tandeau de Marsac, N., and Dittmann, E. (2007). Horizontal gene transfer of two cytoskeletal elements from a Eukaryote to a Cyanobacterium. *Curr. Biol.* 17, R757–R759.

Haeckel, E. (1866). *Generelle Morphologie der Organismen. Allgemeine Grundzüge der Organischen Formen-Wissen-Schaft, Mechanisch Begründet Durch die von Charles Darwin Reformirte Descendenztheorie*, vol. 2. Berlin: Georg Reimer, 574–462.

Haeckel, E. (1874). Die Gastraea-Theorie, die phylogenetische Classification des Tierreichs und die Homologie der Keimblätter. *Z. Naturwiss. Jena* 8, 1–55.

Haegeman, A., Mantelin, S., Jones, J. T., and Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene* 492, 19–31.

Hagemann, S., Miller, W. J., and Pinsker, W. (1992). Identification of a complete P-elements in the genome of *Drosophila bifasciata*. *Nucleic Acids Res.* 20, 409–413.

Halanych, K. M. (2004). The new view of animal phylogeny. *Ann. Rev. Ecol. Evol. Syst*. 35, 229–256.

Halanych, K. M., Bacheller, J. M., Aguinaldo, A. M. A., Liva, S. M., Hillis, D. M., and Lake, J. A. (1995). Evidence from 18S ribosomal DNA that lophophorates are protostome animals. *Science* 267, 1641–1643.

Hamada, M., Kido, Y., Himberg, M., Reist, J. D., Ying, C., Hasegawa, M., and Okada, N. (1997). A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the SmaI family of repeats: possible evidence for the horizontal transfer of SINEs. *Genetics* 146, 355–367.

Hecht, M. M., Nitz, N., Araujo, P. F., Sousa, A. O., Rosa Ade, C., Gomes, D. A., Leonardecz, E., and Teixeira, A. R. (2010). Inheritance of DNA transferred from American trypanosomes to human hosts. *PLoS One* 5:e9181. doi: 10.1371/journal.pone.0009181

Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 464, 908–912.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W. E., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., and Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 276, 4261–4270.

Hotopp, J. C. D., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Muñoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., and Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756.

Hyman, L. H. (1940). *The Invertebrates: Protozoa Through Ctenophora.* New York, NY: McGraw-Hill Book Company, Inc.

International Aphid Genomics Consortium (IAGC). (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*. 8:e1000313. doi: 10.1371/journal.pbio.1000313

Jordan, I. K., Matyunina, L. V., and McDonald, J. F. (1999). Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12621–12625.

Kandler, O. (1994). "The early diversification of life," in *Early Life on Earth, Nobel Symposium*, ed S. Bengston (New York, NY: Columbia University Press), 152–509.

Kandler, O. (1998). "The early diversification of life: a proposal," in *Thermophiles: The Key to Molecular Evolution and the Origin of Life?* eds M. Adams and J. Weigel (London: Taylor and Francis, 19–31.

Kapitonov, V. V., and Jurka, J. (2003). Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6569–6574.

Katzourakis, A., Gifford, R. J., Tristem, M., Gilbert, M. T., and Pybus, O. G. (2009). Macroevolution of complex retroviruses. *Science* 325, 1512.

Katzourakis, K., and Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet*. 6:e1001191. doi: 10.1371/journal.pgen.1001191

Kidwell, M. G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7704–7711.

Kikuchi, T., Cotton, J. A., Dalzell, J. J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I. J., Assefa, S. A., Cock, P. J., Otto, T. D., Hunt, M., Reid, A. J., Sanchez-Flores, A., Tsuchihara, K., Yokoi, T., Larsson, M. C., Miwa, J., Maule, A. G., Sahashi, N., Jones, J. T., and Berriman, M. (2011). Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* 7:e1002219. doi: 10.1371/journal.ppat.1002219

Korbel, J. O., Snel, B., Huynen, M. A., and Bork, P. (2002). SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* 18, 158–162.

Kordis, D., and Gubensek, F. (1995). Horizontal SINE transfer between vertebrate classes. *Nat. Genet.* 10, 131–132.

Kunin, V., Ahren, D., Goldovsky, L., Janssen, P., and Ouzounis, C. A. (2005). Measuring genome conservation across taxa: divided strains and United Kingdoms. *Nucleic Acids Res.* 33, 616–621.

Lam, W. L., Seo, P., Robison, K., Virk, S., and Gilbert, W. (1996). Discovery of amphibian *Tc1*-like transposon families. *J. Mol. Biol.* 257, 359–366.

Lambshead, P. J. (1993). Recent developments in marine benthic biodiversity research. *Oceanis* 19, 5–24.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lester, C. H., Frimodt-Moller, N., Sorensen, T. L., Monnet, D. L., and Hammerum, A. M. (2006). *In vivo* transfer of the vanA resistance gene

froman *Enterococcus faecium* isolate of animal origin to an *E. faecium* isolate of human origin in the intestines of human volunteers. *Antimicrob. Agents Chemother.* 50, 596–599.

Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A., and Sheen, F. M. (1993). Transposons in place of telomeric repeats at a Drosophila telomere. *Cell* 75, 1083–1093.

Li, Z. W., Shen, Y. H., Xiang, Z. H., and Zhang, Z. (2011). Pathogen-origin horizontally transferred genes contribute to the evolution of Lepidopteran insects. *BMC Evol. Biol.* 11, 356.

Lin, J., and Gerstein, M. (2000). Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. *Genome Res.* 10, 808–818.

Linnaeus, C. (1758). *Systema Naturae per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis,* vol. 1. Holmiae: Laurentii Salvii, 824.

Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M. A., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillén, N., Gilchrist, C., Stroup, S. E., Bhattacharya, S., Lohia, A., Foster, P. G., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N. M., Petri, W. A. Jr., Clark, C. G., Embley, T. M., Barrell, B., Fraser, C. M., and Hall, N. (2005). The genome of the protist parasite Entamoeba histolytica. *Nature* 433, 865–868.

Lohe, A. R., Moriyama, E. N., Lidholm, D. A., and Hartl, D. L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.* 12, 62–72.

Mallatt, J., and Winchell, C. J. (2007). Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol. Phylogenet. Evol.* 43, 1005–1022.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237.

Mark Welch, D., and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288, 1211–1215.

Maruyama, K., and Hartl, D. L. (1991). Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus*. *J. Mol. Evol.* 33, 514–524.

Matveev, V., and Okada, N. (2009). Retroposons of salmonoid fishes (Actinopterygii: Salmonoidei) and their evolution. *Gene* 434, 16–28.

Mayer, W. E., Schuster, L. N., Bartelmes, G., Dieterich, C., and Sommer, R. J. (2011). Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evol. Biol.* 11, 13.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216.

Medina, M., Collins, A. G., Silberman, J. D., and Sogin, M. L. (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc. Natl. Acad. Sci. U.S.A.* 98, 9707–9712.

Miller, S. A., and Harley, J. P. (2007). *Zoology*, 7th Edn. New York, NY: McGraw-Hill Higher Education, 297.

Mizrokhi, L. J., and Mazo, A. M. (1990). Evidence for horizontal transmission of the mobile element jockey between distant Drosophila species. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9216–9220.

Muller, H. J. (1932). Some genetic aspects of sex. *Am. Nat.* 66, 118–138.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* 106, 2–9.

Muller, W. E. G. (2003). The origin of metazoan complexity: Porifera as integrated animals. *Integr. Comp. Biol.* 43, 3–10.

Moran, N. A., and Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328, 624–627.

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G., and Sogin, M. L. (2007). Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* 317, 1921–1926.

Nedelcu, A. M., Miles, I. H., Fagir, A. M., and Karol, K. (2008). Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *J. Evol. Biol.* 21, 1852–1860.

Nielsen, C. (2001). *Animal Evolution, Interrelationships of the Living Phyla*, 2nd edn. Oxford: Oxford University Press, 563.

Nikoh, N., McCutcheon, J. P., Kudo, T., Miyagishima, S. Y., Moran, N. A., and Nakabachi, A. (2010). Bacterial genes in the aphid genome: absence of functional gene transfer from Buchnera to its host. *PLoS Genet*. 6:e1000827. doi: 10.1371/journal.pgen.1000827

Ochman, H., Lawrence, J. G., and Grolsman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Pace, J. K. I. I., Gilbert, C., Clark, M. S., and Feschotte, C. (2008). Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17023–17028.

Pechenik, J. A. (2005). *Biology of the Invertebrates.* Boston: McGraw-Hill, Higher Education, 178.

Philippe, J. (1997). "*Craniata. Animals with skulls,*" *in The Tree of Life Web Project*, http://tolweb.org.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., and Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19, 706–712.

Peterson, K. J., and Butterfield, N. J. (2005). Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9547–9552.

Peterson, K. J., and Eernisse, D. J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol. Dev.* 3, 170–205.

Price, T. D., and Bouvier, M. M. (2002). The evolution of F1 postzygotic incompatibilities in birds. *Evolution* 56, 2083–2089.

Priimagi, A. F., Mizrokhi, L. J., and Ilyin, Y. V. (1988). The *Drosophila* mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. *Gene* 70, 253–262.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J. Biol.* 8, 59.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* 2, 745–756.

Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609.

Ragan, M. A., McInerney, J. O., and Lake, J. A. (2009). The network of life: genome beginnings and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2169–2175.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Richards, T. A., Soanes, D. M., Jones, M. D., Vasieva, O., Leonard, G., Paszkiewicz, K., Foster, P. G., Hall, N., and Talbot, N. J. (2011). Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15258–15263.

Rivera, M. C., and Lake, J. A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155.

Robertson, H. M. (1997). Multiple mariner transposons in flatworms and hydras are related to those of insects. *J. Hered.* 88, 195–201.

Robertson, H. M., and Lampe, D. J. (1995). Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. *Mol. Biol. Evol.* 12, 850–862.

Robertson, H. M., and MacLeod, E. G. (1993). Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol. Biol.* 2, 125–139.

Robertson, H. M., Soto-Adames, F. N., Walden, K. K., Avancini, R. M., and Lampe, D. J. (2002). "The mariner transposons of animals: horizontally jumping genes," in *Horizontal Gene Transfer,* eds M. Syvanen and C. I. Kado (San Diego, CA: Academic Press), 173–183.

Robinson, C. J., Bohannan, B. J., and Young, V. B. (2010). From structure to function: the ecology of

host-associated microbial communities. *Microbiol. Mol. Biol. Rev.* 74, 453–476.

Rödelsperger, C., and Sommer, R. J. (2011). Computational archaeology of the Pristionchus pacificus genome reveals evidence of horizontal gene transfers from insects. *BMC Evol. Biol.* 11, 239.

Rogers, J. (1985). The origin and evolution of retroposons. *Int. Rev. Cytol.* 93, 187–279.

Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.

Rot, C., Goldfarb, I., Ilan, M., and Huchon, D. (2006). Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evol. Biol.* 6, 71.

Roule, L. (1891). Considerations sur l'embranchement des Trochozoaires. *Ann. Sci. Nat. (Zool.) 7me Série* 11, 121–178.

Ruppert, E. E., Fox, R. S., and Barnes, R. D. (2004). *Invertebrate Zoology,* 7th edn. Toronto: Thomson Brooks/Cole.

Saunders, N. J., Hood, D. W., and Moxon, E. R. (1999). Bacterial evolution: bacteria play pass the gene. *Curr. Biol.* 9, R180–R183.

Schliep, K., Lopez, P., Lapointe, F. J., and Bapteste, E. (2010). Harvesting evolutionary signals in a forest of prokaryotic gene trees. *Mol. Biol. Evol.* 28, 1393–1405.

Schmidt-Rhaesa, A., Bartolomaeus, T., Lemburg, C., Ehlers, U., and Garey, J. R. (1998). The position of the Arthropoda in the phylogenetic system. *J. Morphol.* 238, 263–285.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207.

Segers, H. (2007). Annotated checklist of the rotifers (Phylum Rotifera), with notes on nomenclature, taxonomy and distribution. *Zootaxa* 1564, 1–104.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244.

Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1443–1448.

Sneath, P. H. A. (1975). "Cladistic representation of reticulate evolution," in *Systematic Zoology*, (London: Taylor and Francis, Ltd.), 360–368.

Snel, B., Bork, P., and Huynen, M. A. (1999). Genome phylogeny based on gene content. *Nat. Genet.* 21, 108–110.

Syvanen, M., and Ducore, J. (2010). Whole genome comparisons reveals a possible chimeric origin for a major Metazoan assemblage. *J. Biol. Syst.*18, 261–275.

Tekaia, F., Lazcano, A., and Dujon, B. (1999). The genomic tree as revealed from whole proteome comparisons. *Genome Res.* 9, 550–557.

Telford, M. J., Bourlat, S. J., Economou, A., Papillon, D., and Rota Stabelli, O. (2008). The evolution of the Ecdysozoa. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1529–1537.

Terzian, C., Ferraz, C., Demaille, J., and Bucheton, A. (2000). Evolution of the Gypsy endogenous retrovirus in the *Drosophila melanogaster* subgroup. *Mol. Biol. Evol.* 17, 908–914.

Thomas, J., Schaack, S., and Pritham, E. J. (2010). Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol. Evol.* 2, 656–664.

Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* 74, 3715–3730.

Udomkit, A., Forbes, S., Dalgleish, G., and Finnegan, D. J. (1995). BS a novel LINE-like element in *Drosophila melanogaster*. *Nucleic Acids Res.* 23, 1354–1358.

Valentine, J. W. (1997). Cleavage patterns and the topology of the metazoan tree of life. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8001–8005.

Valentine, J. W. (2004). *On the Origin of Phyla.* Chicago: University of Chicago Press, 33.

van Oppen, M. J. H., Catmull, J., McDonald, B. J., Hislop, N. R., Hagerman, P. J., and Miller, D. J. (2002). The mitochondrial genome of Acropora tenuis (Cnidaria: Scleractinia) contains a large group I intron and a candidate control region. *J. Mol. Evol.* 55, 1–13.

Vazquez-Manrique, R. P., Hernandez, M., Martínez-Sebastián, M. J., and de Frutos, R. (2000). Evolution of gypsy endogenous retrovirus in the *Drosophila obscura* species group. *Mol. Biol. Evol.* 17, 1185–1193.

Volff, J. N., Korting, C., and Schartl, M. (2000). Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes. *Mol. Biol. Evol.* 17, 1673–1684.

Wallace, R. L. (1998). "Rotifera," in *Encyclopedia of Reproduction*, eds E. Knobil and J. D. Neil (San Diego: Academic Press), 118–129.

Wallace, R. L., Snell, T. W., and Ricci, C. (2006). "Rotifera, biology, ecology and systematics," in *Guides to the Identification of the Microinvertebrates of the Continental Waters of the World*, ed Segers H. Ghent. (Kenobi Productions, Leiden: Backhuys Publishers), 23.

Watkins, R. F., and Gray, M. W. (2006). The frequency of eubacterium-to-eukaryote lateral gene transfer shows significant cross-taxa variation within Amoebozoa. *J. Mol. Evol.* 63, 801–814.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55, 631–661.

Wessler, S. R. (1998). Transposable elements and the evolution of gene expression. *Symp. Soc. Exp. Biol.* 51, 115–122.

Winchell, C. J., Sullivan, J., Cameron, C. B., Swalla, B. J., and Mallatt, J. (2002). Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol. Biol. Evol.* 19, 762–776.

Woese, C. R. (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8742–8874.

Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., Henrissat, B., Coutinho, P. M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R. S., Fulton, L. A., Clifton, S. W., Wilson, R. K., Knight, R. D., and Gordon, J. I. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* 5:e156. doi: 10.1371/journal.pbio.0050156

Zhu, B., Lou, M. M., Xie, G. L., Zhang, G. Q., Zhou, X. P., Li, B., and Jin, G. L. (2011). Horizontal gene transfer in silkworm, *Bombyx mori*. *BMC Genomics* 12, 248.

Zillig, W., Palm, P., and Klenk, H. P. (1992). "A model of the early evolution of organisms: the arisal of the three domains of life from the common ancestor," in *The Origin and Evolution of the Cell,* eds H. Hartman and K. Matsuno (Singapore: World Scientific Publishing), 163–182.

Zinner, D., Arnold, M. L., and Roos, C. (2011). The strange blood: natural hybridization in primates. *Evol. Anthropol.* 20, 96–103.

Zrzavý, J. (2003). Gastrotricha and metazoan phylogeny. *Zool. Scr.* 32, 61–81.

Zupunski, V., Gubensek, F., and Kordis, D. (2001). Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.* 18, 1849–1863.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 February 2012; accepted: 23 March 2012; published online: 11 April 2012.

*Citation: Ramulu HG, Raoult D and Pontarotti P (2012) The rhizome of life: what about metazoa?. Front. Cell. Inf. Microbio. **2**:50. doi: 10.3389/fcimb.2012.00050*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

MINI REVIEW ARTICLE
published: 06 March 2012
doi: 10.3389/fcimb.2012.00027

Lateral gene transfers have polished animal genomes: lessons from nematodes

Etienne G. J. Danchin [1,2,3]* *and* ***Marie-Noëlle Rosso*** [1,2,3†]

[1] *Institut National de la Recherche Agronomique, UMR 1355 ISA, 400 route des Chappes, Sophia-Antipolis, France*
[2] *Centre National de la Recherche Scientifique, UMR 7254 ISA, 400 route des Chappes, Sophia-Antipolis, France*
[3] *Université de Nice-Sophia Antipolis, UMR ISA, 400 route des Chappes, Sophia-Antipolis, France*

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Kalliopi Georgiades, Unité de Recherche en maladies infectieuses tropicales emergentes, France
Pierre Pontarotti, Centre National de la Recherche Scientifique, France

****Correspondence:***
Etienne G. J. Danchin, Institut Sophia Agrobiotech, INRA-CNRS-UNSA, 400 route des Chappes, F-06903 Sophia-Antipolis, France.
e-mail: etienne.danchin@sophia.inra.fr

† Present address:
INRA, Université Aix-Marseille, UMR1163 Biotechnologie des Champignons Filamenteux, F-13288 Marseille, France.

It is now accepted that lateral gene transfers (LGT), have significantly contributed to the composition of bacterial genomes. The amplitude of the phenomenon is considered so high in prokaryotes that it challenges the traditional view of a binary hierarchical tree of life to correctly represent the evolutionary history of species. Given the plethora of transfers between prokaryotes, it is currently impossible to infer the last common ancestral gene set for any extant species. For this ensemble of reasons, it has been proposed that the Darwinian binary tree of life may be inappropriate to correctly reflect the actual relations between species, at least in prokaryotes. In contrast, the contribution of LGT to the composition of animal genomes is less documented. In the light of recent analyses that reported series of LGT events in nematodes, we discuss the importance of this phenomenon in the evolutionary history and in the current composition of an animal genome. Far from being neutral, it appears that besides having contributed to nematode genome contents, LGT have favored the emergence of important traits such as plant-parasitism.

Keywords: lateral gene transfer, horizontal gene transfer, genome, nematode, eukaryote, adaptation, evolution

INTRODUCTION

Lateral gene transfer (LGT), the transmission of genetic material from one species to another by means other than direct inheritance from parents to the offspring is a prevalent phenomenon in prokaryotes. LGT contribution to the composition of today bacterial genomes is so high that it challenges the idea that genome evolution in these species follows a tree-like pattern. Besides contributing to the composition of their genomes LGT events also play significant role in the acquisition of important biological traits. For instance, antibiotic resistance is frequently acquired during the lifespan of a bacterium and transmitted to its offspring rather than inherited from an ancient ancestor. Whether LGT events have significantly contributed to the composition of metazoan genomes and whether they are involved in the acquisition of new traits remains poorly documented, in contrast. This is probably in part because LGT in these species appears conceptually and mechanistically less evident than for prokaryotes. Here, we discuss an ensemble of LGT cases reported in different nematode lineages with different lifestyles, including necromenic, plant-parasitic, and animal-parasitic species. Although it is currently difficult to evaluate the contribution of LGT to the composition of nematode genomes, in the lack of a comprehensive whole genome scan, we can already assume that LGT have probably played important role in the acquisition of some traits such as plant-parasitism.

LGT IN NECROMENIC NEMATODES

The necromenic nematode *Pristionchus pacificus* spends a considerable part of its life cycle associated with beetles and is supposed to feed on bacteria and fungi that decompose the insect once it is dead. The genome sequencing of *P. pacificus* revealed the presence of functional cellulase genes acquired via LGT of microbial origin (Dieterich et al., 2008). A more detailed analysis of these cellulase genes in the *Pristionchus* genus showed that the genes were probably acquired ancestrally as different *Pristionchus* species possessed these genes and their phylogeny matched the species phylogeny (Mayer et al., 2011). The same analysis also showed that these genes underwent high rates of duplications and losses suggesting that the number of cellulase genes in *Pristionchus* species is under selection. Interestingly, the genome of *P. pacificus* also revealed the presence of Diapausin genes otherwise absent from nematodes but present in beetles. A phylogenetic analysis coupled with multi-species comparison of codon usage bias, revealed that Diapausin genes as well as ca. 500 others in the *P. pacificus* genome might have been acquired via LGT of insect origin (Rodelsperger and Sommer, 2011). Most of the putatively transferred genes of insect origin were retrotransposons, and it is hypothesized that these mobile elements might have played themselves a role of vector in the transfer of other genes. These first clear examples of putative transfers between two metazoan animals indicate that our current view of the amplitude of the LGT phenomenon in

animals might be underestimated because most analyzes have been so far restricted to transfers of non-metazoan origin.

LGT IN ANIMAL-PARASITIC NEMATODES

The mutualistic symbiosis between alpha-proteobacteria Wolbachia and Onchocercids has for long stimulated scientific interest due to the possibility to control these causing agents of severe human diseases, such as river blindness and elephantiasis by antibacterial treatments. The interdependence between the symbiont and its host seems to involve the supply by the nematode of amino acids required for Wolbachia growth and the ability of the symbiont to complete the synthesis of compounds crucial to the nematode such as heme, riboflavin and nucleotides (Foster et al., 2005; Ghedin et al., 2007). Several LGTs from endosymbiont to the nematode genome have been described (Fenn et al., 2006; Dunning Hotopp et al., 2007; McNulty et al., 2010) and may have been facilitated by the close association of the bacteria to germline cells in the female stage (Ferri et al., 2011; Fischer et al., 2011). Until recently, it was generally assumed that the presence of the bacterial symbiont in Onchocercids could result from a single colonization event in an ancestor of the lineage and that the bacteria was secondarily lost in a few nematode species able to proliferate free from bacteria. The presence of transcriptionally active bacterial genes in the genomes of two distantly related Wolbachia-free Onchocercids species indicates that these species might have acquired bacterial genes before symbiont loss (McNulty et al., 2010). However, evidence for a recent Wolbachia capture in an Onchocercid species associated with the Japanese black bear modifies the current appraisal on the co-evolution of the endosymbiont and its host (Ferri et al., 2011). In addition, a screen for the presence of the bacteria in yet unexplored genera and species identified 63% of Onchocercids devoid of Wolbachia, notably the ancestral Oswaldofilariinae (Ferri et al., 2011). It will be most interesting to screen the genomes of these bacteria-free nematodes to search for traces of LGT and assess whether they lost their symbiont after acquisition of bacterial genes.

LGT IN PLANT-PARASITIC NEMATODES

The ability to feed on plants arose at least three times independently in the phylum Nematoda (Blaxter et al., 1998). How nematodes evolved toward plant-parasitism is a fascinating question. All phytonematodes feed from the cytoplasm of living plant cells. Ectoparasites perforate plant tissues with a protrusible stylet that reaches the cell layers they feed on whereas endoparasites penetrate plant tissues. Both ecto- and endoparasites, therefore, require active degradation or softening of the physical barrier formed by plant cell walls. The endoparasitic cyst and root-knot nematodes are among the most notorious phytonematodes, causing major damages to crops worldwide. The identification in cyst nematodes of genes encoding cellulases most similar to bacterial enzymes paved the way for the assumption that LGT could have participated in adaptation to plant-parasitism (Smant et al., 1998). Since then, several nematode genes encoding cell wall-degrading or -modifying enzymes were presented as candidates for LGT (Davis et al., 2000). The analysis of root-knot nematode genomes revealed an unprecedented repertoire of cell wall-degrading and -modifying enzymes in an animal that covered six different protein families able to degrade all major polysaccharides of cell walls, i.e., cellulose, hemicellulose and pectin (Abad et al., 2008; Opperman et al., 2008). Homologs for these nematode cell wall-degrading enzymes were searched in public databases, checked for significance and used for systematic robust phylogenetic reconstruction (Danchin et al., 2010). Two noticeable findings from this analysis were that those genes most likely originated from several independent LGT events and that in most cases the closest relatives were found in bacteria from the rhizosphere. Indeed, LGT events have been traced back at different nodes of the nematoda phylogeny. Some LGT events appear so far lineage-specific, for instance, GH28 polygalacturonases were only found to date in root-knot nematodes. In contrast, other genes such as those encoding GH5 cellulases or PL3 pectate lyases are found in multiple plant-parasitic nematode lineages, suggesting a more ancient acquisition in a common ancestor. Interestingly, several different bacterial species appear as potential donors for the different families of plant cell wall-degrading enzymes. For example, the closest relatives of root-knot nematode GH28s are found in *Ralstonia solanacearum*, a plant-pathogenic soil bacterium that shares plant hosts with these nematodes. On the other hand, nematode PL3 pectate lyases cluster with PL3s of *Clavibacter michiganensis*, another plant pathogen evolutionary distant from *Ralstonia solanacearum*. Overall, this analysis suggests that the ecology of the donor is probably more important than its phylogenetic position for a gene transfer to occur. This is consistent with similar conclusions drawn from the analysis of LGT events between bacteria of the human microbiome (Smillie et al., 2011). Another interesting feature of genes encoding cell wall-degrading enzymes in nematodes is their frequent organization in large multigene families. As many as 30 pectate lyases, 21 cellulases and 20 expansin-like genes were identified in the genome of *M. incognita* (Abad et al., 2008) and similar gene family expansions were observed in the genome of *M. hapla* (Opperman et al., 2008). Gene abundance in these families suggest that gene duplications have been under positive selection during the evolution of these plant-parasites.

Several experimental data further indicate that LGT could have participated to adaptation to plant-parasitism. First, transcripts for all cell wall-degrading enzymes have been localized in the esophageal glands of the nematodes (Rosso et al., 2012). These oesophageal glands are specialized cells in which proteins secreted during infection are produced. In line with this finding, the systematic presence of a secretion signal peptide on the predicted protein sequences suggested a role for the enzymes outside the nematode during parasitism. Second, cellulases, pectate lyases, polygalacturonases, and expansin-like proteins were enzymatically active, although examples of potential pseudogenization were identified in migratory nematodes (Bera-Maillet et al., 2000; Popeijus et al., 2000; Jaubert et al., 2002; Qin et al., 2004; Mitreva-Dautova et al., 2006; Haegeman et al., 2010). Immunolocalization studies showed the secretion of cellulases and pectate lyases *in planta* during migration of the juveniles and comforted a role in cell wall degradation or cell wall softening during host invasion (Wang et al., 1999; Doyle and Lambert, 2002; Vieira et al., 2011). Moreover, immunolocalizations in root-knot nematode

females provided a hint for an additional role of cellulases during egg laying at the surface of the root (Vieira et al., 2011). Finally, the reduced virulence of nematodes after knock-down of cellulase genes established the role of these enzymes in the success of infection (reviewed in Rehman et al., 2009; Rosso et al., 2009; Haegeman et al., 2009a).

Besides the clear examples of genes encoding cell wall-modifying enzymes, a series of other genes have been acquired by LGT in phytonematodes and also probably contributed to the emergence and success of plant-parasitism in these species (reviewed in Haegeman et al., 2011). The processes in which the gene products are putatively involved include crucial mechanisms of the parasitic interaction, such as establishment of the nematode's feeding structure, modulation of plant defense pathways and processing of nutrients absorbed from the plant. As for genes encoding cell wall-degrading enzymes, the principal putative donor species for these genes are soil bacteria but for a few other genes fungal species appear as the most likely donors.

CONCLUDING REMARKS

This bunch of indications for LGT occurrences in nematodes adds to other recent findings on LGT in eukaryotes and modifies our viewpoint on the prevalence of gene transfers in animals (for reviews see Keeling and Palmer, 2008; Dunning Hotopp, 2011). In plant parasitic nematodes, several LGT events occurred at different time points of their evolution. In addition, the transferred genes seem to originate from different donor organisms. However, any conclusion on the multiplicity of potential donors should be considered with caution as our view on gene transfers frequency between prokaryotes is evolving rapidly. Identification of the exact donor bacteria may be complicated by several aspects. The true donor bacteria might not exist anymore or might not have been characterized at the sequence level. In these cases the next bacteria sharing the highest similarity with the studied gene might be considered as donor by default. Also, due to the high frequency of transfers between bacteria, we cannot exclude that the gene transferred from a bacterium to a nematode had itself been already transferred before from another bacteria. Indeed, genes transferred multiple times have been described in proteobacteria (Kloesges et al., 2011). The observed transfer might thus be a secondary event.

How could foreign genes be integrated into the genome of nematodes is certainly the main black box here. Gene transfers from the endosymbiont *Wolbachia* in Onchocercids can be facilitated by the presence of the bacteria in germ cells, though the exact genetic mechanism of transfer is unknown.

Noticeably, no symbiont has been identified so far in the root-knot and cyst nematodes mentioned above. Still, ancestors of phytonematodes could have held bacterial symbionts, as suggested by the presence of an endosymbiotic bacterium from the *Wolbachia* supergroup within the reproductive tract of *Radopholus similis* females, plant-parasitic nematodes that share a common ancestor with cyst- and root-knot nematodes (Haegeman et al., 2009b). However, in the case of genes encoding plant cell wall-degrading enzymes, an origin from an ancient Wolbachia-like symbiont appears highly unlikely because in no phylogenetic analysis was a Wolbachia or Wolbachia-like sequence identified as an homolog (Danchin et al., 2010).

The two insect parasitic nematodes Steinernema and Heterorhabditis provide another example of close association between nematodes and bacteria. Gamma-proteobacteria dwell in the gut of infective juveniles of these two nematodes and in rectal glands of Heterorhabditis females (Ciche and Ensign, 2003). The genome of these nematodes is not yet available and if new cases of LGT of bacterial origin were identified here, they could indicate the possibility of transfer to the germline cells without direct contact with the bacteria.

An alternative hypothesis for the origin of LGT events is the "you are what you eat" hypothesis first formulated in (Doolittle, 1998). In root-knot nematodes, four different groups of bacteria can be viewed as potential donors of plant cell wall-degrading enzyme genes. Interestingly, three of these groups include notorious plant pathogens or symbionts having interactions within plant hosts and thereby sympatric with these nematodes (Danchin et al., 2010). One possibility is that an ancestor of plant-parasitic nematodes was initially bacterivorous and used to feed on soil bacteria, including plant-pathogenic bacteria (**Figure 1**). By mechanisms discussed later, genetic material from digested bacteria might have been transferred to the nuclear genome of nematodes. In the case of cell wall-degrading enzymes,

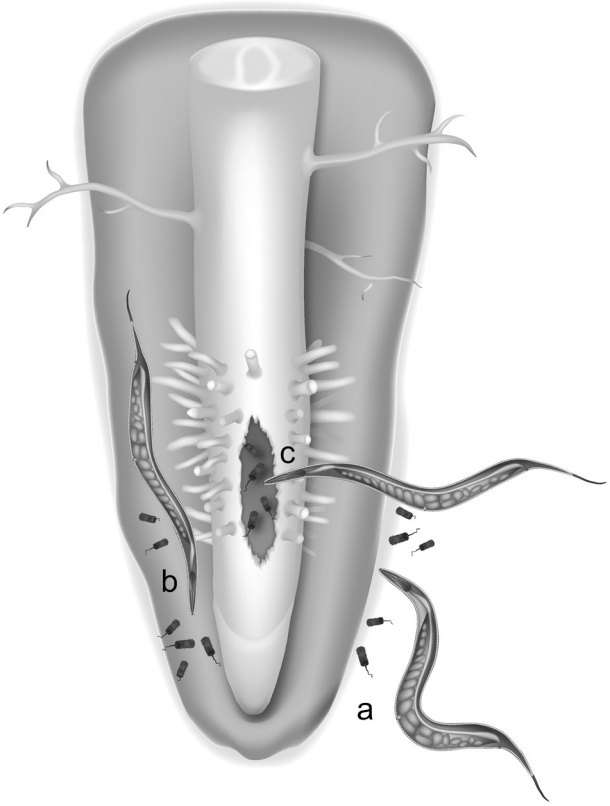


FIGURE 1 | An hypothesis for lateral gene transfer from bacteria to plant-parasitic nematodes. LGT could have been favored in ancestral nematodes that shared their ecological niche with plant-associated bacteria in soil **(a)**, in the rhizosphere **(b)** or in bacteria-infected plant tissues **(c)**.

the transferred genes likely provided the nematode the ability to penetrate plant tissue and in a first time to access root-dwelling bacteria otherwise inaccessible to other bacterivorous nematodes. This selective advantage might have favored individuals that possessed bacterial genes encoding cell wall-degrading enzymes and following a LGT ratchet mechanism, individuals with additional enzymes acquired via LGT might have been positively selected generation after generation and eventually developed a plant-parasitic lifestyle.

Whatever the origin, from ancient endosymbionts or from feeding, different potential mechanisms of transfer can be envisioned. They may involve bacterial secretion systems, bacteriophages, plasmids, or viruses as well as passive or active spread of DNA fragments from digested bacterial cells (Danchin, 2011). Transposable elements can also be considered as potential vectors, particularly in the case of Pristionchus in which retrostransposons have been found associated with transferred genes. In contrast, no evidence for a proximity of transposable elements in the vicinity of transferred genes in plant-parasitic nematodes has been described.

Overall, it appears that LGT, both from bacterial and fungal origins have not only contributed to the current composition of nematode gene repertoires but also had crucial importance in the emergence of new biological capabilities. Although several examples of LGT of non-metazoan origin in nematode genomes have been reported, the total contribution of LGT to the composition of a nematode genome is yet poorly described. To date, no comprehensive list of genes putatively acquired via LGT in a nematode genome has been published. It is also largely unknown whether transfers of animal or plant origin substantially contributed to the composition of nematode genomes. The recent example of transfers from insects to necromenic nematodes suggests this might actually be the case.

The idea that a Darwinian binary tree-like representation incorrectly reflects the evolutionary history of genes and genomes has been raised as soon as 1975 (Sneath, 1975). Based on the observation that LGT occur frequently in bacteria and that reticulate evolution, another phenomenon challenging binary tree-like representation is common in flowering plants, Sneath already proposed that a network-like representation might be more accurate. This idea has found echoes more recently in the light of whole genome analysis, suggesting that, at least in bacteria, evolution more resembles a rhizome than a bifurcating tree (Raoult, 2010). We cannot state at the moment to what extent could LGT events disturb a binary tree-like representation in the phylum nematoda. Regardless the potential contribution of LGT to the composition of nematode genomes, it appears clear, at least in the case of plant-parasitic nematodes that these events have contributed to adaptation to a new life style.

ACKNOWLEDGMENTS

This work benefited from links funded through European COST Action 872.

REFERENCES

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., Van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., and Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* 26, 909–915.

Bera-Maillet, C., Arthaud, L., Abad, P., and Rosso, M. N. (2000). Biochemical characterization of MI-ENG1, a family 5 endoglucanase secreted by the root-knot nematode *Meloidogyne incognita*. *Eur. J. Biochem.* 267, 3255–3263.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., and Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75.

Ciche, T. A., and Ensign, J. C. (2003). For the insect pathogen Photorhabdus luminescens, which end of a nematode is out? *Appl. Environ. Microbiol.* 69, 1890–1897.

Danchin, É. G. J. (2011). What Nematode genomes tell us about the importance of horizontal gene transfers in the evolutionary history of animals. *Mobile Genet. Elem.* 1, 1–5.

Danchin, E. G., Rosso, M. N., Vieira, P., De Almeida-Engler, J., Coutinho, P. M., Henrissat, B., and Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17651–17656.

Davis, E. L., Hussey, R. S., Baum, T. J., Bakker, J., Schots, A., Rosso, M. N., and Abad, P. (2000). Nematode parasitism genes. *Annu. Rev. Phytopathol.* 38, 365–396.

Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S. P., Wilson, R. K., and Sommer, R. J. (2008). The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat. Genet.* 40, 1193–1198.

Doolittle, W. F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 14, 307–311.

Doyle, E. A., and Lambert, K. N. (2002). Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. *Mol. Plant Microbe Interact.* 15, 549–556.

Dunning Hotopp, J. C. (2011). Horizontal gene transfer between bacteria and animals. *Trends Genet.* 27, 157–163.

Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., and Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756.

Fenn, K., Conlon, C., Jones, M., Quail, M. A., Holroyd, N. E., Parkhill, J., and Blaxter, M. (2006). Phylogenetic relationships of the Wolbachia of nematodes and arthropods. *PLoS Pathog.* 2:e94. doi: 10.1371/journal.ppat.0020094

Ferri, E., Bain, O., Barbuto, M., Martin, C., Lo, N., Uni, S., Landmann, F., Baccei, S. G., Guerrero, R., De Souza Lima, S., Bandi, C., Wanji, S., Diagne, M., and Casiraghi, M. (2011). New insights into the evolution of Wolbachia infections in filarial nematodes inferred from a large range of screened species. *PLoS One* 6:e20843. doi: 10.1371/journal.pone.0020843

Fischer, K., Beatty, W. L., Jiang, D., Weil, G. J., and Fischer, P. U. (2011). Tissue and stage-specific distribution of Wolbachia in Brugia malayi. *PLoS Negl. Trop. Dis.* 5:e1174. doi: 10.1371/journal.pntd.0001174

Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., Vincze, T., Ingram, J., Moran, L., Lapidus, A., Omelchenko, M., Kyrpides, N., Ghedin, E., Wang, S., Goltsman, E., Joukov, V., Ostrovskaya, O., Tsukerman, K., Mazur, M., Comb, D., Koonin, E., and Slatko, B. (2005). The Wolbachia genome of

Brugia malayi: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* 3:e121. doi: 10.1371/journal.pbio.0030121

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J. E., Delcher, A. L., Guiliano, D. B., Miranda-Saavedra, D., Angiuoli, S. V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N. M., Wortman, J. R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S. L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G. J., Carlow, C. K., Crawford, M. J., Daub, J., Dimmic, M. W., Estes, C. F., Foster, J. M., Ganatra, M., Gregory, W. F., Johnson, N. M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B. W., Li, W., Lindblom, T. H., Lustigman, S., Ma, D., Maina, C. V., Martin, D. M., McCarter, J. P., McReynolds, L., Mitreva, M., Nutman, T. B., Parkinson, J., Peregrin-Alvarez, J. M., Poole, C., Ren, Q., Saunders, L., Sluder, A. E., Smith, K., Stanke, M., Unnasch, T. R., Ware, J., Wei, A. D., Weil, G., Williams, D. J., Zhang, Y., Williams, S. A., Fraser-Liggett, C., Slatko, B., Blaxter, M. L., and Scott, A. L. (2007). Draft genome of the filarial nematode parasite Brugia malayi. *Science* 317, 1756–1760.

Haegeman, A., Jones, J. T., and Danchin, E. G. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol. Plant Microbe Interact.* 24, 879–887.

Haegeman, A., Kyndt, T., and Gheysen, G. (2010). The role of pseudo-endoglucanases in the evolution of nematode cell wall-modifying proteins. *J. Mol. Evol.* 70, 441–452.

Haegeman, A., Vanholme, B., and Gheysen, G. (2009a). Characterization of a putative endoxylanase in the migratory plant-parasitic nematode *Radopholus similis*. *Mol. Plant Pathol.* 10, 389–401.

Haegeman, A., Vanholme, B., Jacob, J., Vandekerckhove, T. T., Claeys, M., Borgonie, G., and Gheysen, G. (2009b). An endosymbiotic bacterium in a plant-parasitic nematode: member of a new Wolbachia supergroup. *Int. J. Parasitol.* 39, 1045–1054.

Jaubert, S., Laffaire, J. B., Abad, P., and Rosso, M. N. (2002). A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *FEBS Lett.* 522, 109–112.

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618.

Kloesges, T., Popa, O., Martin, W., and Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.* 28, 1057–1074.

Mayer, W. E., Schuster, L. N., Bartelmes, G., Dieterich, C., and Sommer, R. J. (2011). Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evol. Biol.* 11, 13.

McNulty, S. N., Foster, J. M., Mitreva, M., Dunning Hotopp, J. C., Martin, J., Fischer, K., Wu, B., Davis, P. J., Kumar, S., Brattig, N. W., Slatko, B. E., Weil, G. J., and Fischer, P. U. (2010). Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One* 5:e11029. doi: 10.1371/journal.pone.0011029

Mitreva-Dautova, M., Roze, E., Overmars, H., De Graaff, L., Schots, A., Helder, J., Goverse, A., Bakker, J., and Smant, G. (2006). A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. *Mol. Plant Microbe Interact.* 19, 521–529.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., and Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14802–14807.

Popeijus, H., Overmars, H., Jones, J., Blok, V., Goverse, A., Helder, J., Schots, A., Bakker, J., and Smant, G. (2000). Degradation of plant cell walls by a nematode. *Nature* 406, 36–37.

Qin, L., Kudla, U., Roze, E. H., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J. T., Schots, A., Smant, G., Bakker, J., and Helder, J. (2004). Plant degradation: a nematode expansin acting on plants. *Nature* 427, 30.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Rehman, S., Butterbach, P., Popeijus, H., Overmars, H., Davis, E. L., Jones, J. T., Goverse, A., Bakker, J., and Smant, G. (2009). Identification and characterization of the most abundant cellulases in stylet secretions from Globodera rostochiensis. *Phytopathology* 99, 194–202.

Rodelsperger, C., and Sommer, R. J. (2011). Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. *BMC Evol. Biol.* 11, 239.

Rosso, M. N., Hussey, R. S., Davis, E. L., Smant, G., Baum, T., Abad, P., and Mitchum, M. (2012). "Nematode effector proteins: targets and functions in plant parasitism," in *Effectors in Plant-Microbe Interactions*, eds F. Martin and S. Kammoun (Chichester, UK: Wiley-Blackwell), 327–354.

Rosso, M. N., Jones, J. T., and Abad, P. (2009). RNAi and functional genomics in plant parasitic nematodes. *Annu. Rev. Phytopathol.* 47, 207–232.

Smant, G., Stokkermans, J. P., Yan, Y., De Boer, J. M., Baum, T. J., Wang, X., Hussey, R. S., Gommers, F. J., Henrissat, B., Davis, E. L., Helder, J., Schots, A., and Bakker, J. (1998). Endogenous cellulases in animals: isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4906–4911.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244.

Sneath, P. H. A. (1975). Cladistic representation of reticulate evolution. *Syst. Zool.* 24, 360–368.

Vieira, P., Danchin, E. G., Neveu, C., Crozat, C., Jaubert, S., Hussey, R. S., Engler, G., Abad, P., De Almeida-Engler, J., Castagnone-Sereno, P., and Rosso, M. N. (2011). The plant apoplasm is an important recipient compartment for nematode secreted proteins. *J. Exp. Bot.* 62, 1241–1253.

Wang, X., Meyers, D., Yan, Y., Baum, T., Smant, G., Hussey, R., and Davis, E. (1999). In planta localization of a beta-1,4-endoglucanase secreted by *Heterodera glycines*. *Mol. Plant Microbe Interact.* 12, 64–67.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 January 2012; paper pending published: 14 February 2012; accepted: 21 February 2012; published online: 06 March 2012.

Citation: Danchin EGJ and Rosso M-N (2012) Lateral gene transfers have polished animal genomes: lessons from nematodes. Front. Cell. Inf. Microbio. 2:27. doi: 10.3389/fcimb.2012.00027

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 26 July 2012
doi: 10.3389/fcimb.2012.00098

Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution

Vaibhav Bhandari, Hafiz S. Naushad and Radhey S. Gupta*

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada

Edited by:
Yousef Abu Kwaik, University of Louisville School of Medicine, USA

Reviewed by:
Srinand Sreevastan, University of Minnesota, USA
Andrey P. Anisimov, State Research Center for Applied Microbiology and Biotechnology, Russia

***Correspondence:**
Radhey S. Gupta, Department of Biochemistry and Biomedical Sciences, McMaster University, 1200 Main Street West, Health Sciences Center, Hamilton, ON L8N 3Z5, Canada.
e-mail: gupta@mcmaster.ca

The analyses of genome sequences have led to the proposal that lateral gene transfers (LGTs) among prokaryotes are so widespread that they disguise the interrelationships among these organisms. This has led to questioning of whether the Darwinian model of evolution is applicable to prokaryotic organisms. In this review, we discuss the usefulness of taxon-specific molecular markers such as conserved signature indels (CSIs) and conserved signature proteins (CSPs) for understanding the evolutionary relationships among prokaryotes and to assess the influence of LGTs on prokaryotic evolution. The analyses of genomic sequences have identified large numbers of CSIs and CSPs that are unique properties of different groups of prokaryotes ranging from phylum to genus levels. The species distribution patterns of these molecular signatures strongly support a tree-like vertical inheritance of the genes containing these molecular signatures that is consistent with phylogenetic trees. Recent detailed studies in this regard on the Thermotogae and Archaea, which are reviewed here, have identified large numbers of CSIs and CSPs that are specific for the species from these two taxa and a number of their major clades. The genetic changes responsible for these CSIs (and CSPs) initially likely occurred in the common ancestors of these taxa and then vertically transferred to various descendants. Although some CSIs and CSPs in unrelated groups of prokaryotes were identified, their small numbers and random occurrence has no apparent influence on the consistent tree-like branching pattern emerging from other markers. These results provide evidence that although LGT is an important evolutionary force, it does not mask the tree-like branching pattern of prokaryotes or understanding of their evolutionary relationships. The identified CSIs and CSPs also provide novel and highly specific means for identification of different groups of microbes and for taxonomical and biochemical studies.

Keywords: conserved indels, signature proteins, phylogenetic trees, lateral gene transfers, Thermotogae, Archaea, Crenarchaeota, RpoB signatures

INTRODUCTION

The understanding of prokaryotic relationships is one of the most important goals of evolutionary sciences. These relationships have been difficult to understand due to the simplicity and antiquity of prokaryotic organisms and disagreements in viewpoints among evolutionary biologists regarding the importance of different factors when grouping prokaryotes. Although earlier studies in this regard were based on morphology or physiology (Cowan, 1965; Buchanan and Gibbons, 1974; Stanier et al., 1976), the field itself has evolved to account for new information brought about by technological or informational breakthroughs, viz. molecular data, DNA hybridization and 16S rRNA (Zuckerkandl and Pauling, 1965; Woese and Fox, 1977; Woese, 1987). The most recent breakthrough involves rapid and easily available sequencing of entire genomic sequences (Fleischmann et al., 1995; Iguchi et al., 2009; NCBI genomic database, 2012). This has allowed determination of evolutionary relationships among different organisms based upon large numbers of different gene/protein sequences using a variety of approaches (Gupta, 1998; Haggerty et al., 2009; Puigbo et al., 2009; Blair and Murphy, 2011).

The comparative genomic analyses have revealed that phylogenetic relationships deducted based upon different genes and protein sequences are not congruent and lateral gene transfer (LGT) among different taxa is indicated as the main factor responsible for this lack of concordance (Gogarten et al., 2002; Bapteste and Boucher, 2008; Dagan et al., 2008; Puigbo et al., 2009; Swithers et al., 2009; Andam and Gogarten, 2011). This has led to questioning of whether the Darwinian model of evolution involving vertical inheritance of genes from parents to progenies (Darwin, 1859) is applicable to the prokaryotes (Doolittle, 1999; Pennisi, 1999; Gogarten et al., 2002; Dagan and Martin, 2006; Doolittle and Bapteste, 2007; Dagan et al., 2008; Bapteste et al., 2009; Williams et al., 2011). Multiple mechanisms are known to contribute to the evolution of an organism's genomes including genes that are acquired vertically from the parent organism,

evolution of new genes by gene duplication and divergence, gain of new genes by means of LGTs, as well as gene losses in various lineages (Bapteste et al., 2009; Ragan and Beiko, 2009; Treangen and Rocha, 2011; Williams et al., 2011). LGT, in particular, is being increasingly thought to have an overbearing influence on prokaryotic genome composition. Although rRNAs, ribosomal proteins and other genes involved in the information transfer processes are considered less prone to LGTs due to their involvement in complex gene networks (Jain et al., 1999; Sorek et al., 2007), recent studies indicate that no single gene/protein is completely immune to this process (Yap et al., 1999; Doolittle and Bapteste, 2007; Dagan et al., 2008). Some recent studies have estimated that over time most genes (81 ± 15%) have undergone at least one LGT event (Doolittle, 1999; Dagan and Martin, 2007; Doolittle and Bapteste, 2007; Dagan et al., 2008). These studies in large part form the basis of the hypothesis that LGTs have led to abolishment of all signals that can be used for determination of prokaryotic evolutionary relationships and a call for uprooting the tree of life (Martin, 1999; Pennisi, 1999; Doolittle, 2000; Gogarten et al., 2002; Delsuc et al., 2005; Bapteste et al., 2009).

Although the importance of LGTs in genome evolution is widely accepted, there is considerable disagreement concerning the prevalence of LGTs and their impact on prokaryotic evolutionary relationships. While some authors have indicated that LGT is so profuse that its influence disguises the Darwinian mode of evolution involving vertical inheritance of genes (Gogarten et al., 2002; Bapteste et al., 2005b, 2009; Doolittle and Bapteste, 2007; Koonin, 2007), others have inferred that the incidences of LGTs are either very minimal or limited and those genes that are laterally transferred have little impact on prokaryotic phylogeny (Wolf et al., 2002; Kurland et al., 2003; Dutilh et al., 2004; Beiko et al., 2005; Kunin et al., 2005; Kurland, 2005; Galtier, 2007; Puigbo et al., 2009; Gao and Gupta, 2012a). However, there are no standardized methods to assess LGTs and the methods used to infer LGTs are varied and based upon large numbers of often poorly supported assumptions (Koski and Golding, 2001; Koski et al., 2001; Ragan, 2001; Beiko et al., 2005; Boto, 2010). Thus, the prevalence of LGTs differ greatly among different studies and often similar datasets have led to dissimilar conclusions (Koski et al., 2001; Ragan, 2001; Wang, 2001; Lerat et al., 2003; Susko et al., 2006; Zhaxybayeva et al., 2007; Marri and Golding, 2008; Roettger et al., 2009). Therefore, prior to concluding that in view of LGTs the Darwinian mode of evolution is not a suitable model for prokaryotes, reliability of the incidences of LGTs and their overall impact on the evolutionary relationships should be critically examined.

Despite the prevalence of LGTs, phylogenetic trees based upon 16S rRNA as well as numerous single genes as well multi-gene analyses strongly support the existence of large numbers of distinct phyla of bacteria (Ludwig and Klenk, 2005). Additionally, these trees also clearly delineate many discrete taxonomic clades within these phyla (Woese, 1987; Ludwig and Klenk, 2005; Ciccarelli et al., 2006; Wu et al., 2009; Gao and Gupta, 2012a). In a recent detailed study Puigbo et al. (2009) reported construction of phylogenetic trees for 6901 prokaryotic genes. Although there were significant topological differences among these trees, a consistent phylogenetic signal was observed in most of these trees, indicating that the LGT events, which were of random nature, did not obscure the central trend resulting from the vertical transfer of genes. The fact that similar prokaryotic clades at different taxonomic levels (ranging from phyla to genera) are consistently identified in phylogenetic trees based upon different gene/protein sequences strongly indicates that the distinctness of the prokaryotic taxa and their evolutionary relationships are in large part discernible and they have not been obliterated by LGTs (Woese, 1987; Daubin et al., 2002; Kurland et al., 2003; Lerat et al., 2003; Beiko et al., 2005; Kurland, 2005; Ludwig and Klenk, 2005; Ciccarelli et al., 2006; Ragan and Beiko, 2009; Wu et al., 2009; Boto, 2010; Yarza et al., 2010; Gupta, 2010b; Gao and Gupta, 2012a). To account for the above observations and the occurrences of LGTs, it has been suggested that the prokaryotic evolution has both tree-like (at intermediate phylogenetic depths) and non-tree (or net-like) (at the base and tips) characteristics (Dagan et al., 2008; Puigbo et al., 2009, 2010; Swithers et al., 2009; Boto, 2010; Beiko, 2011; Dagan, 2011; Kloesges et al., 2011; Popa et al., 2011).

The availability of genome sequences is also enabling development of novel and independent sequence based approaches for determining the evolutionary relationships among organisms and to assess the impact of LGTs on these relationships. In this review, we provide a summary of our recent work in this area based upon two different types of molecular markers that we have used successfully for understanding the evolutionary relationships among prokaryotes. Based upon these markers it is now possible to identify different prokaryotic taxa ranging from phyla to genera in clear molecular terms and the evolutionary relationships among them can also be reliably deducted (Gupta and Griffiths, 2002; Gupta, 2009, 2010a; Gao and Gupta, 2012b). The relationships revealed by these new approaches strongly support a tree-like branching pattern among prokaryotes and the observed incidences of LGTs, which exhibit no specific pattern or statistical significance, apparently have no major impact on the derived relationships. It is contended that these molecular markers provide valuable means for developing a reliable phylogeny and taxonomy of the prokaryotic organisms.

USEFULNESS OF CONSERVED SIGNATURE INDELS (CSIs) AND CONSERVED SIGNATURE PROTEINS (CSPs) FOR UNDERSTANDING EVOLUTIONARY RELATIONSHIPS AMONG PROKARYOTES

Of the two kinds of molecular markers that we are using for studying prokaryotic evolution, the *c*onserved *s*ignature *i*ndels (inserts or deletions), or CSIs, in protein sequences comprises an important category (Gupta, 1998, 2010a; Griffiths and Gupta, 2001). The CSIs that provide useful molecular markers for evolutionary studies are generally of the same lengths and they are flanked on both sides by conserved regions to ensure that the observed changes are not caused by alignment artifacts (Gupta, 1998; Gupta and Griffiths, 2002; Jordan and Goldman, 2012). When such CSIs are present in the same position in a given protein in a group of related species, their presence is most parsimoniously explained by postulating that the genetic change leading to the CSI occurred in a common ancestor of this group

and then this gene with the indel was vertically transmitted to its progeny (Rivera and Lake, 1992; Baldauf and Palmer, 1993; Gupta, 1998, 2000b; Rokas and Holland, 2000; Cutino-Jimenez et al., 2010). The CSIs that are uniquely shared by organisms of one taxa provide molecular tools for identifying the species from this taxa and consolidating the relationships among bacteria of that taxa by delimiting it in molecular terms (Gupta, 2004). Additionally, depending upon the presence or absence of a given CSI in the outgroup species, it can be determined whether the indel represents an insert or a deletion and based upon this a rooted relationship among the species of interest can be derived. Our earlier work in this regard has led to identification of large numbers of CSIs that are specific for different groups of microbes at various phylogenetic levels (**Table 1**; Gupta and Griffiths, 2006; Gupta, 2009; Gupta and Bhandari, 2011; Gupta and Shami, 2011; Gao and Gupta, 2012b).

The second kind of molecular markers that we have usefully employed in our systematic and evolutionary studies are whole proteins that are uniquely found in particular groups or subgroups of bacteria (Gupta, 2006; Gupta and Griffiths, 2006; Gupta and Mok, 2007; Gao and Gupta, 2012b). Comparative analyses of genomic sequences have indicated that many conserved proteins are uniquely present in all species from particular groups, at different phylogenetic depths (Daubin and Ochman, 2004; Lerat et al., 2005; Gupta, 2006; Gupta and Griffiths, 2006; Gupta and Mok, 2007; Dutilh et al., 2008; Gao and Gupta, 2012b). Because of their unique presence in species from particular phylogenetic clades of species, it is likely that the genes for these CSPs originated once in a common ancestor of these groups and then vertically acquired by all its descendants. Because of their taxa specificity these CSPs again provide valuable molecular markers for identifying different groups of species in molecular terms and for evolutionary studies (Gao and Gupta, 2007; Gupta and Mathews, 2010; Gupta, 2010b). However, when a CSP (or CSI) is confined to certain species/strains, then based upon this information alone, it is often difficult to determine whether these species form a clade in the phylogenetic sense or not. Hence, to understand the evolutionary significance of these signatures, such studies are generally performed in conjunction with phylogenetic analysis, which provides a reference point for evaluating the significance of various CSIs and CSPs (Gao and Gupta, 2007; Gupta and Mathews, 2010; Gupta, 2010b).

Molecular markers in the form of CSIs and CSPs have proven useful for examining or consolidating prokaryotic relationships at domain, phylum as well as intra-phylum levels. **Table 1** provides a summary of some bacterial and archaeal taxa for which CSIs and CSPs have been identified (Gupta, 2010a). Two recent detailed studies based upon CSIs and CSPs have focused upon understanding evolutionary relationships within the phylum Thermotogae and the domain Archaea (Gao and Gupta, 2007; Gupta and Bhandari, 2011; Gupta and Shami, 2011). To illustrate the usefulness of these molecular markers for elucidation of prokaryotic evolutionary relationships, and to assess the influence of LGTs on the derived inferences, results for these two taxonomic groups are reviewed here.

MOLECULAR MARKERS FOR THE THERMOTOGAE

The species of the phylum Thermotogae are a group of hyperthermophilic, anaerobic, gram-negative bacteria recognized by a distinctive toga-like sheath structure and their ability to grow at high temperatures (Huber et al., 1986). The approximately 90 species of this phylum are currently divided into nine Genera within a single family termed the Thermotogaceae (Euzeby, 2011; NCBI Taxonomy, 2012). The Thermotogae species, prospectively, are important tools for industrial and biotechnological applications due to the ecological niche they inhabit and the thermo-stable proteins that they harbor (Conners et al., 2006). With the publication of the genome for *T. maritima*, the first species from this phylum (Nelson et al., 1999), the Thermotogae were brought to the forefront of LGT debate. This was due to the fact that based upon Blast searches it was determined that for about 25% of the genes from *T. maritima* genome, the closest blast hits were from archaeal species rather than any bacteria, leading to the inference that Thermotogae species have incurred high degree of LGTs with the archaeal organisms (Nelson et al., 1999). Upon revisiting this issue, Zhaxybayeva et al. (2009) found that for only about 11% of the Thermotogae proteins Archaea were the closest hits, but that the Thermotogae proteins exhibited maximal similarity (42–48% of genes) to the Firmicutes. Based upon these observations, the Thermotogae species genomes were proposed to be a chimera composed of different bacterial and archaeal sources (Zhaxybayeva et al., 2009). However, these estimates for LGTs have been questioned in other studies which indicate that much less (6–7%) of the Thermotogae genome has been laterally transferred (Garcia-Vallve et al., 2000; Ochman et al., 2000). Further, in view of the fact that Thermotogae species branch in proximity of the Firmicutes phylum (Gupta, 2001; Griffiths and Gupta, 2004b), the observation that a preponderance of the top hits for the Thermotogae species are from Firmicutes is an expected results, and it does not indicate that these genes have been laterally transferred (Zhaxybayeva et al., 2009; Andam and Gogarten, 2011).

Apart from their unique protein toga, the species of the phylum Thermotogae are assigned to this group and divided into its different genera primarily on the basis of their branching in the 16S rRNA trees (Reysenbach, 2001; Huber and Hannig, 2006; Zhaxybayeva et al., 2009; Yarza et al., 2010). Until recently, no unique molecular or biochemical characteristics were known that could distinguish the species of this phylum from other bacteria. For identification of molecular markers that could possibly define this phylum and its sub-taxa, a genome wide analysis was performed on protein sequences from 12 Thermotogae spp. whose genomes were available (Gupta and Bhandari, 2011). The protein sequences from these 12 species as well as species representing other bacteria phyla were aligned and examined for the presence of CSIs that were uniquely present in Thermotogae species or those that were commonly shared with some other bacteria. The analysis identified numerous CSIs specific for all Thermotogae. An example of a CSI consisting of a 3 aa long insert in the ribosomal protein L7 that is exclusively present in all sequenced Thermotogae species, including two recently sequenced species, is shown in **Figure 1A**. The unique presence of this CSI of the same length, at the same position in

Table 1 | Overview of the CSIs and CSPs that have been identified for some major prokaryotic taxa.

Taxonomic group	Number of CSPs/CSIs	References
Archaea	*Archaeal Kingdom specific*: 16 CSPs *Subgroups*: Thaumarchaeota—6 CSIs/201 CSPs, Euryarchaeota—6 CSPs, Thermoacidophiles—77 CSPs, Halophiles—127 CSPs, Methanogens—31 CSPs, Thermococcus-Pyrococcus clade—141 CSPs	Gao and Gupta, 2007; Gupta and Shami, 2011
Crenarchaeota	*Phylum specific*: 6 CSIs, 13 CSPs *Subgroups*: Sulfolobales—3 CSIs/151 CSPs, Thermoproteales—5 CSIs/25 CSPs, Desulfurococcales—4CSPs, Sulfolobales-Desulfurococcales clade—2 CSIs/18 CSPs	Gupta and Shami, 2011
Thaumarchaeota	>200 CSPs	Gupta and Shami, 2011
Thermotogae	*Phylum specific*: 18 CSIs *Subgroups*: Thermotoga genus—13 CSIs, Thermosipho genus—7 CSIs, Thermosipho-Fervidobacterium clade—13 CSIs, Thermotoga-Thermosipho-Fervidobacterium clade—5 CSIs, Petrotoga-Kosmotoga clade—4 CSIs	Gupta and Bhandari, 2011
Cyanobacteria	*Phylum specific*: 39 CSPs/10 CSIs *Subgroups*: Cyanobacterial Clade A—14 CSPs/1 CSI, Other Cyanobacteria (outside clade A)—5 CSPs/4 CSIs, Cyanobacterial Clade C—60 CSPs, Nostocales—65 CSPs, Chroococcales—8 CSPs, *Synechococcus*—14 CSPs, *Prochlorococcus*—19 CSPs, Low B/A type *Prochlorococcus*—67 CSPs	Gupta, 2009; Gupta and Mathews, 2010
Chlamydiae	*Phylum specific*: 59 CSPs/8 CSIs *Subgroups*: Chlamydiaceae—79 CSPs, Chlamydophila—20 CSPs, Chlamydia—20 CSPs	Gupta and Griffiths, 2006
Bacteroidetes, chlorobi and fibrobacteres	*Phylum specific*: 1 CSP/2 CSIs *Subgroup specific*: Bacteroidetes—27 CSPs/2 CSIs, Chlorobi—51 CSPs/2 CSIs, Bacteroidetes and Chlorobi clade—5 CSPs/3CSIs	Gupta, 2004
Actinobacteria	*Phylum specific*: 24 CSPs/4 CSIs *Subgroup specific*: CMN group—13 CSPs, *Mycobacterium* and *Nocardia*—14 CSIs, *Mycobacterium*—24 CSPs, *Micrococcineae*—24 CSPs, Corynebacteriales—4 CSPs/2 CSIs, Bifidobacteriales—14 CSPs/1 CSI	Gao and Gupta, 2005, 2012b; Gao et al., 2006
Deinococcus-thermus	*Phylum specific*: 65 CSPs/8 CSIs *Subgroup specific*: Deinococci—206 SPs	Griffiths and Gupta, 2004a, 2007a
Aquificae	*Phylum specific*: 10 CSPs/5 CSIs	Griffiths and Gupta, 2006b, 2004b
α-proteobacteria	*Class specific*: 6 CSPs/13 CSIs *Subgroups*: Rickettsiales—3 CSPs/2 CSIs, Rickettsiaceae—4 CSPs/5 CSIs, Anaplasmataceae—5 CSPs/2 CSIs, Rhodobacterales-Caulobacter-Rhizobiales clade—2 CSIs, Rhodobacterales-Caulobacter clade—1 CSI, Rhizobiales—6 CSPs/1CSI, Bradyrhizobiaceae—62 CSPs/2CSIs	Gupta and Mok, 2007
γ-proteobacteria	*Class specific*: 4 CSPs/1 CSI *Subgroups*: 20 CSPs, 2 CSIs for various subgroup combinations of subgroups	Gao et al., 2009
ε-proteobacteria	*Class specific*: 49 CSPs/4 CSIs *Subgroups*: *Wolinella-Helicobacter* clade—11 CSPs/2 CSIs, *Campylobacter* genus—18 CSPs/1 CSI	Gupta, 2006
Pasteurellales	*Order specific*: 44 CSIs *Subgroups*: Pasteurellales Clade I—13 CSIs, Pasteurellales Clade II—9 CSIs	Naushad and Gupta, 2012
Clostridia sensu stricto	*Genus specific*: 10 CSPs/3 CSIs	Gupta and Gao, 2009

The table provides general information regarding the number of CSIs and CSPs identified for many taxonomic groups on which genomic studies have been conducted. Further details can be obtained from the corresponding studies.

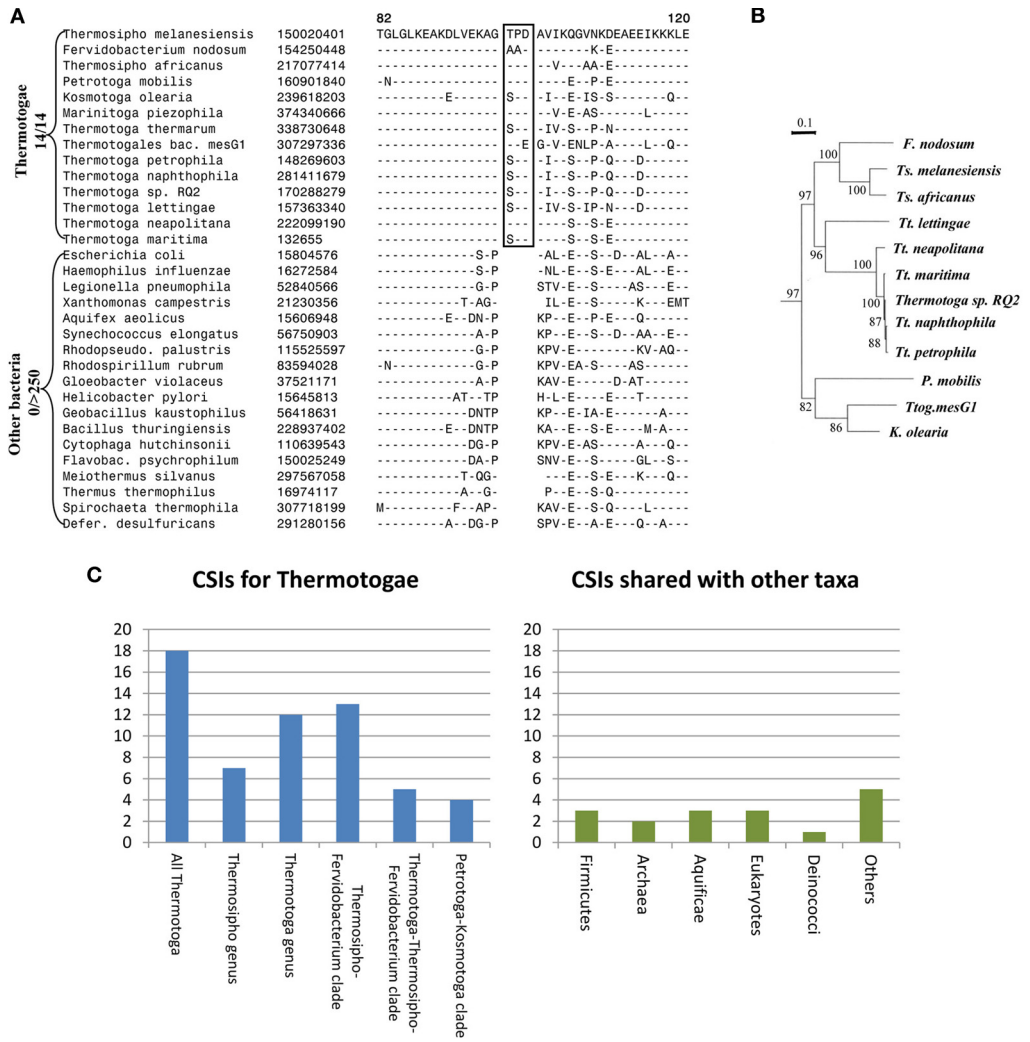


FIGURE 1 | Evolutionary relationships among Thermotogae species based upon CSIs and a Phylogenetic Tree. (A) Partial sequence alignment for the ribosomal protein L7 showing a 3 aa CSI (boxed) that is specific for all detected species of the Thermotogae phylum. The dashes in the alignment (–) indicate amino acid identity with the corresponding residue in the top line; **(B)** A maximum likelihood tree for the 12 sequenced Thermotogae species based upon concatenated sequences for 12 conserved proteins. **(C)** A summary diagram showing the species specificities of different CSIs identified for the Thermotogae group of species. The left panel highlights the CSIs that are specific for the entire Thermotogae phylum or its sub-groups, whereas the right panel indicates the CSIs that were also present in some non-Thermotogae organisms. **Figures 1A,B** modified from Gupta and Bhandari (2011).

this universally distributed protein, in different species from the phylum Thermotogae indicates that the genetic change leading to this CSI occurred once in the common ancestor of the Thermotogae species. In addition to this CSI, this study also identified 17 other CSIs in other important proteins such as DNA recombination protein RecA, DNA polymerase I and tryptophanyl-tRNA synthetase that are also specific for the species from the phylum Thermotogae (Gupta and Bhandari, 2011).

In addition to the large numbers of CSIs that were uniquely present in all Thermotogae species, this study also identified many CSIs that were specific for different sub-groups within the phylum Thermotogae (Gupta and Bhandari, 2011). These included 13 CSIs that were specific for the species of the genus *Thermotoga* and seven others that distinguished species of the genus *Thermosipho* from all others. However, it was observed that the species *Thermotoga lettingae* shared only 1 of 13 CSIs that were otherwise commonly present in other species of this genus. This suggests that *T. lettingae*, which is distantly related to all other *Thermotoga* species, should be assigned to a separate genus. Besides these CSIs that were specific for the species of these two genera, 13 CSIs supported a specific relationships among species of the *Fervidobacterium* and *Thermosipho* genera; 5 CSIs were shared by species from the genus *Thermotoga* and those from the *Fervidobacterium-Thermosipho* clade; and 4 CSIs supported a grouping of the *Petrotoga* and *Kosmotoga* genera along with the species *Thermotogales bacterium MesG1.Ag.4.2* (**Figure 1C**, left panel; Gupta and Bhandari, 2011). Importantly, all of the

relationships indicated by various CSIs were also independently observed in a phylogenetic tree for the Thermotogae species based upon concatenated sequences for 12 conserved proteins (**Figure 1B**).

The CSIs identified in the above study independently and strongly supported different nodes observed in the phylogenetic tree for Thermotogae species all the way from phylum to genus level. If the hypothesis that LGT events have abolished the ability to discern prokaryotic relationships was correct, then it should have been difficult to identify discrete molecular markers supporting distant relationships among these species. At the very least, the Thermotogae species would have shown relationships with species of other prokaryotic groups such as Firmicutes or Archaea as frequently as they did with one another. In this study, in addition to the CSIs that were specific for the Thermotogae species (**Figure 1C**, left panel), several CSIs were also identified that the Thermotogae shared with species from other prokaryotic or eukaryotic organisms (**Figure 1C**, right panel). However, such CSIs, suggesting possible LGT between Thermotogae and other taxa, were far outweighed by CSIs supporting the monophyletic, tree-like relationships among the species of the phylum (left panel) (Gupta and Bhandari, 2011). Assuming that all the CSIs that the Thermotogae shared with other groups are due to LGT, less than 20% (16 of 85) of all Thermotogae genes containing these CSIs have incurred LGTs (Gupta and Bhandari, 2011). Moreover, these presumed LGT events are of random nature and in no case do the Thermotogae species share more than a total of 3 CSIs with any particular phyla of species. Additionally, in most of these cases only a few species from these other taxa contained the indels that were present in most or all Thermotogae species (Gupta and Bhandari, 2011). Thus, these other CSIs, although they are present in a few isolated species from other taxa, are also largely specific for the Thermotogae species and they do not affect the ability of other CSIs to clearly discriminate Thermotogae species from all other bacteria or to deduce the evolutionary relationships amongst species from this phylum.

The shared presence of similar CSI in unrelated taxa can result from two different possibilities, either the gene with the CSI was laterally transferred among the two groups or that independent CSIs owing to two separate genetic events are responsible for these CSIs. After identification of such CSIs, tree-making approaches can be used to test if the presence of the indel in the two groups is due to LGT. Previously, in our work, a number of CSIs in the GlyA and MurA proteins that were commonly shared by the Chlamydiae and a subgroup of Actinobacteria were shown to be due to lateral transfer of genes from Actinobacteria to a common ancestor of the Chlamydiae (Griffiths and Gupta, 2006a). Recently, the shared presence of several CSIs in the bacteriochlorophyll biosynthesis proteins by unrelated phyla of photosynthetic prokaryotes has also been shown to be due to LGTs (Raymond et al., 2002; Gupta, 2012). However, in many other instances phylogenetic analyses have not supported LGT as the possible reason for the presence of a related CSI in unrelated taxa. In these cases, similar CSIs have originated independently in these lineages due to their presumed similar functions in these particular taxa.

MOLECULAR MARKERS FOR THE ARCHAEA AND ITS SUB-GROUPS

Archaea are widely recognized as the third domain of life. They generally inhabit extreme environments such as those of extreme temperature, pH or salinity, where little to no other life exists (Woese et al., 1990). However, recent studies indicate that archaeal species are widespread in the environment and they play a major role in the carbon and nitrogen cycles (Pace, 1997; Herndl et al., 2005; Leininger et al., 2006). Some archaeal species have been found to be commensal organisms residing in human colons (Oxley et al., 2010). The Archaea are generally divided into two main phyla, the Crenarchaeota and Euryarchaeota, based on 16S rRNA data and other phylogenetic data (Woese et al., 1990; Gribaldo and Brochier-Armanet, 2006). The Crenarchaeotes are described as thermophiles with sulfur-reducing capabilities while the Euryarchaeotes are metabolically and morphologically quite diverse (Gribaldo and Brochier-Armanet, 2006; Gupta and Shami, 2011). The mesophilic Crenarchaeota have been recently placed into a separate phylum called the Thaumarchaeota (Brochier-Armanet et al., 2008; Gupta and Shami, 2011).

Despite the importance of Archaea in different environments and in understanding of the evolutionary history of life on earth (Woese et al., 1990; Gupta, 2000a), until recently, very few molecular characteristics were known that are uniquely shared by all Archaea. Additionally, as the higher taxonomic groups within Archaea are described primarily based upon 16S rRNA trees, the characteristics that are unique to different phyla, classes, orders and families of the Archaea have scarcely been elucidated (Boone et al., 2001). The utilization of archaeal genomes for discovery of CSPs as well as CSIs has provided significant information in the form of molecular markers that are distinctive characteristics of Archaea and its taxonomic sub-groups. In 2007, a comprehensive analysis was performed on available archaeal genomes to search for CSPs that were unique to either all Archaea or many of its sub-groups (Gao and Gupta, 2007). Over 1400 such proteins distinctive of Archaea or its main taxa were discovered (**Figure 2**). In the analysis, sixteen proteins specific to all or most Archaea were identified that were not present in any bacterial or eukaryotic organism. Numerous proteins whose homologs were limited to the Crenarchaeota, Euryarchaeota and other sub-groups such as the Thermococci, Thermoplasmata, and Halobacteriales were also detected (**Figure 2**). Significantly, this study also identified 31 proteins that were commonly shared by all methanogenic bacteria (Gao and Gupta, 2007). In the 16S rRNA and other phylogenetic trees, the methanogenic Archaea do not form a monophyletic lineage, but instead are split into a number of distinct clusters separated by non-methanogenic Archaea (Burggraf et al., 1991; Brochier et al., 2004; Bapteste et al., 2005a; Gao and Gupta, 2007). Because most of the proteins that are commonly shared by various methanogens are generally involved in functions related to methanogenesis and their genes are clustered into a few large operons in genomes (Harms et al., 1995; Tersteegen and Hedderich, 1999; Grabarse et al., 2001; Gao and Gupta, 2007), it is likely that the genes for these proteins have been laterally acquired by different Archaea. This could provide a plausible explanation for the observed discrepancy in the branching of methanogenic Archaea in phylogenetic trees and

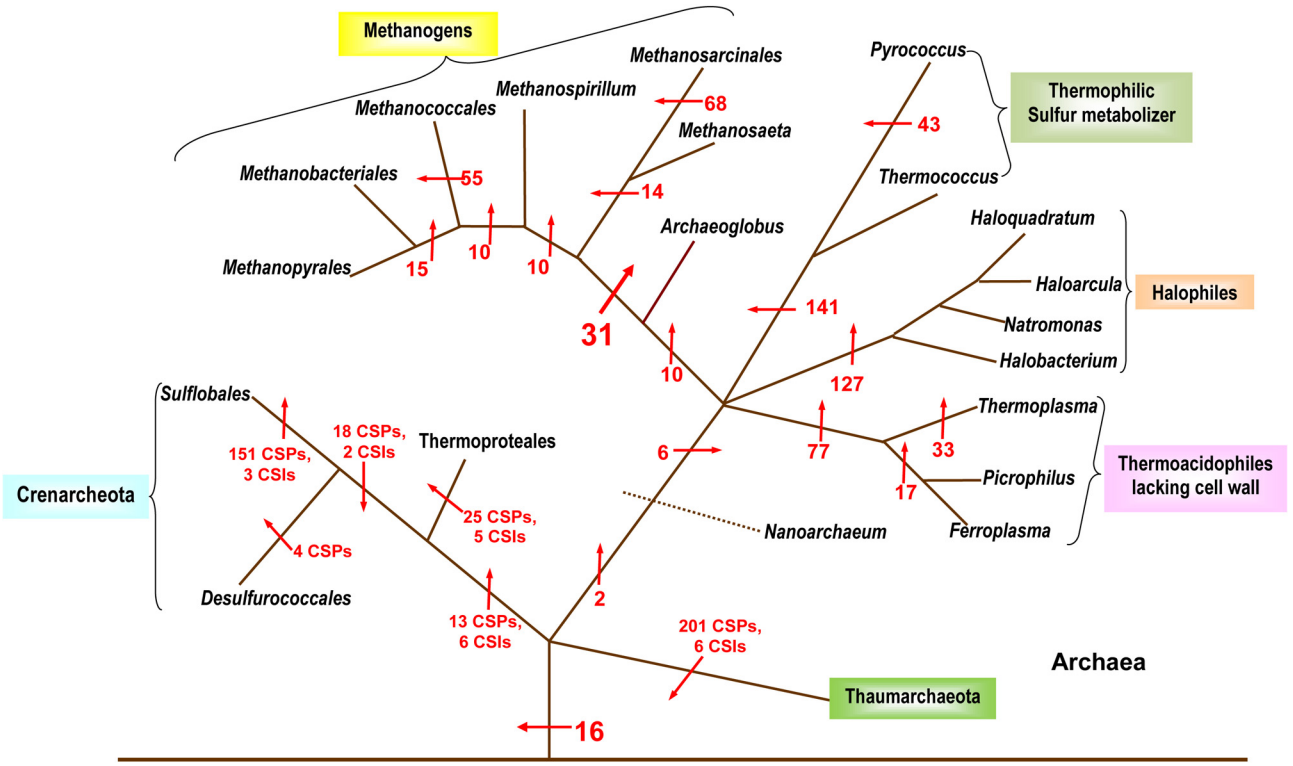


FIGURE 2 | A summary diagram showing the various molecular markers that have been identified for the Archaeal kingdom and its subgroups. The arrows indicate the suggested evolutionary stages where the proteins unique for a particular taxa are proposed to have been introduced. The numbers beside the arrows indicate the number of CSIs and CSPs specific for the various taxa (these numbers indicate CSPs unless otherwise noted). The branching pattern shown is based solely upon the distribution patterns of CSPs and CSIs. Modified from Gao and Gupta (2007) and Gupta and Shami (2011).

their unique sharing of genes for these proteins (Gao and Gupta, 2007).

A recent analysis has further added to the catalogue of molecular signatures for the archaeal organisms (Gupta and Shami, 2011). The focus of this study was on identifying CSIs and CSPs that were specific for the Crenarchaeota and Thaumarchaeota phyla (Gupta and Shami, 2011). Six CSIs and 13 CSPs specific for all species of the phylum Crenarchaeota were identified along with numerous markers for its different orders: the Sulfolobales (151 CSPs, 3 CSIs), Thermoproteales (25 CSPs, 5 CSIs) and the Desulfurococcales (4 CSPs). The study also described the markers (18 CSPs and 2 CSIs) indicative of a close relationship among the Sulfolobales and the Desulfurococcales. The discriminative ability of CSPs is highlighted by the results of blast searches on some CSPs that are specific for the Crenarchaeota or its main groups (Sulfolobales, Thermoproteales, Desulfurococcales and Acidilobales) that are shown in **Table 2**. In these cases, BLASTP searches were carried out on these proteins and the results for all species for whom the observed *E*-values were significant are shown. From the results presented in **Table 2**, it is evident that the first 2 CSPs are specific for the Crenarchaeota phylum, the next two are uniquely found in various species belonging to the orders Desulfurococcales, Acidilobales and Sulfolobales, whereas the last 5 CSPs are distinctive characteristics of species belonging to either the Desulfurococcales (and Acidilobales), the Sulfolobales, or the Thermoproteales orders.

In this study, more than 200 CSPs for various members of the newly defined Thaumarchaeota phylum were also identified (Gupta and Shami, 2011). The Thaumarchaeota are composed of several organisms previously included in the Crenarchaeota (Brochier-Armanet et al., 2008). The two phyla appear as sister groups in phylogenetic analysis and they also share 3 CSIs and 10 CSPs with each other (Gupta and Shami, 2011). Nevertheless, the two groups can be phylogenetically differentiated and numerous markers have been identified for each group that helps to define them molecularly as individual taxa (Gupta and Shami, 2011). A summary diagram depicting the various molecular markers specific for the archaeal species is shown in **Figure 2**. It should be noted that CSIs were only identified for the Thaumarchaeota and the Crenarchaeota and no detailed analysis to identify CSIs has thus far been carried out on the Euryarchaeota.

The two studies noted above have identified numerous CSIs and CSPs for the Archaea, its main phyla (Euryarchaeota, Crenarchaeota, Thaumarchaeota) and a number of its subphylum level taxa (Sulfolobales, Thermococcales, Halobacteriales, etc.; Gao and Gupta, 2007; Gupta and Shami, 2011). Except for the methanogens, the distribution patterns of the identified CSIs and CSPs are also strongly supported by the phylogenetic

Table 2 | A series of proteins specific for the Crenarchaeota and its sub-groups.

	Protein accession # **Protein length**	**NP_147640** **262 aa**	**NP_147284** **143 aa**	**BAA81469** **98 aa**	**NP_147588** **228 aa**	**YP_001041009** **127 aa**	**YP_254810** **228 aa**	**YP_254922** **270 aa**	**NP_559041** **626 aa**	**NP_559897** **113 aa**
Desulfurococcales	*Aeropyrum pernix*	0.0	9e–98	5e–64	7e–161	7e–22	–	–	–	–
	Hyperthermus butylicus	3e–46	9e–43	1e–20	1e–23	3e–25	–	–	–	–
	Ignicoccus hospitalis	3e–41	–	5e–27	4e–19	3e–25	–	–	–	–
	Desulfurococcus kamchatkensis	7e–46	1e–21	2e–20	5e–17	7e–32	–	–	–	–
	Staphylothermus marinus	4e–56	1e–25	3e–21	3e–21	2e–85	–	–	–	–
Acidilobales	*Acidilobus saccharovorans*	9e–56	4e–36	4e–21	1e–46	1e–19	–	–	–	–
Sulfolobales	*Sulfolobus tokodaii*	4e–40	2e–29	3e–20	7e–26	–	1e–77	1e–80	–	–
	Sulfolobus islandicus	4e–42	6e–30	1e–25	1e–15	–	7e–50	8e–65	–	–
	Sulfolobus acidocaldarius	7e–34	3e–23	4e–22	4e–24	–	2e–162	0.0	–	–
	Sulfolobus solfataricus	1e–41	7e–30	5e–26	8e–15	–	5e–50	8e–64	–	–
	Metallosphaera sedula	3e–31	3e–33	3e–20	1e–22	–	4e–39	8e–60	–	–
Thermoproteales	*Pyrobaculum aerophilum*	9e–18	3e–11	–	–	–	–	–	0.0	2e–73
	Pyrobaculum islandicum	3e–18	3e–11	–	–	–	–	–	0.0	6e–54
	Pyrobaculum arsenaticum	1e–18	1e–10	–	–	–	–	–	0.0	2e–63
	Pyrobaculum caldifontis	6e–22	7e–11	–	–	–	–	–	0.0	1e–60
	Thermofilum pendens	1e–35	5e–30	–	–	–	–	–	1e–42	3e–10
	Caldivirga maquilingensis	1e–17	4e–8	–	–	–	–	–	1e–87	2e–22
	Thermoproteus neutrophilus	2e–19	7e–11	–	–	–	–	–	0.0	5e–61
	Thermoproteus tenax	3e–15	6e–10	–	–	–	–	–	0.0	4e–46
Top non-Crenarchaeota hit		*Brucella melitensis* (2e–1)	*Desulfobacterium autotrophicum* (8e–1)	*Aromatoleum aromaticum* (4e–1)	*Serpula lacrymans* (7e–1)	*Clonorchis sinensis* (3e–1)	*Granulicatella elegans* (6e–1)	*Encephalitozoon cuniculi* (7e–1)	*Burkholderia cenocepacia* (9e–1)	*Sordaria macrospora* (1e–1)

Blastp searches were carried out on proteins specific for the Crenarchaeota or its sub-groups and the results for representative species from different sub-groups of the Crenarchaeota are shown with the observed E-values. E-values greater than 1e–3 are considered insignificant hits with lack of homology to the query protein sequence.

The dashes (–) indicate that the homolog for the protein query was not detected in the BlastP searches.

Top non-Crenarchaeota hits indicate detection of species outside the Crenarchaeota that were observed to have the lowest E-value scores.

branching pattern of the archaeal organisms (Gribaldo and Brochier-Armanet, 2006; Gao and Gupta, 2007; Brochier-Armanet et al., 2008; Gupta and Shami, 2011). Considering the specificities of these molecular markers for either all Archaea or different clades of Archaea, these results strongly indicate that LGTs have not obliterated the phylogenetic signal necessary to delineate the evolutionary relationships among this domain of prokaryotes. The discovered CSIs and CSPs also provide novel tools for the identification of different groups of Archaea in various environments.

THE USEFULNESS OF THE CSIs FOR UNDERSTANDING BACTERIAL PHYLOGENY AND TAXONOMY

In addition to the CSIs that are specific for particular prokaryotic taxa, several of the identified CSIs have also proven useful in clarifying the branching order and interrelationships amongst different bacterial phyla (Gupta, 2001, 2011; Gupta and Griffiths, 2002). One example of these kinds of CSIs, which are referred to as the main-line signatures in our work, is shown in **Figure 3A**. In this case, a large ~100 aa insert in the β subunit of RNA polymerase protein (RpoB) is commonly

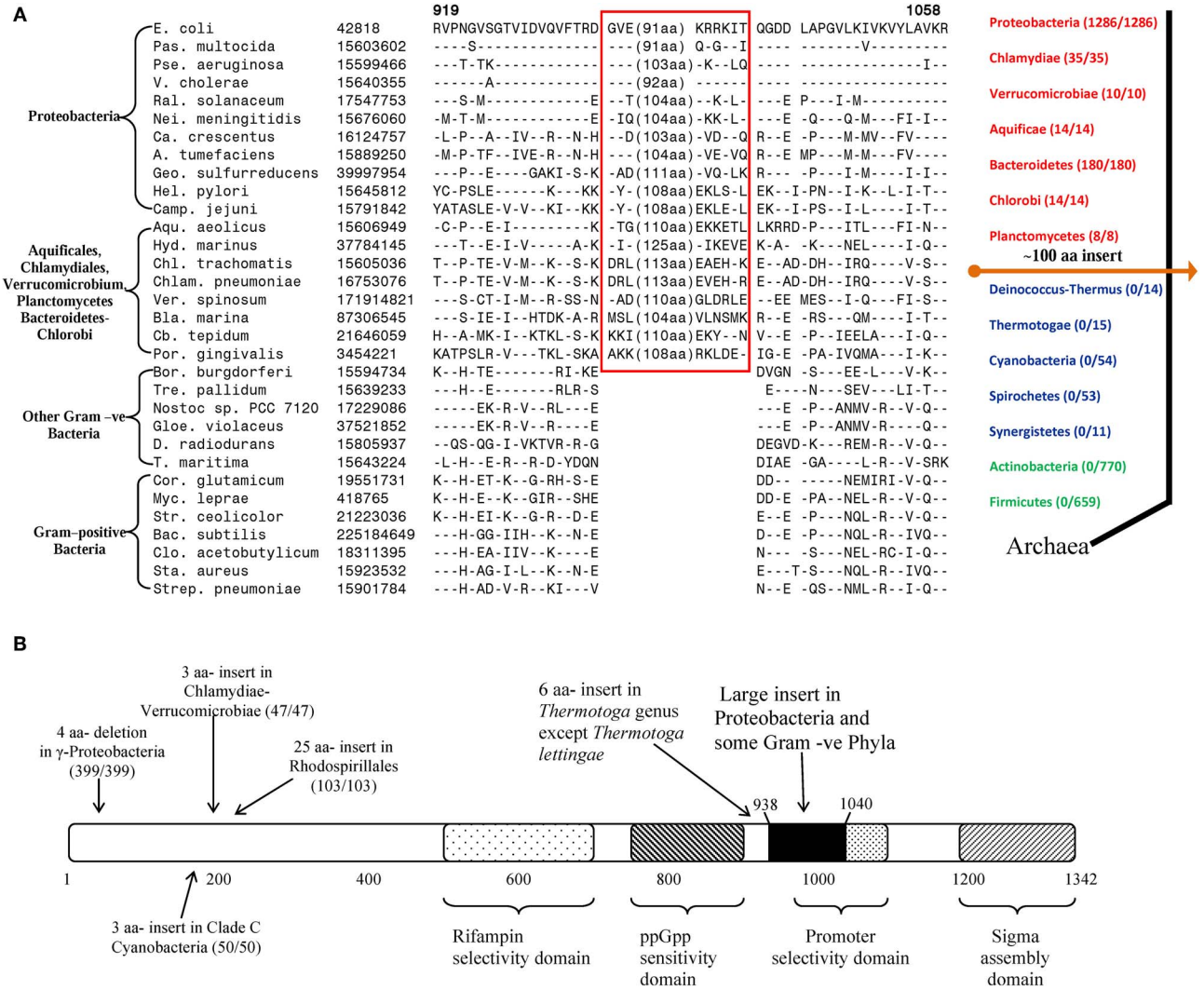


FIGURE 3 | Evolutionary significance of various identified CSIs in the RNA polymerase β subunit. (A) A portion of the RpoB sequence alignment showing a large insert (boxed) that is distinctive characteristic of all Proteobacteria and some Gram-negative phyla (Chlamydiae-Verrucomicrobiae, Aquificales, Planctomycetes, and Bacteroidetes-Chlorobi), but not found in other phyla of bacteria. Due to the large size of the insert, its entire sequence is not shown. Dashes (–) indicate identity with the amino acid on the top line. On the right is a linear representation of prokaryotic relationships based on the presence and absence of this CSI. The numbers in the brackets indicate the species of each phylum, which have been identified to contain the CSI. **(B)** A schematic representation of the sequence for *E. coli* RNA polymerase β subunit (RpoB) showing some functionally important regions and the positions of different lineage-specific inserts that have been identified within this protein. The large insert depicted in **(A)** (≈ 100 aa in *E. coli)* is shown in solid black. The positions of CSIs for different groups are roughly indicated using arrows. The values in the brackets identify the number of organisms in each respective group and the number of these species to harbour the indicated CSI. In all cases no organism outside of the indicated group was identified to contain the indel. The indicated CSIs have been described in earlier work (Griffiths and Gupta, 2004b, 2007b; Gupta and Mok, 2007; Gao et al., 2009; Gupta and Bhandari, 2011; Naushad and Gupta, 2012).

shared by all of the sequenced species belonging to the phyla Proteobacteria (different subclasses), Aquificae, Chlamydiae, Verrucomicrobiae, Bacteroidetes-Chlorobi, and Planctomycetes (Griffiths and Gupta, 2007b). This insert is present in all of the >1500 sequences that are available from species from these phyla. On the other hand, this CSI is not found in any of the >1500 sequences available from various species belonging to the phyla Firmicutes, Actinobacteria, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Synergistetes, Spirochaetes, etc. This insert is also not found in the archaeal RpoB homologs, thus providing evidence that this indel is an insert in the groups of species where it is found (Griffiths and Gupta, 2004b). Based upon its highly specific species distribution pattern, which argues strongly against the lateral transfer of this gene amongst various phyla, the genetic change responsible for this CSI most likely occurred in a common ancestor of the group of species that contain this CSI, after the divergence of other bacterial phyla that lack this indel as indicated in **Figure 3A** (right panel). A number of other mainline CSIs, which based upon their species distribution patterns have occurred at other important branch points in prokaryotic evolution, have been described in our earlier works (Griffiths and Gupta, 2001, 2004b; Gupta and Griffiths, 2002). Based upon these CSIs, it is possible to determine the branching order of most of the bacterial phyla (Gupta, 1998, 2001, 2003; Griffiths and Gupta, 2004b; see also www.bacterialphylogeny.info).

Within the highly conserved RpoB protein, in addition to the large CSI that is commonly shared by a number of bacterial phyla, several other CSIs have been identified that are specific for different groups/phyla of bacteria. The taxon specificities of these CSIs and their positions within in the RpoB polypeptide are shown in **Figure 3B**. These CSIs include a 4 aa deletion that is commonly and uniquely shared by a number of different orders of the γ-proteobacteria (399/399 species), a 3 aa insert that is specifically present in all of the Chlamydiae-Verrucomicrobiae species (47/47), another 3 aa insert that is a distinctive property of the Clade C cyanobacteria (50/50; Gupta, 2009), a 25 aa insert in various species from the order Rhodospirillales (103/103) and a 6 aa insert in all species from the genus *Thermotoga* except *T. lettingae* (Gupta and Griffiths, 2006; Gupta and Mok, 2007; Griffiths and Gupta, 2007b; Gao et al., 2009; Gupta and Bhandari, 2011). It is highly significant that within a single gene/protein multiple highly specific CSIs are present, each of which is specific for a different group of bacteria and help distinguish these groups from all other bacteria. These CSIs are not present in any species outside of the indicated taxa. The presence of these different taxa-specific characteristics in a single gene/protein strongly indicates that the genetic changes responsible for these CSIs occurred in the gene for this key protein at different stages in the evolution of bacterial domain and that no LGT of the gene for the RpoB protein has occurred among these taxa. Similar to the RpoB protein, multiple CSIs that are specific for different groups of prokaryotes have also been identified in many other important genes/proteins. These observations indicate that strong and consistent phylogenetic signals that are very likely not affected to any significant extent by the LGTs are still present in many conserved and universally distributed genes/proteins and these can be used to trace the evolutionary relationships among prokaryotes.

It is important to point out that virtually all of the higher taxonomic clades (above the Genus rank) within prokaryotes are currently identified solely on the basis of their branching in the 16S rRNA trees. Because the phylogenetic trees are a continuum, based upon them it has proven difficult to clearly define or delimit the boundaries of different taxonomic groups. Additionally, for virtually all of the higher prokaryotic taxa, no molecular, biochemical or physiological characteristics are known that are unique to them. Hence, a very important aspect of microbiology that needs to be understood is that in what respects do species from different main groups of bacteria differ from each other and what, if any, unique molecular, biochemical, structural or physiological characteristics are commonly shared by species from different groups? In this context, the large numbers of CSIs and CSPs for different taxonomic clades of bacteria that are being discovered by comparative genomic analyses provide novel and valuable tools for taxonomic, diagnostic, and biochemical studies (Gupta and Bhandari, 2011; Gao and Gupta, 2012b). In view of the specificities of the discovered CSIs and CSPs for different groups of prokaryotes and their retention by all species from these groups of prokaryotes, it is highly likely that these CSIs and CSPs are involved in functions that are essential for prokaryotes (Galperin and Koonin, 2004; Fang et al., 2005; Singh and Gupta, 2009; Schoeffler et al., 2010). Indeed, recent work on several CSIs have shown that they are essential for the group of organisms where they are found and the deletion or substantial changes in them led to failure of cell growth (Singh and Gupta, 2009; Schoeffler et al., 2010). Hence, further studies on understanding the cellular functions of the different taxa-specific CSIs and CSPs could lead to identification of novel biochemical and other functional characteristics that are specific for these groups of organisms.

It should also be noted that the identified CSIs and CSPs generally constitute robust molecular characteristics that exhibit high degree of predictive ability. Many of these CSIs and CSPs were discovered when the sequence information was available for very few prokaryotic species. However, despite the large increase in the number of sequenced genomes, most of these CSIs and CSPs are still specific for the originally indicated groups of prokaryotes (Gupta, 2009, 2011; Gao and Gupta, 2012b). Additionally, for several Chlamydiae-, Aquificae-, Deinococcus-Thermus- and Actinobacteria- specific degenerate primers based on conserved flanking sequences have been designed and they have been used to amplify the sequence regions predicted to contain the CSIs from large numbers of organisms for whom no sequences were available (Griffiths and Gupta, 2004a,b; Gao and Gupta, 2005; Griffiths et al., 2005). In these studies, in almost all cases the expected inserts or deletions were found to be present in previously un-sequenced organisms from the indicated groups, thus providing evidence that these CSIs and CSPs provide powerful new tools for identification of both known as well as novel species from different groups of prokaryotes.

CONCLUSIONS

There is considerable debate at present concerning the impact of LGTs on understanding prokaryotic phylogeny. While there

is little dispute that LGT plays an important role in microbial evolution, the extreme view taken by some that LGTs are so rampant within the prokaryotes that it totally masks the evolutionary signal from vertical transfer of genes (Doolittle, 2000; Gogarten et al., 2002; Doolittle and Bapteste, 2007; Dagan et al., 2008; Bapteste et al., 2009) is not supported by available evidence. As reviewed here, in phylogenetic trees based upon most gene/protein sequences all of the major groups within prokaryotes (from phylum down to genus level) are generally clearly identified, thus indicating that a strong phylogenetic signal emanating from vertical transfer of genes is maintained throughout prokaryotic evolution (Gupta, 1998, 2000b; Dutilh et al., 2004; Ludwig and Klenk, 2005; Ciccarelli et al., 2006; Puigbo et al., 2009). Most of the differences seen amongst these trees are either at the tips (i.e., species/strains levels) or at the base, i.e., relationships among the higher taxonomic clades such as phyla, class, etc. A recent study indicates that the incidence of LGTs shows linear correlation with the genome sequence and the GC content similarities of the donor and recipient organisms (Kloesges et al., 2011). Hence, while many of the observed inconsistencies between different gene trees at the species/strain levels could be due to LGTs (Puigbo et al., 2009; Kloesges et al., 2011), the differences in branching pattern at the higher taxonomic levels are perhaps in large parts due to loss of the phylogenetic signal and the lack of resolving power of the tree-based phylogenetic approaches (Gupta, 1998; Ludwig and Klenk, 2005; Puigbo et al., 2009).

In this review we have discussed the usefulness of CSIs and CSPs, as novel and important class of molecular markers for understanding the evolutionary relationships among prokaryotes. We have presented compelling evidence that based upon the species distribution patterns of these molecular signatures different prokaryotic taxa from phylum down to the genus levels can be clearly identified. Additionally, based upon these markers it is also possible to reliably deduct the evolutionary relationships amongst different prokaryotic taxa, both within a phylum and among different phyla. The evolutionary relationships deduced based upon these molecular markers generally exhibit high degree of congruency with those indicated by 16S rRNA trees or other gene/protein sequences. The analyses based upon these markers have also been able to clarify some relationships that are not resolved in phylogenetic trees. The species distribution patterns of these markers thus provide strong evidence that different clades of bacteria have evolved in a tree-like manner and that the prokaryotic organisms are not an exception to the Darwinian model of evolution. The relatively small numbers of these CSIs where the indel is also present in some unrelated species, which could be due to LGTs, show no specific pattern or relationship, thus they have minimal or no impact on the strong and consistent tree-like branching pattern that is evident from all other identified CSIs. However, it should be acknowledged that all of the work using CSIs and CSPs on understanding the evolutionary relationships among prokaryotes has thus far been carried out at genus level or higher taxa. Hence, it remains to be seen whether this approach will prove equally useful in clarifying the evolutionary relationships at the species or strain levels or not, where the evolutionary flux and the incidences of LGTs are deemed to be the highest (Daubin et al., 2003; Lerat et al., 2003; Dagan et al., 2008; Puigbo et al., 2009; Kloesges et al., 2011).

The molecular markers such as those described here in addition to their usefulness for understanding prokaryotic phylogeny also provide valuable means to address/clarify a number of important aspects of microbiology. (1) Based upon these markers different prokaryotic taxa can now be identified in clear molecular terms rather than only as phylogenetic entities. (2) Based upon them the boundaries of different taxonomic clades can also be more clearly defined. (3) Due to their high degree of specificity and predictive ability, they provide important diagnostic tools for identifying both known and unknown species belonging to these groups of bacteria. (4) The shared presence of these CSIs by unrelated groups of bacteria provides potential means for identifying novel cases of LGTs. (5) Functional studies on these molecular markers should help in the discovery of novel biochemical or physiological properties that are distinctive characteristics of different groups of prokaryotes.

Lastly, it should be acknowledged that the number of genes which harbor rare genetic changes such as these CSIs is generally small in comparison to the total number of genes that are present in any genome. However, the genes containing these CSIs are involved in different essential functions and they are often are amongst the most conserved proteins found in various organisms. Although, the criticism could be levied that the inferences based upon small numbers of genes/proteins containing these CSIs are not representative of the entire genomes (Dagan and Martin, 2006; Bapteste and Boucher, 2008), it should be emphasized that in a number of studies such as those discussed here, the reported CSIs or CSPs represent analyses of the entire genomes. Based upon these CSIs and/or CSPs, no other significant or consistent relationships or patterns among these organisms, other than those indicated here, can be derived from consideration of all of the gene/protein sequences in these genomes using these approaches. In this context it is also helpful to remember that molecular sequences like all other fossils change and disintegrate over long evolutionary periods of time and they lose their information content at different rates. Hence, a well-preserved fossil is generally considered to be far more informative than hundreds or even thousands of disintegrated fossils. Following this analogy, it is expected that not all genes/proteins will prove equally useful for understanding the evolutionary history of prokaryotes, which spans > 3.5 billion years. Thus, the best we can hope for is to find significant numbers of conserved genes/proteins, which contain consistent and reliable signals such as those described in the present work, whose inferences are generally consistent with all/most other available information.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Natural Science and Engineering Research Council of Canada. Hafiz S. Naushad was partly supported by a scholarship from the Islamia University of Bahawalpur.

REFERENCES

Andam, C. P., and Gogarten, J. P. (2011). Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* 9, 543–555.

Baldauf, S. L., and Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11558–11562.

Bapteste, E., and Boucher, Y. (2008). Lateral gene transfer challenges principles of microbial systematics. *Trends Microbiol.* 16, 200–207.

Bapteste, E., Brochier, C., and Boucher, Y. (2005a). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* 1, 353–363.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R. L., and Doolittle, W. F. (2005b). Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* 5, 33.

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupre, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Dir.* 4, 34.

Beiko, R. G. (2011). Telling the whole story in a 10,000-genome world. *Biol. Dir.* 6, 34.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14332–14337.

Blair, C., and Murphy, R. W. (2011). Recent trends in molecular phylogenetic analysis: where to next? *J. Hered.* 102, 130–138.

Boone, D. R., Castenholz, R. W., and Garrity, G. M. (2001). *Bergey's Manual of Systematic Bacteriology*. New York, NY: Springer, 1–721.

Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* 277, 819–827.

Brochier, C., Forterre, P., and Gribaldo, S. (2004). Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox. *Genome Biol.* 5, R17.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third Archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252.

Buchanan, R. E., and Gibbons, N. E. (1974). *Bergey's Manual of Deteminative Bacteriology*. Baltimore, MD: Williams and Wilkins.

Burggraf, S., Stetter, K. O., Rouviere, P., and Woese, C. R. (1991). Methanopyrus kandleri: an Archaeal methanogen unrelated to all other known methanogens. *Syst. Appl. Microbiol.* 14, 346–351.

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.

Conners, S. B., Mongodin, E. F., Johnson, M. R., Montero, C. I., Nelson, K. E., and Kelly, R. M. (2006). Microbial biochemistry, physiology, and biotechnology of hyperthermophilic Thermotoga species. *FEMS Microbiol. Rev.* 30, 872–905.

Cowan, S. T. (1965). Principles and practice of bacterial taxonomy—a forward look. *J. Gen. Microbiol.* 39, 143–153.

Cutino-Jimenez, A. M., Martins-Pinheiro, M., Lima, W. C., Martin-Tornet, A., Morales, O. G., and Menck, C. F. (2010). Evolutionary placement of Xanthomonadales based on conserved protein signature sequences. *Mol. Phylogenet. Evol.* 54, 524–534.

Dagan, T. (2011). Phylogenomic networks. *Trends Microbiol.* 19, 483–491.

Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10039–10044.

Dagan, T., and Martin, W. (2006). The tree of one percent. *Genome Biol.* 7, 118.

Dagan, T., and Martin, W. (2007). Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 870–875.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Daubin, V., Gouy, M., and Perriere, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12, 1080–1090.

Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 14, 1036–1042.

Daubin, V., Moran, N. A., and Ochman, H. (2003). Phylogenetics and the cohesion of bacterial genomes. *Science* 301, 829–832.

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* 284, 2124–2129.

Doolittle, W. F. (2000). Uprooting the tree of life. *Sci. Am.* 282, 90–95.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Dutilh, B. E., Huynen, M. A., Bruno, W. J., and Snel, B. (2004). The consistent phylogenetic signal in genome trees revealed by reducing the impact of noise. *J. Mol. Evol.* 58, 527–539.

Dutilh, B. E., Snel, B., Ettema, T. J., and Huynen, M. A. (2008). Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* 25, 1659–1667.

Euzeby, J. P. (2011). List of prokaryotic names with standing in nomenclature. http://www.bacterio.cict.fr/classifphyla.html. (Ref Type: Generic).

Fang, G., Rocha, E., and Danchin, A. (2005). How essential are nonessential genes? *Mol. Biol. Evol.* 22, 2147–2156.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.

Galperin, M. Y., and Koonin, E. V. (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–5463.

Galtier, N. (2007). A model of horizontal gene transfer and the bacterial phylogeny problem. *Syst. Biol.* 56, 633–642.

Gao, B., and Gupta, R. S. (2005). Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int. J. Syst. Evol. Microbiol.* 55, 2401–2412.

Gao, B., and Gupta, R. S. (2007). Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* 8, 86.

Gao, B., and Gupta, R. S. (2012a). Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek* 101, 45–54.

Gao, B., and Gupta, R. S. (2012b). Phylogenetic framework and molecular signatures for the main clades of the phylum actinobacteria. *Microbiol. Mol. Biol. Rev.* 76, 66–112.

Gao, B., Mohan, R., and Gupta, R. S. (2009). Phylogenomics and protein signatures elucidating the evolutionary relationships among the Gammaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 59, 234–247.

Gao, B., Paramanathan, R., and Gupta, R. S. (2006). Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie Van Leeuwenhoek* 90, 69–91.

Garcia-Vallve, S., Romeu, A., and Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10, 1719–1725.

Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.

Grabarse, W., Mahlert, F., Duin, E. C., Goubeaud, M., Shima, S., Thauer, R. K., Lamzin, V., and Ermler, U. (2001). On the mechanism of biological methane formation: structural evidence for conformational changes in methyl-coenzyme M reductase upon substrate binding. *J. Mol. Biol.* 309, 315–330.

Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1007–1022.

Griffiths, E., and Gupta, R. S. (2001). The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that Fibrobacter diverged at a similar time to Chlamydia and the Cytophaga-Flavobacterium-Bacteroides division. *Microbiology* 147, 2611–2622.

Griffiths, E., and Gupta, R. S. (2004a). Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the deinococcus-thermus phylum. *J. Bacteriol.* 186, 3097–3107.

Griffiths, E., and Gupta, R. S. (2004b). Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales. *Int. Microbiol.* 7, 41–52.

Griffiths, E., and Gupta, R. S. (2006a). Lateral transfers of serine hydroxymethyltransferase (glyA) and UDP-N-acetylglucosamine enolpyruvyl transferase (murA) genes from free-living Actinobacteria to the parasitic chlamydiae. *J. Mol. Evol.* 63, 283–296.

Griffiths, E., and Gupta, R. S. (2006b). Molecular signatures in protein sequences that are characteristics of the phylum Aquificae. *Int. J. Syst. Evol. Microbiol.* 56, 99–107.

Griffiths, E., and Gupta, R. S. (2007a). Identification of signature proteins that are distinctive of the

Deinococcus-Thermus phylum. *Int. Microbiol.* 10, 201–208.

Griffiths, E., and Gupta, R. S. (2007b). Phylogeny and shared conserved inserts in proteins provide evidence that Verrucomicrobia are the closest known free-living relatives of chlamydiae. *Microbiology* 153, 2648–2654.

Griffiths, E., Petrich, A. K., and Gupta, R. S. (2005). Conserved indels in essential proteins that are distinctive characteristics of Chlamydiales and provide novel means for their identification. *Microbiology* 151, 2647–2657.

Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491.

Gupta, R. S. (2000a). The natural evolutionary relationships among prokaryotes. *Crit. Rev. Microbiol.* 26, 111–131.

Gupta, R. S. (2000b). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Rev.* 24, 367–402.

Gupta, R. S. (2001). The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *Int. Microbiol.* 4, 187–202.

Gupta, R. S. (2003). Evolutionary relationships among photosynthetic bacteria. *Photosynth. Res.* 76, 173–183.

Gupta, R. S. (2004). The phylogeny and signature sequences characteristics of Fibrobacteres, Chlorobi, and Bacteroidetes. *Crit. Rev. Microbiol.* 30, 123–143.

Gupta, R. S. (2006). Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (Campylobacterales). *BMC Genomics* 7, 167.

Gupta, R. S. (2009). Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *Int. J. Syst. Evol. Microbiol.* 59, 2510–2526.

Gupta, R. S. (2010a). "Applications of conserved indels for understanding microbial phylogeny," in *Molecular Phylogeny of Microorganisms*, eds A. Oren and R. T. Papke (Norfolk, UK: Caister Academic Press), 135–150.

Gupta, R. S. (2010b). Molecular signatures for the main phyla of photosynthetic bacteria and their subgroups. *Photosynth. Res.* 104, 357–372.

Gupta, R. S. (2011). Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek* 100, 171–182.

Gupta, R. S. (2012). Origin and spread of photosynthsis based upon conserved sequence Features in key bacteriochlorophyll biosynthesis proteins. *Mol. Biol. Evol.* PMID: 22628531. [Epub ahead of print].

Gupta, R. S., and Bhandari, V. (2011). Phylogeny and molecular signatures for the phylum Thermotogae and its subgroups. *Antonie Van Leeuwenhoek* 100, 1–34.

Gupta, R. S., and Gao, B. (2009). Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridium sensu stricto (cluster I). *Int. J. Syst. Evol. Microbiol.* 59, 285–294.

Gupta, R. S., and Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor. Popul. Biol.* 61, 423–434.

Gupta, R. S., and Griffiths, E. (2006). Chlamydiae-specific proteins and indels: novel tools for studies. *Trends Microbiol.* 14, 527–535.

Gupta, R. S., and Mathews, D. W. (2010). Signature proteins for the major clades of Cyanobacteria. *BMC Evol. Biol.* 10, 24.

Gupta, R. S., and Mok, A. (2007). Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. *BMC Microbiol.* 7, 106.

Gupta, R. S., and Shami, A. (2011). Molecular signatures for the Crenarchaeota and the Thaumarchaeota. *Antonie Van Leeuwenhoek* 99, 133–157.

Haggerty, L. S., Martin, F. J., Fitzpatrick, D. A., and McInerney, J. O. (2009). Gene and genome trees conflict at many levels. *Philos. Trans. R. Soc. Lond. B Biol Sci.* 364, 2209–2219.

Harms, U., Weiss, D. S., Gartner, P., Linder, D., and Thauer, R. K. (1995). The energy conserving N5-methyltetrahydromethanopterin: coenzyme M methyltransferase complex from Methanobacterium thermoautotrophicum is composed of eight different subunits. *Eur. J. Biochem.* 228, 640–648.

Herndl, G. J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., and Pernthaler, J. (2005). Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl. Environ. Microbiol.* 71, 2303–2309.

Huber, R., and Hannig, M. (2006). "Thermotogales," in *The Prokaryotes*, eds M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer), 899–922.

Huber, R., Langworthy, T. A., Konig, H., Thomm, M., Woese, C. R., Sleytr, U. B., and Stetter, K. O. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 C*. *Arch. Microbiol.* 144, 324–333.

Iguchi, A., Thomson, N. R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I. R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., Kenny, B., Quail, M. A., Thurston, S., Dougan, G., Hayashi, T., Parkhill, J., and Frankel, G. (2009). Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127, H6 strain E2348/69. *J. Bacteriol.* 191, 347–354.

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806.

Jordan, G., and Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol.* 29, 1125–1139.

Kloesges, T., Popa, O., Martin, W., and Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.* 28, 1057–1074.

Koonin, E. V. (2007). The Biological Big Bang model for the major transitions in evolution. *Biol. Dir.* 2, 21.

Koski, L. B., and Golding, G. B. (2001). The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* 52, 540–542.

Koski, L. B., Morton, R. A., and Golding, G. B. (2001). Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* 18, 404–412.

Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C. A. (2005). The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* 15, 954–959.

Kurland, C. G. (2005). What tangled web: barriers to rampant horizontal gene transfer. *Bioessays* 27, 741–747.

Kurland, C. G., Canback, B., and Berg, O. G. (2003). Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9658–9662.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C., and Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809.

Lerat, E., Daubin, V., and Moran, N. A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol.* 1:E19. doi: 10.1371/journal.pbio.0000019

Lerat, E., Daubin, V., Ochman, H., and Moran, N. A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* 3:e130. doi: 10.1371/journal.pbio.0030130

Ludwig, W., and Klenk, H.-P. (2005). "Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics," in *Bergey's Manual of Systematic Bacteriology*, eds D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (Berlin: Springer-Verlag), 49–65.

Marri, P. R., and Golding, G. B. (2008). Gene amelioration demonstrated: the journey of nascent genes in bacteria. *Genome* 51, 164–168.

Martin, W. (1999). Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* 21, 99–104.

Naushad, H. S., and Gupta, R. S. (2012). Molecular signatures (conserved indels) in protein sequences that are specific for the order Pasteurellales and distinguish two of its main clades. *Antonie Van Leeuwenhoek* 101, 105–124.

NCBI genomic database. (2012). http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi. (Ref Type: Electronic Citation).

NCBI Taxonomy. (2012). http://www.ncbi.nlm.nih.gov/taxonomy. (Ref Type: Electronic Citation).

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C., and Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Oxley, A. P., Lanfranconi, M. P., Wurdemann, D., Ott, S., Schreiber, S., McGenity, T. J., Timmis, K. N., and Nogales, B. (2010). Halophilic Archaea in the human intestinal mucosa. *Environ. Microbiol.* 12, 2398–2410.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740.

Pennisi, E. (1999). Is it time to uproot the tree of life? *Science* 284, 1305–1307.

Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J. Biol.* 8, 59.

Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* 2, 745–756.

Ragan, M. A. (2001). On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol. Lett.* 201, 187–191.

Ragan, M. A., and Beiko, R. G. (2009). Lateral genetic transfer: open issues. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2241–2251.

Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y., and Blankenship, R. E. (2002). Whole-genome analysis of photosynthetic prokaryotes. *Science* 298, 1616–1620.

Reysenbach, A.-L. (2001). "Phylum BII. Thermotogae phy. nov," in *Bergey's Manual of Systematic Bacteriology* eds G. M. Garrity, D. R. Boone, and R. W. Castenholz (Berlin: Springer), 369–387.

Rivera, M. C., and Lake, J. A. (1992). Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257, 74–76.

Roettger, M., Martin, W., and Dagan, T. (2009). A machine-learning approach reveals that alignment properties alone can accurately predict inference of lateral gene transfer from discordant phylogenies. *Mol. Biol. Evol.* 26, 1931–1939.

Rokas, A., and Holland, P. W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459.

Schoeffler, A. J., May, A. P., and Berger, J. M. (2010). A domain insertion in *Escherichia coli* GyrB adopts a novel fold that plays a critical role in gyrase function. *Nucleic Acids Res.* 38, 7830–7844.

Singh, B., and Gupta, R. S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genomics* 281, 361–373.

Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., and Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452.

Stanier, R. Y., Adelberg, E. A., and Ingraham, J. L. (1976). *The Microbial World*. Englewood Cliffs, NJ: Prentice-Hall Inc., 1–871.

Susko, E., Leigh, J., Doolittle, W. F., and Bapteste, E. (2006). Visualizing and assessing phylogenetic congruence of core gene sets: a case study of the gamma-proteobacteria. *Mol. Biol. Evol.* 23, 1019–1030.

Swithers, K. S., Gogarten, J. P., and Fournier, G. P. (2009). Trees in the web of life. *J. Biol.* 8, 54.

Tersteegen, A., and Hedderich, R. (1999). Methanobacterium thermoautotrophicum encodes two multisubunit membrane-bound [NiFe] hydrogenases. Transcription of the operons and sequence analysis of the deduced proteins. *Eur. J. Biochem.* 264, 930–943.

Treangen, T. J., and Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284

Wang, B. (2001). Limitations of compositional approach to identifying horizontally transferred genes. *J. Mol. Evol.* 53, 244–250.

Williams, D., Fournier, G. P., Lapierre, P., Swithers, K. S., Green, A. G., Andam, C. P., and Gogarten, J. P. (2011). A rooted net of life. *Biol. Dir.* 6, 45.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. (2002). Genome trees and the Tree of Life. *Trends Genet.* 18, 472–479.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., Goodwin, L., Wu, M., Tindall, B. J., Hooper, S. D., Pati, A., Lykidis, A., Spring, S., Anderson, I. J., D'Haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J. F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E. M., Kyrpides, N. C., Klenk, H. P., and Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056–1060.

Yap, W. H., Zhang, Z., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete Thermomonospora chromogena and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* 181, 5201–5209.

Yarza, P., Ludwig, W., Euzeby, J., Amann, R., Schleifer, K. H., Glockner, F. O., and Rossello-Mora, R. (2010). Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst. Appl. Microbiol.* 33, 291–299.

Zhaxybayeva, O., Nesbo, C. L., and Doolittle, W. F. (2007). Systematic overestimation of gene gain through false diagnosis of gene absence. *Genome Biol.* 8, 402.

Zhaxybayeva, O., Swithers, K. S., Lapierre, P., Fournier, G. P., Bickhart, D. M., DeBoy, R. T., Nelson, K. E., Nesbo, C. L., Doolittle, W. F., Gogarten, J. P., and Noll, K. M. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5865–5870.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theor. Biol.* 8, 357–366.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 April 2012; accepted: 27 June 2012; published online: 26 July 2012.

Citation: Bhandari V, Naushad HS and Gupta RS (2012) Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution. Front. Cell. Inf. Microbio. 2:98. doi: 10.3389/fcimb.2012.00098

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 01 May 2012
doi: 10.3389/fcimb.2012.00054

Selfishness, warfare, and economics; or integration, cooperation, and biology

Emiliano Salvucci*

Consejo Nacional de Investigaciones Cientificas y Técnicas, Argentina

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Kalliopi Georgiades, Unité de Recherche en maladies infectieuses tropicales emergentes, France
Jean-Marc Rolain, Centre National de la Recherche Scientifique, France
Fernando Baquero, Ramón y Cajal Institute for Health Research, Spain

***Correspondence:**
Emiliano Salvucci, Consejo Nacional de Investigaciones Cientificas y Técnicas, Argentina.
e-mail: salvucci.emiliano@gmail.com

The acceptance of Darwin's theory of evolution by natural selection is not complete and it has been pointed out its limitation to explain the complex processes that constitute the transformation of species. It is necessary to discuss the explaining power of the dominant paradigm. It is common that new discoveries bring about contradictions that are intended to be overcome by adjusting results to the dominant reductionist paradigm using all sorts of gradations and combinations that are admitted for each case. In addition to the discussion on the validity of natural selection, modern findings represent a challenge to the interpretation of the observations with the Darwinian view of competition and struggle for life as theoretical basis. New holistic interpretations are emerging related to the Net of Life, in which the interconnection of ecosystems constitutes a dynamic and self-regulating biosphere: viruses are recognized as a macroorganism with a huge collection of genes, most unknown that constitute the major planet's gene pool. They play a fundamental role in evolution since their sequences are capable of integrating into the genomes in an "infective" way and become an essential part of multicellular organisms. They have content with "biological sense" i.e., they appear as part of normal life processes and have a serious role as carrier elements of complex genetic information. Antibiotics are cell signals with main effects on general metabolism and transcription on bacterial cells and communities. The hologenome theory considers an organism and all of its associated symbiotic microbes (parasites, mutualists, synergists, amensalists) as a result of symbiopoiesis. Microbes, helmints, that are normally understood as parasites are cohabitants and they have cohabited with their host and drive the evolution and existence of the partners. Each organism is the result of integration of complex systems. The eukaryotic organism is the result of combination of bacterial, virus, and eukaryotic DNA and it is the result of the interaction of its own genome with the genome of its microbiota, and their metabolism are intertwined (as a "superorganism") along evolution. The darwinian paradigm had its origin in the free market theories and concepts of Malthus and Spencer. Then, nature was explained on the basis of market theories moving away from an accurate explanation of natural phenomena. It is necessary to acknowledge the limitations of the dominant dogma. These new interpretations about biological processes, molecules, roles of viruses in nature, and microbial interactions are remarkable points to be considered in order to construct a solid theory adjusted to the facts and with less speculations and tortuous semantic traps.

Keywords: Darwinism, natural selection, evolution, paradigm, virus, hologenome, autopoiesis

"I do not write for those who examine
new books quickly, often
with the intention of finding in them their ideas
preconceived, but for the few who read,
who meditate deeply, who love
study of nature and are capable of
even sacrificing their own interests,
for the knowledge of a new truth."
J. B. Lamarck (1744–1829)

INTRODUCTION

While the Modern Synthetic Theory is the most widely accepted evolutionary theory, many authors consider that it is necessary to evaluate its explanatory power and a self-criticism of orthodoxy from the inability to explain the phenomena and discoveries daily observed (Ehrlich and Birch, 1967; Goldsmith, 1989; Margulis and Sagan, 1995; Kampis, 1997; Sandín, 1997; Abdalla, 2006).

This dominant paradigm based on a conception of the transmission of strictly Mendelian characters has as basic tenets: (1) Evolution is a gradual process of substitution of alleles within a population. The source of variability in these alleles would be the point mutations or micromutations. (2) The genetic material is only the raw material. What drives the evolutionary process is natural selection (Mayr, 1966; Dobzhansky et al., 1977; Sandín, 1997).

However, with the data provided by different areas of biology, this theoretical framework based on natural selection appears weak to explain the complex evolutionary processes. At least, it is necessary to discuss the explaining power of the dominant paradigm. It is common that new discoveries bring about contradictions that are intended to be overcome by adjusting results to the dominant reductionist paradigm using all sorts of gradations and combinations that are admitted for each case (Sandín, 1997; Forterre, 2010). Nowadays there are new interpretations about biological processes, new approaches and perspectives that are remarkable points to be considered in order to construct a solid theory more adjusted to the facts, and with less speculations and tortuous semantic traps.

The present work is a humble contribution to that discussion with the intention of enriching it by providing new perspectives in evolution related to complex systems.

THE KIDNAPPING OF BIOLOGY

Darwinism grew out of the Malthusian concepts and vision that disease and food shortages act as regulators of the population favoring the fittest in a continuous struggle for life. Darwin wrote his book "On the Origin of Species by Means of Natural Selection, or the maintenance of favored races in the struggle for existence" (1859) based on Malthus theory and then in the expressions of Herbert Spencer: "As more individuals are produced which may survive, there must be necessarily a struggle for existence (. . .) is the doctrine of Malthus applied with multiplied force to the nature" (Darwin, 1869). And elsewhere he writes: "I call this principle which preserves all small variation, it is useful, natural selection mark your faculty with man's selection. But the expression used by Herbert Spencer that the fittest survive is more accurate." In words of Sandín: "The idea expressed more forcefully in the work of Darwin is the extrapolation of the activities of ranchers and farmers to the phenomena of nature" (Sandín, 1997).

It is not the purpose of this paper to review the historical injustice done to those scientists who built the evolutionary scientific basis and began the studies of evolutionary mechanisms. But, it is important to briefly remember some facts. Darwin was not the "inventor" of evolution, neither the idea of evolutionary process was in the "air" before him. On the contrary, it was in a much more solid basis.

Jean Baptiste Pierre-Antoine de Monet, Chevallier de Lamarck, published the more structured Theory of Evolution in 1809 in his book Philosophie Zoologique. He was a disciple of Georges-Louis Leclerc, comte de Buffon and Professor of the Natural History Museum. In 1800, he gave a lecture exposing a coherent theory on the transformation and laid the foundations of epigenesis and organism-environment interaction derived from the mechanism of adaptation. Buffon was the author of an encyclopedia on nature, in 44 volumes (only 36 of them were published in life), the "Histoire naturelle, générale et particulière," where he mentioned that the species observed were transformed and links between organisms. Frederic Gerard in "Theorie de l'evolution des formes organiques," (1841–1849) exhibited a clear distinction between micro and macroevolution based on thorough paleontological studies. To these works are added those from Agassiz, Geoffroy Saint Hilaire (stating "teratologies" abrupt morphological changes that occur during the development), von Zittel, von Baer, Tremaux, who developed the "allopatric speciation" and "punctuated equilibrium long before Darwin and Gould" (Wilkins and Nelson, 2008), and others. The idea of evolution was known and studied among naturalists. Later, after receiving a letter from Wallace set forth the concept of natural selection independently, Darwin published his famous book "From Origin of Species by Means of Natural Selection, or the conservation of Favoured Races in the Struggle for Life."

The idea of natural selection was noted by many philosophers and scientists before Darwin, from the ancient Greek philosophers Empedocles and Aristóteles (third and fourth centuries BC) to Edward Blyth (1810–1873) and Wallace. From 1835 to 1837, Blyth published some articles in The British Magazine of Natural History (Vols. 8, 9, and 10) dealing with natural selection, adaptive radiation, and the struggle for life. It is known that Darwin recieved copies of this magazine while in Peru in 1835 during his voyage on the Beagle. In 1750, the concept of natural selection was noted by Pierre-Louis Moreau de Maupertuis in his "Essay on Cosmology." Also, it was defined by Denis Diderot (1713–1784), William Charles Wells (in an essay from 1813, "Two Essays . . . with Some Observations on the Causes of the Differences of color and form Between the white and black races of men. By the Late WC Wells . . . with a Memoir of His Life, written by himself"), Patrick Matthew (1790–1874) as well as by James Cowles Prichard and William Lawrence.

In the midst of industrial revolution, Darwin observed the growth of misery and poverty. He was influenced and linked to *laissez faire* policies, propelled by Adam Smith, who proposed the lowest state intervention (it was postulated, among others things, to stop creating schools) so as to "naturally" remove the homeless through a free competition. Gertrude Himmelfarb noted that Darwinism was a biological justification of the status of the victorian society as the "fittest": "The theory of natural selection, it is said, could only have originated in England, because only *laissez faire* England provided the atomistic, egotistic mentality necessary to its conception. Only there could Darwin have blandly assumed that the basic unit was the individual, the basic instinct selfinterest, and the basic activity struggle. Spengler, describing the *Origin* as: "the application of economics to biology," said that it reeked of the atmosphere of the English factory . . . natural selection arose . . . in England because it was a perfect expression of Victorian "greed-philosophy" of the capitalist ethic and Manchester economics" (1962, p. 418). In that place and time, there was a social predisposition for that kind of evolution theory.

History often tells us that Darwin found rejection in society of the time and among the church hierarchy. However, Darwin

found great support among the most influential scientists and their ideas were welcomed by the X-club. This elite society of the time consisted of a group made among others by Joseph Dalton Hooker, Thomas Henry Huxley, John Lubbock, Herbert Spencer, who propelled Darwinian ideas and had remarkable power to control the Royal Society (Barton, 1998).

It is well known the discussion between Huxley, defender of Darwin, and the Bishop of Oxford, Wilberforce. While the church defended the fixity, that species did not change, they also questioned the weaknesses of Darwinian proposal, which assumed the transformation of species as a fact but without proof that it occurs by the proposed mechanisms. Wilberforce was right on some points: the first question, in the course of human history there was no evidence of any new species development. Secondly, the selective pressures, although it is true that they have an effect, they do not cause a change of species. Finally, the phenomenon of hybrid sterility was a strong evidence in favor of the fixity of species. Thus, this well known dispute raises a dichotomy that is useful to both dogmas nowadays. It is stating that any challenge to the Darwinian "science" is a "creationist" attack and avoids a scientific discussion about the weaknesses of Darwinism and a recognition that this is also a dogma. The really important issue is not the creationist critics, because this is faith, but the scientific criticisms and what we can do to build a more scientific evolutionary theory.

Since the inception of darwinist natural selection (the cornerstone of the dominant theory), the acclamation was not a unanimous reaction. Among the *scientific* criticisms received, we can mention those of Charles Darwin[1], Adam Sedgwick, Aldous Huxley, Karl von Baer, Louis Agassiz, Richard Owen, Cherles Lyell, Richard Lewontin, St. George Mivart, Albert von Kölliker, Clémence Royer, Robert Peters, etc. This was, not because there was a naive resistant to the science from creationists, but Darwin's theory had huge gaps in its "pure state" and did not explain the complexities observed in organisms and it did not fit the fossil record available at the time (and less to the present). Therefore, from the beginning, this was a theory *scientifically* problematic (Abdalla, 2006).

THE DISCUSSION ABOUT NATURAL SELECTION

In Darwin's work, natural selection takes many forms and nuances. Darwin postulates this "mechanism" generator of new species in a scenario of continued competition. It also takes other definitions as a determinant of character preservation, general process, survival of the fittest, agent, power, cause of extinction, strength. The definitions given for this invisible arm is also varied and their use to explain it all leads to acquire a stunning conceptual flexibility (Cervantes, 2011a).

In words of Futuyma: "*Natural selection is the only* ***mechanism*** *known to cause the evolution of adaptations, so many biologists would simply define an adaptation as a characteristic that has evolved by natural selection*" *and* "*any consistent* ***difference in fitness*** *among phenotypically different classes of biological entities*" (Futuyma, 2009). We cannot define it as a mechanism, given that in a mechanism there are elements known and arranged to ensure a predictable performance. Furthermore, as natural selection would be the generator of species and the insurer of the survival of the fittest, it must also generate morphological novelties (Cervantes, 2011a).

For Dawkins "*there is of course no 'architect.' The DNA instructions have been assembled by natural selection.*" However, "*Natural selection* ***is not an external force or agent****, and certainly not a purposeful one. It is a name for* ***statistical differences*** *in reproductive success among genes, organisms, or populations, and nothing more.*" (Dawkins, 1976). But, "*natural selection, i.e., survival and differential reproduction of organisms, is the main* ***controlling agent*** *of evolutionary change* (Dobzhansky et al., 1977).

Also, *Natural selection is at one and the same time a blind and creative* ***process*** (Dobzhansky, 1973). The idea of selection implies a teleological residue. Selection implies intention since this term refers to a deliberate action of men. If we consider that natural selection is a process, we are allowed to associate it with any natural phenomenon, and we would be allocated for this purpose or intentional phenomenon. The phenomenon (evolution) is confused with the concept that seeks to explain (the selection) (Cervantes, 2011a,b).

Schluter (2009), who did not define natural selection, writes: "The main question today is how selection leads to speciation (. . .) what are the mechanisms of natural selection (. . .)." It is assumed that mechanisms of natural selection (that is not a mechanism) are not known. Even though it is generally accepted that natural selection could not only generate all species but also "drive" the evolution, i.e., the generation of new structures, the cause of appearance of the existing body systems.

At the same time, natural selection is a statistical difference, cause of adaptations, process, mechanism, the assembler of DNA, the agent that acts over DNA, the result of the adaptations (reproductive success once adapted to the environment), the difference in fitness, the result of that difference, the differential survival of entities. For Cervantes, it is a semantic ghost. A concept that is many things at the same time is probably nothing (Cervantes, 2011a). We can agree that everybody understands natural selection as survival and differential reproduction of organisms. However, the term refers both the causes as the effects and takes lot of nuances along literature. This indefiniteness made everything seem to be explained but nothing is explained actually. Everything leads us to confirm the existence of pliable natural selection with the existence of living organisms (survivors) and that they are adapted to their environment. That leaves us still at the starting point of evolutionary research.

Linguistic traps of Darwinism began in Darwin's work but continued through time and spread more confusion. During 70 s there was a discussion about the tautological nature of natural selection. Initially, natural selection claimed that in nature not only a few survive, but also that the fittest survive. That is, those that survive are the fittest to survive, because survival means that not all of them do it, surviving means ability to survive and they survive precisely because they are the fittest. It's a circular reasoning that does not represent any advance in knowledge. What any evolutionary theory should prove is what the laws of evolution are

[1] "I admit . . . that in the earlier editions of my *Origin of Species* I probably attributed too much to the action of natural descent of the survival of the fittest." —*Charles Darwin, The Descent of Man*, Vol. 1 (1871 1st ed.), p. 152.

and do not say that the fittest survive. Peters argues that given its inability to make predictions it cannot be called a scientific theory (Peters, 1976). Natural selection is currently used to explain relationships among organisms, without being used in the context of the evolutionary process, i.e., major organizational, morphological, physiological changes and the origin of species. The core of the problem is that, despite the defenses that can be done in favor of natural selection, it does not add any knowledge or information to contribute to the explanation of the process.

From the point of view of the renowned philosopher and epistemologist Karl Popper the criterion of demarcation, i.e., a rule that defines when a theory is scientific or not, is its falsifiability. If a proposition is not falsifiable it is not scientific, and his rebuttal is determined by experimentation, the scientific method. As a tautology, natural selection is not falsifiable, and then, with this criterion, it is not a scientific theory (Popper, 1963).

For Ehrlich and Birch, in agreement with Popper, Darwinism *"cannot be refuted by any possible observations and it is thus outside empirical science"* (. . .). It is *"an evolutionary dogma accepted by most of us as part of our training"* (Ehrlich and Birch, 1967). A concept that was very vague from the beginning, in a text with little scientific rigor and a lot of ambiguity, was sustained over time and forced to fit the new discoveries.

Taking natural selection as correct, it can also lead to inconsistencies in the theory (Bouchard and Rosenberg, 2004). With knowledge of the complexity of the microbial world (natural selection arises from the observation of domestic animals) and the complexity revealed by genetics until today, the excessive eagerness to believe in natural selection is striking.

NEW FINDINGS, OLD PARADIGM

Besides the semantic problems, another questionable aspect of the dominant theory is the important place occupied by random mutation. Mutation is not a solid explanation neither at levels of generation of new structures that constitutes the evolution nor in the generation of new species (Bernhard, 1967; Schützemberger, 1967). In Bacteria, mutation rates are subject to complex regulation that we are now just beginning to understand (Wright, 2000). Furthermore, bacterial populations tend to have low mutation rates which give stability to their genomes and avoid lethal mutations (Martinez et al., 2009a).

Darwinian reductionism in which everything is reduced to the sum of the parts leads to determinism according to which if we know the parts we can understand the whole. In this regard, it is believed that the complexity of life can be explained by the mechanical interaction of the fundamental molecules, mainly nucleic acids (DNA and RNA). For Abdalla, one of the facets of the potential crisis of paradigm in biology is related to this reflection on the complexity. The neo-Darwinian paradigm eventually leads to a reductionist approach that believes life is a result of localized phenomena in the DNA molecule, subjected to random changes and natural selection (Abdalla, 2006).

Throughout these decades several mechanisms and biological processes have been described that are difficult to frame within the Synthetic Theory: the mobile elements, repeat DNA sequences, the homeotic genes, regulatory sequences, the implication of endogens virus in the regulation and control of embryonic development, morphogenetic fields with incredible precision in the spatial and temporal process of the formation (Harrison, 1937; Weiss, 1939; Child, 1941). A lot of processes control cell functioning and self-regulate each other conforming complex networks, molecular memory, gene-gene communication, and multitasking of eukaryotic genomes (Ball, 2001; Mattick and Gagen, 2001). The evidence provided by evolutionary ontogeny, those provided by the fossil record, the "Evo-Devo," the morphological novelties, horizontal transfer, the integration of genomes, the presence of a high percentage of bacterial and viral genes in eukaryotic genomes, the response to the environment and epigenetic phenomena, self-organizing systems are some of the aspects that constitute a body of knowledge that points out the limitations of the theory of competition, natural selection, and random mutations.

The evidence shows that genetic moving elements through changes in location and duplication, chromosomal rearrangements, cause changes in gene expression and regulation. These sequences also are a constituent part of the structures. For example, more than one gene sequence expressed in 37 human tissues have been identified as belonging to endogenous retroviruses (Johnson and Coffin, 1999; Mattick and Gagen, 2001; Vitali et al., 2003; Mallet et al., 2004; Hamilton, 2006). Furthermore, with the new discoveries, it is necessary to redefine gene, that is far away to be the gene in which it is sustained the Darwinist theory (Gerstein, 2007; Ledford, 2008; Buchanan et al., 2009). It is doubt the existance of the common ancestor (and the known domains Bacteria, Eukaria, and Archaea needs to be redefined Boyer et al., 2010).

The idea of natural selection is powerful because of being so simple. The embryological and genomic remodeling observed in evolution (Gilbert et al., 1996) seem not at all explained by the survival of the fittest (the less fit can also survive) and with that warlike scenario in which even the genes competes and where living beings are used by their own genes (Sandín, 1997).

Maybe, it is time "to resynthesize biology, put organism back into its environment; connect it again to its evolutionary past" (Woese, 2004).

OTHER PERSPECTIVES

Since there are basic facts of evolution that are the most difficult to "fit" in the framework of conventional theory, it is necesary to evaluate the explanatory power of the central dogma. The study of the dominant paradigm shortcomings in the light of the continuous discoveries involve sociological, biological, and epistemological aspects leading to a kind of Kuhnian revolution which other sciences such as physics have already experienced. On the contrary, in many reports the continuous discoveries are adjusted to the paradigm that the results contradict.

The Darwinian perspective does not take into account that reductionism leads to study living things, or partial aspects of them as if they were independent entities. Also it is common to refer that natural selection acts at "different levels," and each character, molecule or process is explained (or assumed to be explained) by action of this strength/mechanism/differential reproduction/etc. Organisms clearly do not exist as isolated organisms but in terms of its environment consisting of living

and non-living forms at different levels between which there are interconnections and interdependencies. Living organisms are in intense exchanges with their environment and are capable of self-organization forming a dynamic ecosystem. The interconnection of ecosystems constitutes a dynamic and self-regulating biosphere: the Net of Life (Sandín, 1997; Maturana and Varela, 1999). Even when these concepts appear to be well studied, it persists the intention to explain everything by a selfishness and warfare view (Nedelcu et al., 2011; Vannier-Santos and Lenzi, 2011) that does not take into account that organisms evolved intertwined and the coexistence is the result of what we are studying. As one example, Nedelcu et al. (2011) deal with *the problem* of altruism. The authors conclude that "active death in single-celled organisms is a maladaptive trait maintained as a byproduct of selection on pro-survival functions, but that could—under conditions in which kin/group selection can act—be co-opted into an altruistic trait" (Nedelcu et al., 2011). In this case, even when the authors are assuming that it is necessary a new paradigm, they just create new tortuous semantic traps and metaphors to explain by economical terms the phenomenon studied. A new theoretical basis is necessary a that takes into account the integrative and associative process that are observed in nature and the evolution of organism in association with their partners and the environment instead of maintaining economic prejudices and speculations that organisms live and exist thanks to cost-benefits and selfish transactions.

In the times of the origin of the theoretical basis of population genetics (basis of the "Theory Modern Synthetic") the existing genetic knowledge about the processes and mechanisms were very limited. Although the concept of transmission characters according to Mendelian inheritance type was a simplification of some processes today we know that they are really much more complex (Buchanan et al., 2009). Evolution of life is a process of complex systems integrating to other systems, integrating higher levels (Kauffman, 1993; Margulis and Sagan, 1995; Johnson and Coffin, 1999; Doolittle, 2000; Gupta, 2000; Davidson and Erwin, 2006). The components of basic units of bacteria that would have all the processes and mechanisms of cellular life appear to have been preserved with very few changes along the evolutionary process. Viruses, by chromosomal integration mechanism, which would, either individually or through their combination, introduce new sequences responsible for controlling embryonic development of new tissues and organs, as well as regulating its operation. It seems that association and cooperation have been underrated in the biology that only sees a battle for life in nature.

Since the inception of natural selection and the Darwinian view of nature, the definition of life is skewed. Nowadays, the discover of giant viruses, mimivirus, the description of amoebae as genitors of new microorganisms, the attempt to understand the evolutionary history of eukaryotic Nucleocytoplasmic Large DNA Viruses (NCLDV), focuses the attention on the fundamental question of the definition of life (Raoult, 2010a).

The concept of autopoiesis was introduced by Humberto Maturana and Francisco Varela (Varela et al., 1974). It considers a living system as a dynamic composite entity, a unity as a closed network of productions of components in away through their interactions in composition and decomposition, the components: (1) recursively constituted the same network of production that produced them, and (2) specify the extension of the network and constitute operational boundaries that separate it as a dynamic unity in a space defined by elements of the kind of those that compose it. It is an autopoietic system (Maturana, 2002). The word autopoiesis connotes the organization of living systems as closed networks of molecular production. Living systems exist only as long as their autopoietic organization is conserved. "Autopoiesis is the actual manner of being as the organization that constitutes living systems as singular entities in the molecular space" (Maturana, 2002).

Structural changes in the living system are foreign to the characterization of an observer and external or internal, but they occur contingent on structural meeting with the environment. In forming a lineage of living beings, what defines the lineage is the maintenance of autopoiesis over generations. For Maturana, biodiversity is the result of the formation and transformation of lineages in a continuous *phylogenetic coderiva.* That is, during the continuity of a lineage of living beings, an ontogenic phenotype is conserved in a reproductive sequence. This occurs in a systemic dynamics and not in a genetic one. The systemic genotype may change but the lineage may be kept. The new lineage will emerge, depending on the conditions that are systemic to this effect, as a variant of the original whenever the new ontogenic phenotype is preserved systemically (Maturana and Mpodozis, 1999). A system, facing a profound environmental change, may respond with a structural quantum leap or collapsing (general theory of systems). Organisms arise due to a structural dynamics independent of them. Nothing happens during this diversification that can be called selective force or pressure. An observer may notice a differential survival of different kinds of organisms that constitute a population (we can remember the dubious experiment of the peppered moth), but the observer cannot affirm that what led to this survival differential was a selection. This historic result from the phylogenetic deriva is the consequence of a systemic process in which there is no "pressure." To the extent that living beings are autopoietic systems that exist in ontogenic structural coderiva and breed in conditions of conservation of organization and adapt or die, producing lineages and ontogenetic phenotypic variations, are spontaneous and inevitable processes (Maturana and Mpodozis, 1999).

To this holistic new perspective, it is possible to enumerate different topics in which the results interpreted within the Darwinist preconceptions arise other interpretations that allow a better assessment of the facts observed.

ANTIBIOTICS

Antibiotics, which are the main molecules used by microorgamisms as weapons in the Darwinian view, are now re-studied as molecular signals (Linares et al., 2006; Fajardo and Martínez, 2008; Jayaraman, 2009).

It is known that subminimal inhibitory concentrations of antibiotics could produce subtle changes in bacterial physiology. The behavior of the bacterial population is an integrated response to different cell-to-cell signals (Martinez et al., 2009a). At concentrations found naturally in the environment where the organism

lives as producer, the main effects are general metabolism, changing patterns of transcription in a dose-dependent (Tsui et al., 2004; Yim et al., 2006a,b; Fajardo and Martínez, 2008; Martinez et al., 2009a). Also, some of these changes are antibiotic-specific. Antibiotics that inhibit bacterial topoisomerases might, at low concentrations, trigger SOS response or enhance RNA stability and produce changes in DNA supercoiling (Linares et al., 2006), responses that are beneficial for the microorgamisms involved (Linares et al., 2006; Martinez et al., 2009a).

Under the traditional view, the generation of resistance to antibiotics used in clinical medicine would be an evolutionary strategy of pathogenic microorganisms, calling these resistance mechanisms as pathogenicity or virulence factors. The limited interpretation of antibiotics as weapons results in misinterpretation of resistance mechanisms as specific shields that confer the protection against the weapons. This is the implicit belief of many reports related to this topic but the mechanisms involved in this resistance are more complex and they are being elucidated.

Alternative functional roles for resistance elements are now being proposed. Firstly, the presence of an antibiotic resistance gene does not necessarily imply that its original role was to help resist the action of the antibiotic (Martinez et al., 2009a). Also, the incidence of bacteria carrying multidrug resistance (MDR) pumps is not limited to environments with a high antibiotic load. Pumps that extrude antibiotics instead of being "bacterial strategies" against humans appear to have the function of detoxification of intracellular antibiotics rather than resistance to external ones (Martinez et al., 2009b). Furthermore, it is necessary to remark that some of these MDR pumps can efflux signal compounds indicating that signaling networks may be important in triggering antibiotic resistance (Martinez et al., 2009a). In many cases, the expressions of MDR pumps are related in regulating Quorum Sensing homeostasis (Martinez et al., 2009b).

Microbial cell-signaling is a result of an integrated system, interrelated ecosystem that it is far away of be weapons against their neighbors. Horizontal gene tansfer (HGT) is related to developing of competence and these both processes could be triggered by agressive and stressing conditions. High (toxic and stressing) concentrations of antibiotics (that are rarely found in nature) are the consequence of human activity. This artificial selection results in the disseminationan of resistance by HGT (for example, the spread of integrons). This process (that the traditional dogma could called "exaptation") is a result of the antropogenic activity and the complex mechanisms involved demonstrate that these interrelated process change against a stressful condition to restore the homeostasis of the whole system.

Antibiotics are produced in normal conditions in nature by microorganisms that are in a physiological state similar to what in the laboratory is called stationary phase. At these levels these molecules are signals that mantain the homeostasis. This state (so-called "stress") triggers a signal and those genes related to resistance genes are activated, acting as extrusion of the signal molecules. In the case of other mechanism of resistence that involves enzymatic modification of antibiotics, implies the synthesis of enzymes that have a metabolic function primarily as precursors phosphorylate "antibiotic."

BACTERIOCINS

Bacteriocins are defined as antimicrobial proteinaceous compounds synthesized ribosomally by bacteria (Diep and Nes, 2002). Even though authors report that the ecological function of these peptides is not yet fully understood, they decide that bacteriocins represent an important component of the warfare in nature (Riley, 1998; Gillor et al., 2008; Desriac et al., 2010). However, this is a limited interpretation since it is a directional result of an experiment that is searching for inhibition. And this function, in great amount, is assumed as the ecological function.

Bacteriocins are defined only by a partial and forced effect while their role in the microbial ecology context is forgotten. But, a look to the reports related to bacteriocins reflect their function as signals in a more complex context that the darwinian view of bacterial compounds as weapons in the fight between microbial competitors for colonizing of the same niche (Riley, 1998; Riley and Wertz, 2002; Desriac et al., 2010). The evaluation, and screening of bacteriocins is achieved in a "five stars restaurant" of a broth in lab, and we obtain great amount of bacteriocins that are used to inhibit the growth of another strain. Subclass IIa bacteriocins, for example, recognize mannose phosphotransferase system in the membrane of producer and "target" strains. Different strains display different expression levels of a man-PTS gene that corresponded to the variation in bacteriocin sensitivity (Kjos et al., 2009). These peptides act as signals among cells of a species and could interact with their environment, that is, abiotic and biotic factors around (Perry et al., 2009a,b). The bacteria and other organisms respond and send other signal to the bacteriocin producer. As with antibiotics, the search for bacteriocins were performed in order to obtain inhibitors, and many researchers tend to consider that microbes use them with the same function (Cotter et al., 2005; Papagianni and Anastasiadou, 2009). This is allegorical, because the obsessive behavior to produce profitability replaces the study of the phenomenon itself. As the essence of Darwinism was born in liberal trade and prejudices are so ingrained, it is easy to fall into confusion and move an ideology and human behavior to nature.

VIRUSES

A key aspect is the role of viruses in evolution (Rohwer et al., 2009). Viruses have acquired a new interpretation based on their capacity to insert genomes in cells and they are recognized as a macroorganism with a huge collection of genes, most unknown that constitute the major planet's gene pool. The continuing sequenciation of phages and virus is a way to the unknown (Rosario et al., 2009). Continued virus gene rearrangements derived from virus particles have formed a mosaic gene that underlies the creation of new structures and the generation of new species (Tristem et al., 1995; Johnson and Coffin, 1999; Tristem, 2000; Casjens, 2003; Johnson, 2008).

Genomes of all living organisms are mosaic of genes. Eukariotyc genome has genes from bacterial, archaeal and viral origins. Similarly, organelles like mitochondria do not have a single common ancestor but likely have numerous ancestors, including proto-Rickettsiales, proto-Rhizobiales, and proto-Alphaproteobacteria, as well as current alphaproteobacterial species (Georgiades and Raoult, 2011). Lateral Gene Transfer

among intracellular bacteria allows the gene exchange between phylogenetically very different bacteria (Saisongkorh et al., 2010). The representation of the evolutionary pathway as a tree leading to a single common ancestor is incorrect and obsolete. Raoult suggests that the evolution of species looks much more like a rhizome (Raoult, 2010b). The evolutionary history of intracellular bacteria *Rickettsia felis* and mitochondria from *Reclinomonas americana, Homo sapiens, Pediculus humanus,* and *Saccharomyces cerevisiae* were represented in the form of a rhizome (Georgiades and Raoult, 2011; Merhej et al., 2011). It was also affirmed that "the tree of life is not sufficient to explain the chimeric structure of current genomes, and the theory of a single common ancestor and a top-down tree does not reflect our current state of knowledge" (Georgiades and Raoult, 2011). The integration of complex systems (von Bertalanffy, 1950) is an alternative to build a strong theoretical framework more adjust to facts and recent discoveries, since the recent comprehension of genome complexity it is not possible to be explained by a tautology of natural selection and random mutation.

Organisms arise from the integration of complex systems into one another. In these processes viruses play a fundamental role since their sequences are capable of integrating into the genomes in an "infective" way and become an essential part of multicellular organisms. There is evidence that viral sequences in the genome of complex organisms have content with "biological sense" i.e., appear as part of normal life processes, and have a serious role of carrier elements of complex genetic information (Sandín, 1997; Mattick and Gagen, 2001; Vitali et al., 2003; Mallet et al., 2004; Hamilton, 2006; Hunter, 2008; Forterre, 2010). The simultaneous sequence integration in several individuals (i.e., the integration of a complex system within another) changes radically not only the process and the identity of character-creating agent, but also the meaning of this process. These sequences are involved in regulating gene expression or codifying very similar proteins in different animal groups (Medstrand and Mag, 1998; Mi et al., 2000; Villareal and De Filippis, 2000; Jamain et al., 2001). In addition, there are clear differences between the endogenous retroviral populations (ERVs) of reptiles, birds, and mammals (Tristem et al., 1995) and between primate specific (Johnson and Coffin, 1999), which implies specificity in functional sequences.

A comparative study of virome, the viral community associated with human hosts, from cystic fibrosis and non-cystic fibrosis individuals host have revealed that disease and and non-diseased states are defined by metabolism and not by taxonomy. The non-diseased airway virome contains a set of shared core metabolic functions, which deviate strongly in the face of chronic disease (Willner et al., 2009). This represents that integration of viruses goes beyond the genetic record level but also at individual levels with great importance in metabolic processes and the adaptation of the host.

The presence of viral genome in a big percentage in prokaryotes and eukaryotes and their essential roles is a common phenomenon that highlights the great evolutionary importance. For instance, the action and expression of a gene derived from an ERV allows the formation of placenta in mammals (Mallet et al., 2004). The virus and ERVs are implicated in the most of the adaptative mutations in the last 500 millions of years. Retrotransposons have been identified involved in the regulation of genes related to the histocompatibility (McDonald, 1995), with expression in tissues of different tetra1-alphaglobulins human (Kim et al., 1989) as well as in other mammals and invertebrates (Dnig and Lipshitz, 1994).

In bacterial cells, viruses are related to the generation of micro-compartiments (Yeates et al., 2007) where they have regulatory and structural functions. Organelles as carboxisoma consist of thousands of protein subunits assembled in a viral-like structure or scaffold (Kerfeld et al., 2005) and genes that codify it are present in both autotrophic and heterotrophic bacteria. They are also found in bacteria considered pathogenic. The insertive nature of virus fits these observations. These findings would not be so "mysterious" if one could think from another perspective. However, the only interest seems to be, again, developing a strategy to fight against bacteria (Yeates et al., 2007). Integrative capacities of the virus added to their great genetic diversity of this extraordinary gene pool (Bruüssow and Hendrix, 2002) ($>10^{30}$ tailed phages in the biosphere) constitutes an opportunity to strengthen the observation of its role in the evolutionary process. The authors remark that "micro-compartments could have evolved by divergent evolution with bacteria 'capturing' a virus and using both its genes and structural features for its own ends." Under this teleological explanation it seems that selfishness of nucleotide sequences and bacteria (which are sometimes selfish, sometimes exploited, sometimes exploitative) lead them in some remote past to capture and exploit virus. But scientific evidence can ensure that the demonstrated ability of the virus to insert itself into chromosomes (integrating complex systems) is what allowed the structural, morphological change, in this case is the appearance of a carboxisoma. The structural changes that imply evolution and the mechanisms are viral insertion that permits an evolutionary quantum leap.

MICROBIOME AND HOLOGENOME

Microbiome is the collective genome of our indigenous microbes (microbiota). The term also applies as a synonym of microbiota since "biome" refers to "ecosystems" in ecology (Lederberg and McCray, 2001; Dominguez-Bello and Blaser, 2008). Gut microbiome is taxonomically complex, it constitutes an ecologically dynamic community and it influences development, maturation, regulation (stimulation and suppression) of the immune system (Mazmanian et al., 2005; Smits et al., 2005; Hattori and Taylor, 2009; Mai and Draganov, 2009; Kau et al., 2011). Microorganisms have also been implicated in vitamin production, digestion, energy homeostasis, integrity of intestinal barrier, and angiogenesis in the human body (Dominguez-Bello and Blaser, 2008; Kau et al., 2011; Rosenberg and Zilber-Rosenberg, 2011; Slonczewski and Foster, 2011). Works with gnotobiotic mice (also known as germ-free mice, i.e., mice that are born in aseptic conditions and reared in a sterile or microbially controlled laboratory environment) demonstrate that the painstaking separation of a mammal from its associated microbiome results in an underdeveloped immune system, longer digestion times, and lower metabolic rates than those that have been normally colonized (Wostmann, 1981). Alterations of this microbiome could potentially affect human health and promote disease state or disbiosis (Rogler, 2010).

Either by cell number or by genome size the microbiota outnumbers their host. The hologenome theory considers that the holobiont, an organism and all of its associated symbiotic microbes, including parasites, mutualists, synergists, and amensalists as a result of *symbiopoiesis*, or codevelopment of the host and symbiont (Margulis and Fester, 1991; Rohwer et al., 2009; Gilbert et al., 2010; Rosenberg and Zilber-Rosenberg, 2011). This evolutionary approach that considers any organism as a result of integration with microorganisms has many implications and it is related to the Bioma Depletion Theory (also called "hygiene hypothesis") that considers that humans (and all mammals) and their microbiome evolved as a "superorganism" (Kinross et al., 2008; Rook, 2009). The immune system can be seen as having evolved as an interface with symbiotic organisms more than as a defense against invading organisms. The widely appreciated medical care in combination with technology, increased the occurence of allergic disorders, autoimmune diseases, and left us an over-reactive immune response caused by a loss and separation of our partners, our microbiome that normally interact with our immune system (**Figure 1**) (Garn and Renz, 2007; Kau et al., 2011). These partners involve not only the commensal bacteria, but metazoans "parasites" and millions of virus. Bacteria comprising the microbiome have mobile elements that include plasmids, transposons, integrons, bacteriophages (Jones, 2010) that constitute the *mobilome* (Siefert, 2009). This genetic pool and the HGT within the microbiome is a key factor of the microbiome activity and constitute the dynamic response to the environment leading to the adaptation of the holobiont. It fuels the adaptive potential of the whole holobiont (**Figure 1**). The metabolism of microbiome and the host are intertwined constituting an integrated organism. In multicellular eukaryotes, transposition, genome reorganizations, retrovirus extrusion, or insertion, etc., must be taking place in the germ line to result in a structural or metabolic change. Somatic cells have an intragenomic dynamics in response to environmental conditions.

Vannier-Santos and Lenzi (2011) explain that taking into account that organisms identified as "parasites" are almost the 80% of known species and considering that all the theoretical explanation obtained are based in just a little part of the total organisms that exist (Windsor, 1998), we can refer to parasites as cohabitants, since the association drives the evolution and existence of the organisms (Vannier-Santos and Lenzi, 2011). Microbes, helmints, that normally are understood as parasites have cohabited with their host and they are even greater than the

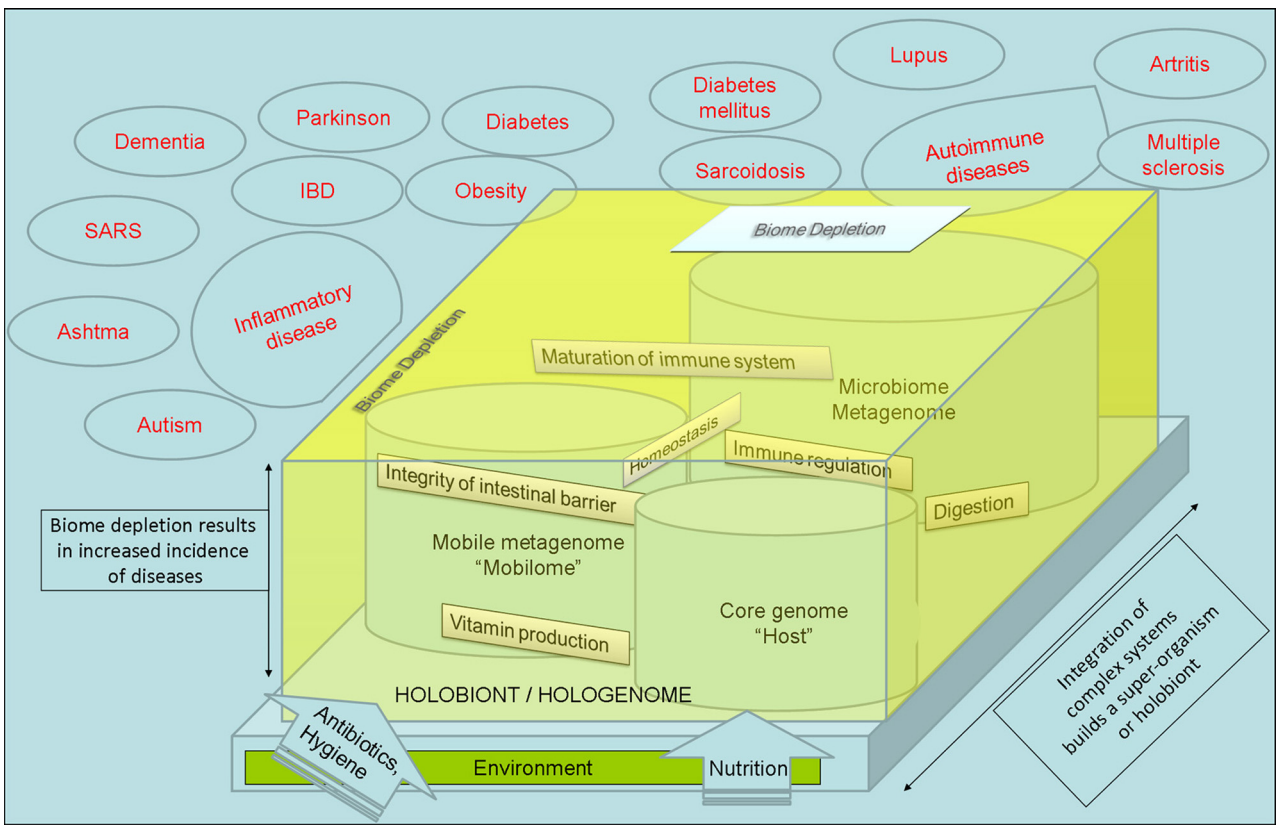


FIGURE 1 | The Integration of Complex Systems considers that any superorganism or holobiont is the result of integration of pre-existing systems. Mobile elements or "mobilome" respond to the environmental factors with dynamic movement between genomes that constitutes a key mechanism for metabolic and structural changes on microbiome. The metabolism of microbiome and the host are intertwined constituting an integrated organism. The medical care, use of antibiotics, technology, and western way of life, resulted in a change and lost of our microbiome and an increased occurence of autoimmune and metabolic diseases that are related with an immune disbalance.

host. If nature is a continuous battle, bacteria and parasites should have won a long time ago. Considering that Life exists as a net, as a process (Maturana and Varela, 1999) it is possible to say that no organism is a free-living specie in *sensu stricto*.

The host and its symbiotic microbiota with its hologenome, acts in cooperation (that becomes cooperation a priority instead of competition) and suggests that it should be considered a unit of selection in evolution (Zilber Rosenberg and Rosenberg, 2008). Even when the authors remark that the theory is in agreement with Darwinism, the hologenome theory represents a holistic approach that considers each specie or organism as a result of an integration and this is a mechanism that is observed at every level of nature: integration of virus, endosymbiotic relationships, and holobionts. This paradigm (like symbiogenesis from Merenchovzky and Margulis) contrasts the observable facts in nature against the individualistic, selfish, and economist conception of Darwinism.

The hologenome theory and these holistic approaches are in agreement with the autopoiesis concept of Maturana and Varela (Varela et al., 1974; Maturana and Varela, 1999; Maturana, 2002) and it could be interpreted as a continuity of the Lynn Margulis endosymbiotic theory (Margulis and Fester, 1991): the existence of each organism is the consequence of integration of pre-existing organisms. The genome of each organism is the result of combination of bacterial, virus, and eukaryotic DNA. Finally, the organism is the result of the interaction of their own genome with the genome of the microbiota (the hologenome), and their metabolism was and are intertwined (as a "superorganism") along evolution (Zilber Rosenberg and Rosenberg, 2008; Gazla and Carracedo, 2009; Kau et al., 2011; Tilg and Kaser, 2011; Vannier-Santos and Lenzi, 2011) (**Figure 1**).

CONCLUSION

We cannot ignore that competition exists, but giving it a creative sense, as an evolutionary engine is an overestimation. Darwin based his theory on economic thoughts of liberal trade put forward by Adam Smith and also the theories of Malthus and Spencer (Weikart, 2009). To this, he added the projection of social and cultural values and worldview of their own time on Nature. Thus, an economic system (and an ideology) was projected on nature. Everything is understood according to cost-benefit and organisms are in a warfare where they are exploiting each other, they produce "weapons," they have social dilemmas and cooperation is a consequence of a "mafia strategy" (Dawkins, 1976; Nogueira et al., 2009).

Natural selection is a linguistic trap. It has many definitions and nuances along the literature and just adds more confusion to the interpretations of facts. It appears that many biologists seem to be unaware that in their anti-creationism, they have replaced one dogma for another, the dogma of the all-powerful natural selection to which they cling with so much faith.

In order to fit the continuous discoveries innumerable metaphors were created, based mainly on economic relations of society (Ball, 2011). Nowadays, the abuse of "personification" (for example, speaking of selfish genes) and metaphors to explain the components and phenomena of nature are common (Ball, 2011). Many hypotheses, concepts and terms that were purely speculative became unquestioned concepts which were welded to the scientific language. They were used systematically to explain everything. The abuse of terms such as competition between proteins and between genes, the selection pressure, fitness, cost-benefit ratios, arsenal, weapons, war, exploitation, self-serving punishment, coercive strategies, mafia, policing (Boyd, 2006; Cant and Johnstone, 2006; Lehmann and Keller, 2006), the destruction of others, the "problem" of altruism, and many others expressions that attempt to explain the relationships between organisms, denote a continuation of a theory with an important ideological basis and a lot of subjective and moral categorizations. Even when a metaphor clearly based on market could be used to explain a relationship between organisms, it assumes as true that life follows the capitalism rules. They are antropocentric projections of dogmas and social economic models.

The data shows us that integration of complex systems into other complex systems as a result of a property of life: autopoiesis (Varela et al., 1974) is a priority instead of competition as the engine of evolution, stressing the importance of self-organization and symbiosis. Integration is a pattern that it is observed at every level: virus and phages "living" in an intracellular state, where they participate actively in the metabolism and in the plasticity of the genome, bacteria forming complex populations, bacteria living inside eucariotic hosts while existing metabolic and genomic exchanges, bacteria and "parasites" have cohabited for thousands of years with their host/cohabitant and co-evolving constituting an holobiont with deep and complex metabolic intertwined. Integration, partnership, symbiosis, viral insertion, etc., are mechanisms that cause evolutionary steps. A change in the approach and the appraisement of these processes will have no need for twisted excuses to explain the "strange phenomenon" of cooperation.

Viruses and bacteria share the double condition of pathogen and the basic unit of life. They have been fundamental in the origin of complex living beings. Their "negative" aspect would be the result of some factor breaking the natural balance of its activities (release of endogenous virus particles, expression of virulence genes) (Gabus et al., 2001; Kho et al., 2004; Seifarth et al., 2005).

Holistic perspectives are emerging strongly based on experimental data but a stride is still necessary to remove of our biological language, many metaphors and prejudices based on market theories that do not reflect what actually occurs in nature. Gaining a comprehensive understanding of the human being as an organism resulting from the integration of systems and understanding the processes of life within the framework of Systems Theory (von Bertalanffy, 1950) can make a better approach to the pathologies that result from the imbalance of our biome.

The presented new interpretations of different facts and discoveries are just a few examples that could be enumerated, but only a deeper interdisciplinary work can go further in the development of a new perspective on the theoretical foundations of evolutionary theory. Autopoiesis, symbiopoiesis, and evolution of biological systems by integration of complex systems are emergent theories that take into account facts and biological properties instead of economical transactions and are

plausible explanations to understand biological diversity and evolutionary process. This could make possible more accurate interpretations of biological processes as well as a new perception and attitude toward nature. It is necessary that biology allow the emergence of other points of view and alternative analysis, otherwise it is a dogmatic discipline of unique thinking and with a great deal of faith.

ACKNOWLEDGMENTS

To Dr. Máximo Sandín for having read this work.

REFERENCES

Abdalla, M. (2006). La crisis latente del darwinismo. *Asclepio* 63, 1.

Ball, P. (2001). Ideas for a new biology. *Nature*. http://www.fractal.org/Life-Science-Technology/Publications/Ideas-for-a-new-biology.htm. Last access April 18, 2012.

Ball, P. (2011). A metaphor too far. *Nature*. doi: 10.1038/news.2011.115. http://www.nature.com/news/2011/110223/full/news.2011.115.html. Last access April 18, 2012.

Barton, R. (1998). Huxley, Lubbock, and half a dozen others: professionals and gentlemen in the formation of the X-club, 1851–1864. *Isis* 89, 410–444.

Bernhard, R. (1967). Heresy in the halls of biology: mathematicians question darwinism. *Sci. Res.* 2, 59–66.

Bouchard, F., and Rosenberg, A. (2004). Fitness, probability and the principles of natural selection. *Br. J. Philos. Sci.* 55, 693–712.

Boyd, R. (2006). Reciprocity: you have to think different. *J. Evol. Biol.* 19, 1380–1382.

Boyer, M., Madoui, M. A., Gimenez, G., La Scola, B., and Raoult, D. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLoS One* 5:e15530. doi: 10.1371/journal.pone.0015530

Bruüssow, H., and Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell* 108, 13–16.

Buchanan, A. V., Sholtis, S., Richtsmeier, J., and Weiss, K. M. (2009). What are genes "for" or where are traits "from"? What is the question? *Bioessays* 31, 198–208.

Cant, M., and Johnstone, R. (2006). Self-serving punishment and the evolution of cooperation. *J. Evol. Biol.* 19, 1383–1385.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49, 277–300.

Cervantes, E. (2011a). Charles Darwin, o el origen de la máquina incapaz de distinguir. *Digital CSIC*. http://digital.csic.es/handle/10261/35958

Cervantes, E. (2011b). Economía semántica para la manipulación del conocimiento: la palabra Evolución y su uso como trampa en "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life". *Digital CSIC*. http://digital.csic.es/handle/10261/31352

Child, C. M. (1941). *Patterns and Problems of Development*. Chicago, IL: University of Chicago Press.

Cotter, P. D., Hill, C., and Ross, R. P. (2005). Bacteriocins: developing innate immunity for food. *Nat. Rev. Microbiol.* 3, 777–788.

Darwin, C. (1869). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 5th Edn. London: John Murray.

Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Desriac, F., Defer, D., Bourgougnon, N., Brillet, B., Le Chevalier, P., and Fleury, Y. (2010). Bacteriocin as weapons in the marine animal-associated bacteria warfare: Inventory and potential applications as an aquaculture probiotic. *Mar. Drugs* 8, 1153–1577.

Diep, I. B., and Nes, I. F. (2002). Ribosomally synthesized antibacterial peptides in Gram positive bacteria. *Curr. Drug Targets* 3, 107–122.

Dnig, D., and Lipshitz, H. D. (1994). Spatially regulated expression of retrovirus-like transposons during Drosophila melanogaster embryogenesis. *Genet. Res.* 64, 167–181.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129.

Dobzhansky, T., Ayala, F., Stebbins, G., and Valentine, J. (1977). *Evolution*. San Francisco, CA: W.H. Freeman.

Dominguez-Bello, M., and Blaser, M. (2008). Do you have a probiotic in your future? *Microbes Infect.* 10, 1072–1076.

Doolittle, W. F. (2000). Uprooting the tree of life. *Sci. Am.* 282, 90–95.

Ehrlich, P., and Birch, L. (1967). Evolutionary history and population biology. *Nature* 214, 349–352.

Fajardo, A., and Martínez, J. L. (2008). Antibiotics as signals that trigger specific bacterial responses. *Curr. Opin. Microbiol.* 11, 161–167.

Forterre, P. (2010). Giant viruses: conflicts in revisiting the virus concepts. *Intervirology* 53, 362–378.

Futuyma, D. J. (2009). *Evolution*. 2nd ed. Sunderland, MA: Sinauer Associates.

Gabus, C., Derrington, E., Leblanc, P., Chnaiderman, J., Dormont, D., Swietnicki, V., Morillas, M., Surewicz, M., Marc, D., Nandi, P., and Darlix, J. (2001). The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. *J. Mol. Biol.* 307, 1011–1021.

Garn, H., and Renz, H. (2007). Epidemiological and immunological evidence for the hygiene hypothesis. *Immunobiology* 212, 441–452.

Gazla, I. N., and Carracedo, M. C. (2009). Effect of intracellular Wolbachia on interspecific crosses between *Drosophila melanogaster* and *Drosophila simulans*. *Genet. Mol. Res.* 8, 861–869.

Gerstein, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681.

Georgiades, K., and Raoult, D. (2011). The rhizome of *Reclinomonas americana*, *Homo sapiens*, *Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. *Biol. Direct* 6, 55.

Gilbert, S., McDonald, E., Boyle, N., Buttino, N., Gyi, L., Mai, M., Prakash, N., and Robinson, J. (2010). Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 671–678.

Gilbert, S. F., Opitz, J. M., and Raffs, R. A. (1996). Resynthesizing evolutionary and developmental biology. *Dev. Biol.* 173, 357–372.

Gillor, O., Etzion, A., and Riley, M. A. (2008). The dual role of bacteriocins as anti and probiotics. *Appl. Microbiol. Biotechnol.* 81, 591–606.

Goldsmith, E. (1989). Gaia and evolution. *Ecologist* 19, 147–153.

Gupta, R. S. (2000). The natural evolutionary relationships among prokaryotes. *Crit. Rev. Microbiol.* 26, 111–131.

Hamilton, G. (2006). Virology: the gene weavers. *Nature* 441, 683–685.

Harrison, R. G. (1937). Embriology and its relations. *Science* 85, 3691–374.

Hattori, M., and Taylor, T. (2009). The human intestinal microbiome: a new frontier of human biology. *DNA Res.* 16, 1–12.

Hunter, P. (2008). Not so simple after all. A renaissance of research into prokaryotic evolution and cell structure. *EMBO Rep.* 9, 224–226.

Jamain, S., Girondot, M., Leroy, P., Clergue, M., Quach, H., Fellous, M., and Bourgeron, T. (2001). Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. *Genomics* 78, 38–45.

Jayaraman, R. (2009). Antibiotic resistance: an overview of mechanisms and a paradigm shift. *Curr. Sci.* 96, 1475–1484.

Johnson, W. (2008). A proviral puzzle with a prosimian twist. *PNAS* 10, 20051–20052.

Johnson, W., and Coffin, J. (1999). Constructing primate phylogenies from ancient retrovirus sequences. *PNAS* 96, 10254–10260.

Jones, B. (2010). The human gut mobile metagenome. A metazoan perspective. *Gut Microbes* 1, 415–431.

Kampis, G. (1997). "Evolution as its own cause and effect," in *Evolutionary Systems*, eds S. Salthe and G. van de Vijver (Kluwer, Dordrecht), 255–266.

Kau, A., Ahern, P., Griffin, N., Goodman, A., and Gordon, J. (2011). Human nutrition, the gut microbiome and the immune system. *Nature* 474, 327–336.

Kauffman, S. A. (1993). *Origins of Order*. New York: Oxford University Press.

Kerfeld, C., Sawaya, M., Tanaka, S., Nguyen, C., Phillips, M., Beeby, M., and Yeates, T. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science* 309, 936–938.

Kho, A., Zhao, Q., Cai, Z., Butte, A., Pomeroy, S., Rowitch, D., and Kohane, I. (2004). Conserved mechanisms across development and tumorigenesis revealed by a mouse

development perspective of human cancers. *Genes Dev.* 18, 629–640.

Kim, J., Yu, C., Bailey, A., Hardison, R., and Shen, C. (1989). Unique sequence organization and erythroid cell-specific nuclear factor-binding of mammalian theta-1 globin promoters. *Nucleic Acids Res.* 17, 5687–5700.

Kinross, J., von Roon, A., Holmes, E., Darzi, A., and Nicholson, J. (2008). The human gut microbiome: implications for future health care. *Curr. Gastroenterol. Rep.* 10, 396–403.

Kjos, M., Nes, I. F., and Diep, D. B. (2009). Class II one-peptide bacteriocins target a phylogenetically defined subgroup of mannose phosphotransferase systems on sensitive cells. *Microbiology* 155, 2949–2961.

Lederberg, J., and McCray, A. (2001). "Ome Sweet" Omics-A Genealogical Treasury of Words. *Scientist* 15, 8.

Ledford, H. (2008). Human genes are multitaskers. *Nature* 456, 9.

Lehmann, L., and Keller, L. (2006). The evolution of cooperation and altruism-a general framework and a classification of models. *Evol. Biol.* 19, 1365–1376.

Linares, J., Gustafsson, I., Baquero, F., and Martinez, J. (2006). Antibiotics as intermicrobial signaling agents instead of weapons. *PNAS* 103, 19484–19489.

Mai, V., and Draganov, P. (2009). Recent advances and remaining gaps in our knowledge of associations between gut microbiota and human health. *World J. Gastroenterol.* 7, 81–85.

Mallet, F., Bouton, O., Prud'homme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *PNAS* 101, 1731–1736.

Margulis, L., and Fester, R. (1991). *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. Boston, MA: MIT Press.

Margulis, L., and Sagan, D. (1995) *What is Life?* New York, London: Simon and Schuster.

Martinez, J., Fajardo, A., Garmendia, L., Hernández, A., Linares, J., Martínez-Solano, L., and Sánchez, M. (2009a). A global view of antibiotic resistance. *FEMS Microbiol. Rev.* 33, 44–65.

Martinez, J., Sanchez, M., Martinez-Solano, L., Hernandez, A., Garmendia, L., Fajardo, A., Garmendia, L., Fajardo, A., and Alvarez-Ortega, C. (2009b). Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol. Rev.* 33, 430–449.

Mattick, J., and Gagen, M. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* 18, 1611–1630.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybern. Hum. Knowing* 9, 3–4.

Maturana, H., and Mpodozis, J. (1999). *De l'origin des spèces par voie de la dèrive naturelle*. Eds. Presses Universitaries de Lyon.

Maturana, H., and Varela, F. (1999). *The Tree of Knowledge. The Biological Basis of Human Knowledge* (3rd Edn.). Madrid: Debate.

Mayr, E. (1966). *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.

Mazmanian, S., Liu, C., Tzianabos, A., and Kasper, D. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122, 107–118.

McDonald, J. (1995). Transposable elements: possible catalysts of organismic evolution. *Trends Ecol. Evol.* 10, 123–126.

Medstrand, P., and Mag, D. L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72, 9782–9787.

Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2011). The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences. *Mol. Biol. Evol.* 28, 3213–3223.

Mi, S., Lee, X., Li, X., Veldman, G., Finnerty, H., Racie, L., Lavallie, E., Tang, X., Edouard, P., Howes, S., Keith, J., and Mccoy, J. (2000). Syncitin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789.

Nedelcu, A., Driscoll, W., Durand, P., Herron, M., and Rashidi, A. (2011). On the paradigm of altruistic suicide in the unicellular world. *Evolution* 65, 3–20.

Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., and Rocha, E. P. (2009). Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.* 20, 1683–1691.

Papagianni, M., and Anastasiadou, S. (2009). Pediocins: the bacteriocins of Pediococci. Sources, production, properties and applications. *Microb. Cell Fact.* 8, 3.

Perry, J. A., Cvitkovitch, D. G., and Lévesque, C. M. (2009a). Cell death in Streptococcus mutans biofilms: a link between CSP and extracellular DNA. *FEMS Microbiol. Lett.* 299, 261–266.

Perry, J. A., Jones, M. B., Peterson, S. N., Cvitkovitch, D. G., and Lévesque, C. M. (2009b). Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence. *Mol. Microbiol.* 72, 905–917.

Peters, R. H. (1976). Tautology in evolution and ecology. *Am. Nat.* 110, 11–12.

Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Barcelona: Paidós, 1981.

Raoult, D. (2010a). Giant viruses from amoeba in a post-Darwinist viral world. *Intervirology* 53, 251–253.

Raoult, D. (2010b). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Riley, M. (1998). Molecular mechanisms of bacteriocin evolution. *Annu. Rev. Genet.* 32, 255–278.

Riley, M. A., and Wertz, J. E. (2002). Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.* 56, 117–137.

Rogler, G. (2010). The importance of gut microbiota in mediating the effect of NOD2 defects in inflammatory bowel disease. *Gut* 59, 153–154.

Rohwer, F., Prangishvili, D., and Lindell, D. (2009). Roles of viruses in the environment. *Environ. Microbiol.* 11, 2771–2774.

Rook, G. (2009). Review series on helminths, immune modulation and the hygiene hypothesis: the broader implications of the hygiene hypothesis. *Immunology* 126, 3–11.

Rosario, K., Nilsson, C., Lim, Y., Ruan, Y., and Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* 11, 2806–2820.

Rosenberg, E., and Zilber-Rosenberg, I. (2011). Symbiosis and development: the hologenome concept. *Birth Defects Res. C Embryo Today* 93, 56–66.

Saisongkorh, W., Robert, C., La Scola, B., Raoult, D., and Rolain, J. M. (2010). Evidence of transfer by conjugation of type IV secretion system genes between Bartonella species and Rhizobium radiobacter in amoeba. *PLoS One* 13:5. doi: 10.1371/journal.pone.0012666

Sandín, M. (1997). Synthetic theory: crisis and revolution. *Arbor* 158, 623–624.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science* 323, 737–741.

Schützemberger, M. (1967). "Algorythms and the neo-darwinian theory of evolution," in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, eds P. S. Moorhead and M. M. Kaplan (Filadelfia: Wistar Institute Press), 73–80.

Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A. D., Hehlmann, R., and Leib-Mösch, C. (2005). Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *J. Virol.* 79, 341–352.

Siefert, J. (2009). "Defining microbiome," in *Horizontal Gene Transfer Genomes in Flux*, eds M. B. Gogarten, J. P. Gogarten, and L. Olendzenski, Vol. 532, (New York, NY: Humana Press), 13–27.

Slonczewski, J., and Foster, J. (2011) *Microbiology: An Evolving Science*. New York, NY: W. W. Norton and Co.

Smits, H. H., Engering, A., van der Kleij, D., de Jong, E., Schipper, K., van Capel, T., Zaat, B., Yazdanbakhsh, M., Wierenga, E., van Kooyk, Y., and Kapsenberg, M. (2005). Selective probiotic bacteria induce IL-10-producing regulatory T cells *in vitro* by modulating dendritic cell function through dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin. *J. Allergy Clin. Immunol.* 115, 1260–1267.

Tilg, H., and Kaser, A. (2011). Gut microbiome, obesity, and metabolic dysfunction. *J. Clin. Invest.* 121, 2126–2131.

Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* 74, 3715–3730.

Tristem, M., Myles, T., and Hill, T. (1995). A highly divergent retroviral sequence in the tuatara (Sphenodon). *Virology* 210, 1.

Tsui, W., Yim, G., Wang, H., McClure, J., Surette, M., and Davies, J. (2004). Dual effects of MLS antibiotics: transcriptional modulation and interactions on the ribosome. *Chem. Biol.* 11, 1307–1316.

Vannier-Santos, M., and Lenzi, H. (2011). Parasites or cohabitants: cruel omnipresent usurpers or creative "éminences grises"? *J. Parasitol. Res.* doi: 10.1155/2011/214174. [Epub ahead of print].

Varela, F., Maturana, H., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* 5, 187–196.

Villareal, L., and De Filippis, V. (2000). A hypothesisi for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* 74, 7079–7084.

Vitali, P., Royo, H., Seitz, H., Bachellerie, J. P., Hüttenhofer, A., and Cavaillé, J. (2003). Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.* 31, 6543–6551.

von Bertalanffy, L. (1950). An outline of general system theory. *Br. J. Philos. Sci.* 1, 139–164.

Weikart, R. (2009). Was Darwin or Spencer the father of laissez-faire social darwinism? *J. Econ. Behav. Organ.* 71, 20–28.

Weiss, P. (1939). *Principles of Development*. New York, NY: Holt.

Wilkins, J. S., and Nelson, G. J. (2008). Trémaux on species: a theory of allopatric speciation (and punctuated equilibrium) before Wagner. *Hist. Philos. Life Sci.* 30, 179–206.

Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F. E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., and Rohwer, F. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 4:e7370. doi: 10.1371/journal.pone.0007370

Windsor, D. (1998). Most of the species on earth are parasites. *Int. J. Parasitol.* 28, 1939–1941.

Woese, C. (2004). A new biology for a new century. *Microbiol. Mol. Biol. Rev.* 68, 173–186.

Wostmann, B. (1981). The germ-free animal in nutritional studies. *Annu. Rev. Nutr.* 1, 257–279.

Wright, B. E. (2000). A biochemical mechanism for nonrandom mutations and evolution. *J. Bacteriol.* 182, 2993–3001.

Yeates, T., Tsai, Y., Tanaka, S., Sawaya, M., and Kerfeld, C. (2007). Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells. *Biochem. Soc. Trans.* 35, 508–511.

Yim, G., de la Cruz, F., Spiegelman, G., and Davies, J. (2006a). Transcription modulation of Salmonella enterica serovar Typhimurium promoters by sub-MIC levels of rifampin. *J. Bacteriol.* 188, 7988–7891.

Yim, G., Wang, H., and Davies, J. (2006b). The truth about antibiotics. *Int. J. Med. Microbiol.* 296, 163–170.

Zilber Rosenberg, I., and Rosenberg, E. (2008). The hologenome theory of evolution. *FEMS Microbiol. Rev.* 32, 723–735.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 October 2011; accepted: 06 April 2012; published online: 01 May 2012.

Citation: Salvucci E (2012) Selfishness, warfare, and economics; or integration, cooperation, and biology. Front. Cell. Inf. Microbio. ***2****:54. doi: 10.3389/fcimb.2012.00054*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 12 July 2012
doi: 10.3389/fcimb.2012.00088

Intelligibility in microbial complex systems: Wittgenstein and the score of life

***Fernando Baquero*[1,2]* *and Andrés Moya*[2,3]**

[1] *Department of Microbiology, IRYCIS, Ramón y Cajal University Hospital, and Division of Extremophily and Evolutionary Biology, Centre for Astrobiology, INTA-CSIC, Madrid, Spain*
[2] *Network Research Center for Epidemiology and Public Health (CIBERESP), Madrid, Spain*
[3] *Cavanilles Institute on Biodiversity and Evolutionary Biology, Genomics and Health Area, Centre for Public Health Research, University of Valencia, Conselleria de Sanitat, Valencia, Spain*

Edited by:
Didier Raoult, Université de la Méditerranée, France

Reviewed by:
Didier Raoult, Université de la Méditerranée, France
Kalliopi Georgiades, Unité de Recherche en maladies infectieuses tropicales emergentes, France

***Correspondence:**
Fernando Baquero, Department of Microbiology, Ramón y Cajal University Hospital, Carr. Colmenar km 9,100, 28034 Madrid, Spain.
e-mail: baquero@bitmailer.net

Knowledge in microbiology is reaching an extreme level of diversification and complexity, which paradoxically results in a strong reduction in the intelligibility of microbial life. In our days, the "score of life" metaphor is more accurate to express the complexity of living systems than the classic "book of life." Music and life can be represented at lower hierarchical levels by music scores and genomic sequences, and such representations have a generational influence in the reproduction of music and life. If music can be considered as a representation of life, such representation remains as unthinkable as life itself. The analysis of scores and genomic sequences might provide mechanistic, phylogenetic, and evolutionary insights into music and life, but not about their real dynamics and nature, which is still maintained unthinkable, as was proposed by Wittgenstein. As complex systems, life or music is composed by thinkable and only showable parts, and a strategy of half-thinking, half-seeing is needed to expand knowledge. Complex models for complex systems, based on experiences on trans-hierarchical integrations, should be developed in order to provide a mixture of legibility and imageability of biological processes, which should lead to higher levels of intelligibility of microbial life.

Keywords: intelligibility, complex systems, Wittgenstein, metaphors, epistemology

INTRODUCTION

Life is a highly complex system, including the most complex objects in the known universe (Bedau, 1996). The genomics revolution has catapulted molecular biology, and particularly microbiology (Westerhoff and Palsson, 2004), into the realms of systems biology approaches to complex systems. Such a trend was based on the growing compelling intuition of the need of scaling-up molecular biology, in a new age of synthesis requiring formal integrative tools (Baquero, 2004, 2009). Biochemistry and lately Molecular Biology have shown that certain distinctive carbon-based macromolecules play a crucial role in the vital processes of all known living entities, but life seems to be more in the nature of a process (Bedau, 1996). The epistemological problem is how to cross the gap transitions between successive levels of understanding that corresponds to the different hierarchical levels of the complex system of life. Microbiologists are the best positioned scientists to respond to such a challenge, as they are familiar with "multiple-levels biology," dealing simultaneously with microbial collectives and collective genomes, as in metagenomics (Moya et al., 2012), cell-to-cell interactions (including pathogenesis, Desnues et al., 2010), the flowing biology of subcellular mobile genetic elements (Beiko et al., 2005) and finally with the wealth of gene-gene epistatic interactions (Babu et al., 2009).

Microbial communities, species, clones, plasmids, transposons, integrons, and genes are evolutionary individuals tracing their evolutionary trajectories at different hierarchical levels (Baquero, 2011). Such trans-hierarchical network-like complexity simply eliminates the possibility of identification of simple causal structures, if they ever exist out of our ways of representation (Schrodinger, 1957). The hope of a simple answer to the classic Baconian question in science "What is the cause of…" has no sense any more, and in fact the complex structure of biological processes constitute the major challenge for Biological Theory (Callebaut and Laubichler, 2007).

The challenge is not only to deal with quantitative integration of elements across these major hierarchical transitions in microbiology, or in biology at large (Maynard Smith and Szathmáry, 1997), but to eventually discover general principles of microbial life rather of just keep on descriptions (Westerhoff and Palsson, 2004). Such scaling-up process of understanding resembles the escalation from the *forms* of a language (as lexical or syntactic) toward its *meaning* (semantics) (Steels, 2004, 2010; Rosen, 2004). Interestingly, the trade-off between these hierarchical levels in a shared world has been defined as *intelligibility* in linguistic theory (Komarova and Niyogi, 2004). In his primary sense, the word *intelligibility* reflects the possibility of such a trans-hierarchical understanding. St. Thomas even

derives the Latin word *intelligere* from *intus legere*, or "reading into"; even if the origin were *inter legere*, the term stresses the need of reading beyond the words and sentences, in another cognitive dimension, to reach the meaning. Accordingly to the Cato's classic sentence, "*legere, et non intelligere, neglegere est,*" that is, "as good not read, as not to understand." In this work, we use the word "*intelligibility*" as the construction of meaningful (thinkable) models in response to the assimilation of knowledge (clear-and-distinct or fuzzy), and able to reflect to a certain extent the reality of complex natural systems, as those which are the objects of biological sciences. The conversion of data into knowledge constitutes a great challenge for future biological research (Brenner, 2010). In fact intelligibility is a prerequisite to developing modern biology grounded on a sound epistemology (Dougherty and Bittner, 2010). "*Legere, et non intelligere, neglegere est*". Two thousand years after Cato, Albert Einstein formulated essentially the same idea: "*Science without epistemology is—insofar as it is thinkable at all—primitive and muddled*" (Einstein, 1949).

Metaphors are frequently used by scientists as "non-logic" epistemological aids to think about reality (De Man, 1978). Biologists have long made use of linguistic metaphors in describing and naming cellular processes, and in particular involving from DNA as language to genome as a "book of life." The current questions are: (1) if these apparently immediate analogies might result in a deeper possibility of analysis of genetic-genomic structures using methods that have been developed in linguistic research; and (2) if such an analysis will enable to understand the general principles and processes of life, and even (not entering here again in the universal's problem) life itself as a intelligible entity.

The image of genome as a "book of life" has attracted popular imagination, but it is obvious that the knowledge of the entire genomic sequence of *Haemophilus influenzae* (Fleischmann et al., 1995) or *Homo sapiens* (McPherson et al., 2001; International Human Genome Sequencing Consortium, 2004) has not resulted in a much deeper understanding of the "life" clues of these organisms. This occurs not only because of our gaps in understanding the function of all genes and the complexities of regulatory and epigenetic interactions between genes and other meaningful sequences. Probably the human way of reading a text is simpler than the cell way of reading. Human language texts are read in one way only, sequentially and involving all characters. Genetic texts are "read" by cellular mechanisms in several different ways, each time using a different selection of the characters of the same text while skipping others. Indeed the "score of life" could be a better image of the genomic language. A score is a series of staves on which all the different instrumental and/or vocal parts of a musical work are written, one under the other in vertical alignment, so that the parts may be read simultaneously.

But in this report we consider there exist even bigger difficulties to predict how relevant is deciphering the language of genes, the "book of life." One key epistemological problem is to discuss about our ability to clarify the possible relations between the structure of a possible language (genetic and genomic sequences) and the characteristics of life of particular organisms, which seem to be determined by this language. Obviously this is a problem of reductionism (Wimsatt, 1976)—might the understanding of life be reduced to the understanding of the genetic-genomic language?

In the way of thinking of Ludwig Wittgenstein, and even more sharply in his friend and commenter, Moritz Schlick (Schlick, 1936), the meaning of the word "life" can only be *shown*, not understood and consequently not *clearly* expressed in propositions. It might sound paradoxical to attach to this statement in the age of glory of genomics, proteomics, and metabolomics. Life is a *fact* that can be *shown* (but not defined by) as something like a moving and loosely integrated complex of contingent structures, each one of them (and the complex itself) tending to be sequentially replaced by similar forms, and displaying various degrees of changes in variability and complexity both during almost instantaneous and long-term periods of time. Note that because we are here only *showing* life, this description does not assure that we are not confronted with non-living structures with similar properties, and certainly that any notion of progress or purposiveness cannot be considered here. Nevertheless, as the human *observers,* we are not neutral in the process of selecting what we would like to *show*, as frequently we are confronted with a non-descriptible feeling of sharing a common quality ("animation?") with what we tend to show as living things.

THE "SCORE OF LIFE" METAPHOR

The Ludwig Wittgenstein's "*Tractatus logico-philosophicus,*" published in English in 1922 under the guidance of Bertrand Russell, is widely recognized as one of the main post-kantian approaches devoted to explore the possibilities of human knowledge of natural world (Wittgenstein, 1921). In its theorem 4.0141, Wittgenstein compares music scores and gramophone (DVDs, in our times) with music.

4.0141

> In the fact that there is a general rule by which the musician is able to read the symphony out of the score, and that there is a rule by which one could reconstruct the symphony from the line on a gramophone record and from this again—by means of the first rule—construct the score, herein lies the internal similarity between these things which at first sight seem to be entirely different. And the rule is the law of projection which projects the symphony into the language of the musical score. It is the rule of translation of this language into the language of the gramophone record.

The order and qualities of the musical notes in the score, the grooves' irregularities in the gramophone record, in summary, the "language" from which music might be reproduced, is not music, but has an *internal similarity* with music. Much longer before the discovery of the genetic code, the Wittgenstein's theorem 4.0141 recalls the main structural feature of living organisms. The process of reading the score (genetic language), produces music (life); conversely, music can be converted, translated, by a "law of projection" into a musical score, and from this again music might be reconstructed. Without internal similarities, these transitions

between series of objects "that at first sight seem to be entirely different" should be simply impossible.

Interestingly in music, as in life, the description of "what is said" beyond the individual sounds is obscure. As the classic question of Erwin Schrodinger what is Life? The question: what is Music? refuses precise answers. No propositions are transmitted by music to describe clear and distinct facts, able to be thought (logically considered) by human mind. There is, as in genetics, a certain "arithmetic order" of notes that is required to produce obscure final effects. In Leibnitz words: "*exercitium arithmeticae occultum nesciendis se numere animi*" (*Leibnitii epistolae, collectio* Kortholdi, ep. 154), that is, music is as an unconscious arithmetic's exercise in which mind do not know what is being counted.

Maybe one of the difficulties of thinking life using linguistic structures is the fluid, dynamic nature of life. Languages, music score or genetic-genomic sequences, are essentially static. A book, or a musical score, or a genome sequence can be indefinitely stored without any alteration, and even more, without producing any effect (except covering a small parcel of physical space). On the contrary, speech, music, or life, are essentially dynamic; without movement they ceases existing. The fact that linguistic structures "contain" potential dynamicity does not make them dynamic at all; indeed they are practically nothing by themselves. The key-fact is that between languages and dynamic phenomena an *interpretative intermediary* should be interposed. The music score gives rise to music only if interpreters are available, musicians (four in a string quartet) able to read the language and converting it into sounds. Indeed a music score has an ordered internal structure, for instance following the rules of harmony, but, at first sight, we could conclude that in the absence of correct interpretation, a music score cannot be differentiated from a random sequence of notes.

Let us now imagine an out-of-Earth scientist examining a music score. He has no idea about notes, instruments, or sounds, even about the existence of music at all. Probably he will be able to differentiate a music score from a random sequence of notes. Some notations (notes) are preferentially linked to other ones, some conserved and iterated sequences are recognizable, the role of black and white notes seems not identical, some occur more frequently than others when accompanying the name a particular instrument (unknown). The note's frequency per decimeter of score apparently depends on some mysterious words at the margin as "*Andante scherzoso quasi alegretto*," that nevertheless might provide a "living equivalent." He could conclude that the musical score has a linguistic structure, potentially leading to an unknown type of dynamic behavior. If the out-of-Earth scientist could had access to a high number of different scores, he could even trace different schools, authors, influences, even a history of this unknown language—and probably he will not be much far from reality. In summary, an analytical "science" of this language could be built, and that in the total absence of knowledge about the nature of music.

Now note that the mirror process of analysis is also possible. In that case our second out-of-Earth scientist is observing the performance of a music group playing the Schubert's Piano Trio in B flat, D. 898. Unfortunately, he does not know about the existence of music, as he is unable to hear any sound, but he is able to distinguish the keys, bows, and strings of the different instruments and he can precisely record any movement of the player's arms and fingers on these structures. A representation of these movements during time should produce something similar to the musical score of the Piano Trio. Indeed the precise record of these movements might substitute the musical score, and when applied to the instruments should reproduce the music. As in the previous case, a collection of this type of records could lead to tracing schools or authors, or a history, or even a science of this language—without knowing what music is at all.

But we can also conceive a third out-of-Earth scientist, able to hear the sounds and to correlate them with the instruments and the movements of the players. It might well happen that the scientist could perceive the separate sounds, but he is either unable to link them in his mind as significant ensembles (melodies), or the sounds are so different in his brain than in ours, that our harmony is totally useless for his sensibility, and out of any esthetical possibility. As in the previous cases, this scientist could be able to study the history of music, without understanding at all what music is.

The essential is to discuss if a particular sequence of written musical notations, or sounds, or hands and finger movements, has only the meaning of "music" when understood by a particular type of sensibility. Even more: we can replace the "out-of-Earth" scientists by musicians, which will be able to reproduce the music without knowing anything about its nature, and without experimenting any of the effects that music might cause in the appropriate sensibility. They are in a "Chinese room" situation, in which the (considered to be intelligent) intermediate within the closed room receives below the door messages in an unknown language, but accordingly to a set of rules, he is able to produce responses in the same unknown language (Searle, 1984). It is obvious that the music score, or the genome sequence, is totally unaware of its function in the process of life, and the same is true for other possible intermediaries, for instance, involved in protein translation. In the words of Sydney Brenner, "genomes do not contain in any explicit form anything at a higher level than genes" (Brenner, 1999), or, paraphrasing Leibnitz when describing monads, "genomes do not have windows."

Therefore, neither from the outside, in which life can only be *shown* (and even that, without certitude), nor from the inside (life is invisible for life-determining structures), life seems to be thinkable. "We feel that even if *all possible* scientific questions be answered, the problems of life have still not been touched at all. Of course there is then no question left and just this is the answer" (Theorem 6.52). In other words, the answer is that to ask ourselves for the meaning of life is a false question, that is, there is nothing to think about. "For an answer which cannot be expressed the question too cannot be expressed" (6.5). And, as stated in the last sentence of the *Tractatus*, "Where of one cannot speak, thereof one must be silent" (6.54). The main interest of the "score of life" metaphor is probably that both life and music can be *shown* (as something that seems to impose a reality), *but not thought* (we cannot say anything about its reality), as the genomic sequences or the musical scores are mere representations of these obscure realities.

THINKABILITY OF LIFE

If life can be only *shown*, is Ludwig Wittgenstein right? Have the terms "*Bio*logy" and their derivatives, as "Micro*bio*logy," intrinsic epistemological contradictions? Are they non-sense proposals? We arrive now to an apparent contradiction. On one hand, we could reach the notion that life is unthinkable. But, as stated in 3.02, "the though contains the possibility of the state of affairs which it thinks. What is thinkable is also possible." On the other hand, life is *perceived* as a *fact*, therefore not only possible, but a realized entity. Obviously, if it exists, it should be thinkable. If life is not thinkable, either life does not exist, or has an unveiled, hidden reality. This antinomy is a clear variation of the Kantian ones, based on the confusion between the spheres of *phenomena* and *noumena*, and encapsulates the main problem that is discussed in this assay. Life is perceived, even *experienced* as a fact, but it is not a fact, it is not an entity. If that proposition were true, life should not be an object of natural science. Limiting the thinkable and thereby the unthinkable, philosophy limits the disputable sphere of natural science (Wittgenstein again, see 4.113–4.114).

Natural science should be thinkable and speakable. Everything that can be thought at all can be thought clearly. Everything that can be said can be said clearly (4.116). If life is unthinkable, but if it had a reality, we should think on "pictures of life" using articulated propositions, which are "models of reality as we think it is" (4.01). "The proposition constructs a world with the help of a logical scaffolding, and therefore one can actually see in the proposition all the logical features possessed by reality *if* it is true" (4.023). Wittgenstein's propositions could be considered as something derived from the combinatorial and ordered nature of structures as the musical score (3.141), or, in the life context genomes, again "representing" the (suspected) reality, but *unable to represent what they have in common* with reality (4.12). The gap is maintained between the unthinkable but presumed reality (life) and the thinkable picture of it (proposition).

Life is as thinkable as music is thinkable. In both cases, there is what Wittgenstein calls a certain "experience of meaning." Understanding life and music is to perceive "fine shades" of meaning. Intuitively both music and life seems to be meaningful, but in both cases they seem to be resistant to any "semantic" treatment. It has always been difficult to see how "meaning" could be fruitfully ascribed to music, as this notion is applied to language (Bar-Elli, 2006). Obviously the understanding of music has nothing to do with the ability to recreate in mind a memorized melody, or to foresee in the concert hall the next variation of a musical theme. We can say the same for life. We can predict, with a certain confidence, what will happen in the next step, based on our experiences, but that is not to understand—and even not to think about life. Experiences might provide a flavor of causality, following Hume, regularities in structured observations leads to expectation (Dougherty and Bittner, 2010), but that is not real understanding. We are just following something that we can only *show*, in a sense, as the conductor of the orchestra is showing with its baton, a kind of mixture of performance and unthinkable matter.

This simultaneous experimentability and unthinkability of music was analyzed in detail by Arthur Schopenhauer, in one of the chapters of his seminal book "*The World as Will and Representation*" (Schopenhauer, 1833). He stated that the music, ignores the world of concrete phenomena, and therefore only *resembles* some original reality than cannot be copied. Schopenhauer believes that music resembles, represents, "is a copy of the *will itself*, the objectivity of which are the Ideas." He will underline after some paragraphs his convergence with Leibniz. As stated before, "*Musica est exercitium metaphysics occultum nesciendis se philosophari animi*," that is, music is an unconscious exercise of metaphysics where the mind does not know what is thinking about. Of course Schopenhauer is a vitalist, and biologists will immediately recognize here the relation between *will* and the obscure dynamics of life, and might immediately reach the intuition that the *will* of music might be a copy, a representation of the *will* of life. The effects of music on humans could be derived from the *recognition* (let us accept here this platonic term!) of something common between our obscure perception of life and the music itself. Reinterpreting Schopenhauer, the *will* represented in music is a representation of the *will* of life, that is, a common *will* is independently perceived (can be *shown*) in both the music and life. In principle, we cannot speak about representations between two entities of the same hierarchical order, in our case, two equally unthinkable entities. It could be suggested that, even among unthinkable entities, there may also exist a hierarchy, so that entities that are lower in the hierarchy (music) might be able to represent higher ones (life).

MAJOR TRANSITIONS: A TRANS-HIERARCHICAL CYCLE OF REPRESENTATIONS AND REPRESENTED ENTITIES

Between the groove pattern and the music there is a "major transition," essentially a qualitative transition. Similarly, between the genes and the life of a bacterial cell, or between cells and the complex living expressions of an ecologically-integrated community of cells, there are major transitions, as those identified by Maynard Smith and Szathmàry in evolutionary processes (Maynard Smith and Szathmáry, 1997). In both cases, there is a collection of "small" parts that assemble to produce qualitative different larger wholes. It is obviously tempting to propose that the lower levels of the hierarchy blindly "represent" the higher levels, as the way they are assembled (its order) has something formal, a correspondence, with the higher level activities. Evolutionary biologists will be prone to accept that there are "levels of life" as the development of life seems to occur from single replicating molecules, which provided the bases for reproduction at higher levels, as cells or organisms, in a successive series of "major evolutionary transitions." But it will be difficult to accept that "lower levels" will be able to "represent" the higher levels. That occurs because, if there is an general evolutionary flow from the lower to the higher hierarchical levels, the different entities at the higher levels of life categorically imposes particular organizations of lower levels. Indeed processes as speciation depends on the imperatives of higher over lower hierarchical levels.

The continuous interplay between hierarchical levels is a trademark of life (Campbell, 1974). Some kind of unity based on reciprocal trans-hierarchical effects occurs there between what is represented and the representation itself. Indeed it is easy to imagine that the life of a particular bacterial organism is to a

certain extent represented in the organization of ifs genome, as the music of a particular Mozart's string-quartet is represented in the organization of its musical score. But note that in both cases the representation is a "generational representation" for the represented entity, that is, the re-production of a particular bacterial organism or string quartet is entirely dependent from the representation, the genome, or the score, respectively. On the other hand, the score is meaningless in producing music in the absence of an instrument. In modern times (the origin of life might be another case), the genetic sequences (the representation) are only *meaningful* if a specific living system (what is represented) is present. The meaning of the representation should be perceived by the living system, which produces a biological scaffold (an instrument) at its turn generates its own representation, what is required for reproduction. The higher hierarchical level (the living system) has the lead in the process, as there is no representation without anything to be represented. That fits with the common wisdom in biology: the content and order of sequences in the genome corresponds to what has been selected by life itself, the complex and dynamic living interplay between the cell and its environment. For instance, speciation requires the dominance of "what it is represented," the adapted phenotype, over a particular genome. Essentially genetic plasticity and modularity expresses such subordination. We found here an interesting trans-hierarchical cyclic correlation between representation and reproduction, a correlation that is produced blindly in both senses, as in the metaphor of the messages crossing a Chinese room (see above). Indeed we suspect that in ancient evolutionary times there was no difference between what was represented and the representation itself, the evolution of life consisting in digging a "major hierarchical transition" between both entities.

How that model applies for the music score? We should admit that the score, being the representation of music, produces music only if there is an instrument that converts notes into sounds, and only if there is somebody with an "experiment, in a sense, to understand (even in an unthinkable way) the *meaning* of the music. Because of its experiential capacity, and the derived effects of such kind of understanding, the human being listening music (higher in the hierarchy) is able to produce suitable instruments for reproducing again something *from himself*, perhaps the obscure *will* of Schopenhauer. The metaphor resists: the music score represents something unthinkable, this will, and the subject of the will (life in general, or a human mind) produces representations to perpetuate the re-presence of the effects associated with the will.

THINKING ON REPRESENTATIONS

The major task of science is to explore and expand the limits of intelligibility. The main current task of biology is to understand the correspondences between genomes and life. (Lewontin, 1974; Ferrada and Wagner, 2012; Wagner, 2012a,b). A representation is based on *correspondences* between what is represented and the representation itself. The representation cannot be at the same hierarchical level than that what is being represented. What might be the correspondences between life or music, and the artifactual representation of these presumed realities, as genomic sequences or musical scores? Let us go back to the *Tractatus*. "It is clear that, however, different from the real one an imagined world may be, it must have something—a form—in common with the real world" (2.022). What it is common are the *forms*. "We make (our logic, our language) to ourselves pictures of facts" (2.1). This picture is a representation of the facts, in which the elements of the picture correspond to the objects (2.13), linked in a definite way accordingly to what it is imposed by their forms (2.14), and logically indicates the possible non-existence of some facts (2.11). The picture, the representation of the reality, is in itself a fact (2.141), "as the elements of the picture are combined with another in a definite way, representing that real things are so combined with another" (2.15; 2.15.14). These *co-ordinations* are as it were the feelers of its elements with which the picture touches reality (2.1515). "What the picture must have in common with reality in order to be able to represent it after is manner—rightly or falsely—is its form of representation" (2.17). In summary, "the picture has the logical form of representation in common what it pictures" (2.2). All that certainly has a certain platonic flavor, as the picture, the representation, can be something made by drawing the shadow of the reality on our mind's screen. Biology is the art and science of finding the *forms* able to *represent* life.

As any other type of knowledge, biology should be based on propositions. Only propositions have sense; only in the context of a proposition has a name meaning (3.3). Propositions do not have meaning, but sense, reflecting all possible situations they represent. Only the relations, the order matter, not the things, not the objects themselves. A proposition is the description of a fact (4.023), a fact that involves objects. The possibility of representation of the reality (the task of sciences) is based, accordingly to Wittgenstein, in the theorem (1.1) of the *Tractatus*: the world is constituted by the totality of facts, not of things (Theorem 1.1). The fact, if it is the case, exist as an elementary (atomic) fact (1.21-2), resulting from a (minimal) particular combination of objects (2.01). Objects are simple (2.02), elementary, fixed (2.026) entities. But the *objects by themselves, outside facts, are only possibilities of facts*, nothing in reality (2.011) except their forms, qualities to be part of facts (2.0141). These qualities determine the possibility of facts (2.012): "objects contain the possibility of all states of affairs" (2.014). The object is the fixed, the existent; the configuration is the changing, the variable (2.0271). In the elementary fact objects hang one in another, like the links of a chain (2.03), combined in a definite way (2.031). The way in which objects hang together in the fact is the structure of the fact (2.032). And finally: the form (of objects) is the possibility of the structure (2.033).

Examining these Wittgenstein's theorems, a microbiologist will immediately be attracted and even moved by the idea that the philosopher is speaking about life, with all its unveiled evolutionary possibilities, based on alternative molecular configurations that give rise to different facts, a game in which molecules themselves are nothing: the *objects by themselves, outside facts, are only possibilities of facts*. If Biology is the art and science of *representing* life, such representation should be sufficiently faithful to reflect the complexity and the dynamics of facts in the unthinkable real life, and, understanding the links in the representation (the model) we should assume that something "similar" should occur, at least in part, in the true life.

THE EFFECTS OF REPRESENTATIONS FROM ZEUXIS TO SCHUBERT

In the fifth century BC, Zeuxis depicted the grapes so realistically that birds flew down to peck them: *Ars simiae Naturae*. In this example, grapes are considered by birds to be alive, so that the representation not only faithfully corresponds to life, but produces the *same effects*. Modifying in the model (the picture) the shape or color of the grapes, the birds will not be attracted anymore, so that we could presume which is the attractive properties of real grapes. Of course the result of the experiment can be wrong. For instance, birds could be attracted in the picture by the odor of oil or egg used as a solvent of a particular dry used to paint the grapes, and not by its realistic color. But the experiment might also be true, and the birds could effectively be attracted by the painted grapes. The representation has an *effect*, which might be similar, or even identical, to the effect caused by what is represented. Now let us birds to examine for a little longer the famous Zeuxis's grapes. Certainly they will be soon disappointed, and if challenged again by the image, they will not be attracted any more. As in the famous Türing metaphor, life will recognize life, provided a certain period of examination.

The idea of "music of life" has been developed recently by the famous physiologist Dennis Noble (Noble, 2006). The authors of the present essay were simultaneously disappointed and flattered when he found that Noble used in his book "The Music of Life" almost exactly the same example that we also used in the first version of our manuscript, produced years ago, namely the Schubert's Piano Trio in E flat major, D.929; the author's choice was its ancestor, the Piano Trio in B flat, D.898. Noble even considered the space travelers metaphor, even though not entirely in our way. The important thing is that all of us were *impressed* emotionally by that piece of music. "As the music entered into the slow movement," "I cried" confess Noble. Beyond any possible doubt, music produces *experiences* and *effects*. If music were an unthinkable representation of life, we are experiencing effects because this representation provides an obscure perception of the *will* of life, in the Schopenhauer sense. But note that if this perception will be very difficult or impossible to obtain just looking at the score, or the irregularities of the grooves of a gramophone record.

HALF-THINKING, HALF-SEEING

Ludwig Wittgenstein evolved in his posthumously published book "Philosophical Investigations" (Wittgenstein, 1953; McGuinn, 1997) to the proposal of some possible ways for understanding the realities that can be *shown*, but not *thought*, or at least, not entirely thought. The experience of seeing something is converted in a perceptual experience, and thus regarded as indistinguishable from a thought (Bar-Elli, 2006). As Wittgenstein says, "-is it a case of seeing and thinking? Or an amalgam of the two, as I should almost like to say?" (PI 197). Let us imagine a complex feature that can be only shown, as a human face. The image of the face depends of a huge network of anatomical interactions involving the shape of bones, the volume of the muscles, the amount and distribution of subcutaneous fat, and many other factors. But also these features reflect the age, the sex, ethnicity, the diet, or even the character of the underlying human being, so that the looking at this face might produce *effect*s. A full description of the dynamic network of elements giving rise to a particular recognizable face will be almost impossible. Of course, rough approximations might be attempted, as anthropologists trying to reconstruct from the bones and the presumed diet the face of Lucy, our hominid ancestor. But the example that we are discussing here is that the aspect of the face that we *see* is "condensed information" of a complex network of elementary facts. Even more, without need of knowing almost anything of the generational network of interactions giving rise to the face, the face can be remembered and *compared* with other faces at an extremely specific level of discrimination. The person recognizing a face has an "experience of meaning." As the face might reveal family resemblances, Wittgenstein suggests a kinship between *seeing* an aspect and the *experience of meaning*. The experience of meaning is half-thinking (Bar-Elli, 2006).

Of course the musical metaphor is exploited by the later Wittgenstein of "Last writings on the Philosophy of Psychology" in an identical sense (Wittgenstein, 1949–1951; Worth, 1997). Music results from a complex dynamic interplay of elements, has not a clear semantic structure and, as the face, produces effects, an *experience of meaning* that enable *connections* and *comparisons*. For instance, he says about a musical theme: "I could compare it with something else which has the same rhythm (I mean *the same pattern*)" (I.382). The understanding of a musical theme is based on the experience of what he defines as "internal relations" occurring in the otherwise "unthinkable" musical stuff. Bar-Elli has pointed out the critical importance of the concept of the *experience of meaning* in Wittgenstein as a part of a synoptic view (*übersicht*) of understanding (Bar-Elli, 2006). Music is an excellent equivalent of life in terms of exploring intelligibility of complex systems, probably superior to language, precisely because we seem to lack here any grip on an idea of semantic units, which is so often conceived as the basis of linguistic analogy. As the book of music cannot be reduced to the music score, and then music is music, "my central argument is that the book of life is life itself" says Dennis Noble (Noble, 2006). In conclusion, our understanding of complex systems as life or music depends *both* on: (1) the understanding of their representations (as genomes or scores), and (2) the understanding (under the form of experience of meaning) of something that can be only seen or show, but still compared or connected. Half-thinking, half-seeing: the HT-HS strategy.

IMPLEMENTING THE HT-HS STRATEGY: COMPLEX MODELS OF COMPLEX SYSTEMS

How the HT-HS, half-thinking, half-seeing strategy could be applied to increase the intelligibility of complex biological systems, and life in general? For the half-thinking part, it is obvious that we should maintain a high-level descriptive research as it is being done in genomics, proteomics, metabolomics, or transcriptomics of particular organisms, complemented by the dynamic approach provided by fluxomics, all within the frame of more and more computationally advanced systems biology. All that is research on the composition of Wittgensteinian *atomic facts* and *propositions*; of course that includes certain level of synthesis, what one of us (Moya et al., 2009) proposed to call *synthetic view one*. This level is the level of anatomy and physiology, or, in

linguistic terms, *legibility*. All these approaches essentially would serve to provide material (organized material) to feed complex models, able to move into the *synthetic view two* (Moya et al., 2009). Possibly the advances in Systems Biology will provide useful integrative models, but they will be insufficient to provide by themselves full intelligibility of functional processes. Brenner contend that this approach is insufficient, as deducing models of function from the behavior of a complex system is an inverse problem that is impossible to solve (Tarantola, 2006; Brenner, 2010).

The half-seeing part might start when we could be able of developing more powerful model tools to run simultaneously all data and processes generated by these—omics, in a comprehensive and integrated way. For that a purpose we need "Big Science," based on the convergent interactions among scientists of many disciplines, and not only from biology (Nurse, 2008). If we were able to represent the holistic result of such synthetic approach, we will be near something as a complex image of a living structure, able to be seen or to be showed. And more importantly, able to be "physiognomically" compared and related with other images obtained from other organisms. At this stage we should reach *imageability*. The more advanced part of the half-seeing part should be based on multi-hierarchical understanding of life, and the expected appearance of *emergent* qualities, particularly if *communication* strategies between levels are assured.

Modeling trans-hierarchical complex levels is certainly one of the biggest challenges we have (Campbell, 1974; Martínez and Moya, 2011). These models would eventually provide different levels of predictability. Indeed predictability is the best touchstone to validate the reality of complex models for complex systems (Martínez et al., 2007). Maybe even complex models might be able to predict just the next steps of biological processes, and only in close space and time compartments, just as meteorological predictions based on cumulative empirical observations. Also, and as we mentioned before (at least for non-atonal music) the experience allows to predict the evolution of a melody, at least for a few compasses. In any case, complex models will serve to continuously provide material to be tested and rejected when not mating with reality (Brenner, 2010). We should in any case be aware that to validate models by comparing them with "reality" might be a circular problem, as the "reality" could be only defined by models. That is why we should be able to "see" the reality, even in an obscure, fuzzy way. Paraphrasing Albert Einstein (Einstein, 1944; Dougherty and Bittner, 2011) the propositions, the rationality of the models (half-thinking) should be "firmly connected with sensory experiences" (half-seeing).

As Denis Noble says, life should be considered in a variety of levels; life is "a kind of music, a symphonic interplay between genes, cells, organs, body, and environment," what can be only examined under the views of synthetic biology (Noble, 2006). Microbiologists are among the best placed scientists to mature these concepts, as they have daily experience of the complex interplay of genetic sequences and domains, operons, genes, proteins, macromolecular complexes, signaling networks, adaptive and regulatory functions, different classes of nested mobile genetic elements, clones, species, communities, integrated microbiotic ensembles, and microbial ecology at large. We have to deal with huge diversity of *facts* or *pieces*, constantly offering in a trans-hierarchical way new complex patterns to evolutionary processes (Baquero, 2004, 2009, 2011; Wagner, 2012a,b). In short, we hope that the future will allow scientists to cover the three phases of this epistemological process will be: legibility, imageability, and intelligibility of complex biological systems. Mixing half-thinking and half-seeing, the scientific method that might be we should apply to understand the complexities of microbial life.

ACKNOWLEDGMENTS

Work in the laboratory of Fernando Baquero is funded by research grants from the European Commission (PAR-241476-FP-HEALTH-2009 and EvoTAR-FP-HEALTH-F3-2011-282004), the CIBERESP Research Network in Epidemiology and Public Health (CB06/02/0053), and the Regional Government of Madrid (PROMPT-CM-CM S2010/BMD-2414), and Andrés Moya by the grants SAF2009-13032-C02-01 of Ministry of Economy and Innovation and Prometeo /2009/092 of the Regional Government of Valencia, Spain.

REFERENCES

Babu, M., Musso, G., Díaz-Mejía, J. J., Butland, G., Greenblatt, J. F., and Emili, A. (2009). Systems-level approaches for identifying and analyzing genetic interaction networks in *Escherichia coli* and extensions to other prokaryotes. *Mol. Biosyst.* 5, 1439–1455.

Bar-Elli, G. (2006). Wittgenstein on the experience of meaning and the meaning of music. *Philos. Invest.* 29, 217–249.

Baquero, F. (2004). From pieces to patterns: evolutionary engineering in bacterial pathogens. *Nat. Rev. Microbiol.* 2, 510–518.

Baquero, F. (2009). Predictions: evolutionary trajectories and planet medicine. *Microb. Biotechnol.* 2, 130–132.

Baquero, F. (2011). The Garrod Lecture. The dimensions of evolution in antibiotic resistance: ex unibus plurum et ex pluribus unum. *J. Antimicrob. Chemother.* 66, 1659–1672.

Bedau, M. A. (1996). "The nature of life," in *The Philosophy of Artificial Life,* ed Margaret Boden, (Oxford: Oxford University Press), 332–357.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14332–14337.

Brenner, S. (1999). Theoretical biology in the third millennium. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354, 1963–1965.

Brenner, S. (2010). Sequences and consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 207–212.

Callebaut, W., and Laubichler, M. D. (2007). Biological complexity as a challenge for biological theory. *Biol. Theory* 2, 1–2.

Campbell, D. T. (1974). "Downward causation in hierarchically organized biological systems," in *Studies in the Philosophy of Biology: Reduction and Related Problems,* eds F. Ayala and T. Dobzhansky (London: MacMillan), 179–186.

De Man, P. (1978). *The Epistemology of Metaphor. Critical Inquiry*, Vol. 5, No. 1, Special Issue on Metaphor (Autumn, 1978), 13–30.

Desnues, B., Al Moussawi, K., and Raoult, D. (2010). Defining causality in emerging agents of acute bacterial diarrheas: a step beyond the Koch's postulates. *Future Microbiol.* 12, 1787–1797.

Dougherty, E. R., and Bittner, M. L. (2010). Causality, randomness, intelligibility, and the epistemology of the cell. *Curr. Genomics* 11, 221–237.

Dougherty, E. R., and Bittner, M. L. (2011). "Epistemology of the cell: a systems perspective on biological knowledge," Series: IEEE Press Series on Biomedical Engineering, Institute of Electrical and Electronics Engineers, (Hoboken, NJ: John Wiley and Sons, Inc).

Einstein, A. (1944). "From the philosophy of Bertrand Russell," in *The Library of Living Philosophers,* Vol. 5, ed P. A. Schilpp (Greensboro: Tudor Publishers), 277–294.

Einstein, A. (1949). "Remarks concerning the essays brought together

in this co-operative volume," in *Albert Einstein: Philosopher-Scientist Library of Living Philosophers,* Vol. 7, ed P. A. Schilpppp (Evanston, IL), 683–684.

Ferrada, E., and Wagner, A. (2012). A comparison of genotype-phenotype maps for RNA and proteins. *Biophys. J.* 102, 1916–1925.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.

Human Genome Sequencing Consortium International. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Komarova, N., and Niyogi, P. (2004). Optimizing the mutual intelligibility of linguistic agents in a shared world. *Artif. Intell.* 154, 1–42.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York, NY: Columbia University Press.

Martínez, J. L., Baquero, F., and Andersson, D. I. (2007). Predicting antibiotic resistance. *Nat. Rev. Microbiol.* 5, 958–965.

Martínez, M., and Moya, A. (2011). Natural selection and multi-level causation. *Philos. Theor. Biol.* 3, e202.

Maynard Smith, J., and Szathmáry, E. (1997). *The Major Transitions in Evolution*. Oxford: Oxford University Press.

McGuinn, M. (1997). *Routledge Philosophy Guidebook to Wittgenstein and the Philosophical Investigations*. London: Routledge.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A., Holden, J. L., McCann, O. T., Dunham, A., Soderlund, C., Scott, C. E., Bentley, D. R., Schuler, G., Chen, H. C., Jang, W., Green, E. D., Idol, J. R., Maduro, V. V., Montgomery, K. T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J. H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P. J., Catanese, J. J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V. G., Kirsch, I. R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J. F., Hawkins, T., Myers, R. M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N. E., Cox, D. R., Haussler, D., Kent, W. J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H. S., Sakaki, Y., Shimizu, N., Asakawa, S., Kawasaki, K., Sasaki, T., Shintani, A., Shimizu, A., Shibuya, K., Kudoh, J., Minoshima, S., Ramser, J., Seranski, P., Hoff, C., Poustka, A., Reinhardt, R., Lehrach, H., and International Human Genome Mapping Consortium. (2001). A physical map of the human genome. *Nature* 409, 934–941.

Moya, A., Krasnogor, N., Peretó, J., and Latorre, A. (2009). Synthetic biology: Goethe's dream. Challenges and opportunities for synthetic biology. *EMBO Rep.* 10, S28–S32.

Moya, A., Cantón, R., and Raoult, D. (eds.). (2012). Microbiome: deciphering the last human body organ. *Clin. Microbiol. Infect.* 18, 1–73.

Noble, D. (2006). *The Music of Life: Biology Beyond the Genome*. Oxford: Oxford University Press.

Nurse, P. (2008). Life, logic and information. *Nature* 454, 424–426.

Rosen, R. (2004). *Life Itself. A comprehensive inquiry into the nature, origin, and fabrication of life,* 2nd Edn. (New York, NY: Columbia University Press).

Schlick, M. (1936). Meaning and verification. *Philos. Rev.* 45, 339–369.

Schopenhauer, A. (1833). "The music in the hierarchy of the arts," in *The World as Will and Representation*. I:52. Trans. A. Payne and E. F. J. Payne, (Indian Hills: Falcon's Wing Press), (1958).

Schrodinger, E. (1957). *Science Theory and Man*. New York, NY: Dover.

Searle, J. R. (1984). *Minds, Brains, and Science*. (Cambridge, MA: Harvard University Press.) 13th Printing, 2003.

Steels, L. (2004). "Analogies between genome and language evolution," in *Artificial Life IX,* ed J. Pollack (Cambridge, MA: Harvard University Press), 200–206.

Steels, L. (2010). Can evolutionary linguistics become a science? *J. Evol. Linguist.* 1, 1–34.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nat. Phys.* 2, 482–484.

Wagner, A. (2012a). The role of robustness in phenotypic adaptation and innovation. *Proc. R. Soc. B Biol. Sci.* 279, 1249–1258.

Wagner, A. (2012b). The role of randomness in Darwinian evolution. *Philos. Sci.* 79, 95–119.

Westerhoff, H. V., and Palsson, B. O. (2004). The evolution of molecular biology into systems biology. *Nat. Biotechnol.* 22, 1249–1252.

Wimsatt, W. C. (1976). "Reductionism, levels of organization and the mind-body problem," in *Consciousness and the Brain: A Scientific and Philosophical Inquiry,* eds G. G. Globus, G. Maxwell and I. Savodnik (New York, NY: Plenum Press), 199–267.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus. Versions*. Trans. C. G. Luckhard and M. A. E. Aue (Basil Blackwell), 1980. http://www.kfs.org/∼jonathan/witt/mapen.html

Wittgenstein, L. (1949–1951). *Last Writings on the Philosophy of Psychology*. Trans. C. G. Luckhard and M. A. E. Aue (Basil Blackwell), 1982.

Wittgenstein, L. (1953). *Philosophical Investigations*. Trans. G. E. M. Anscombe (Basil Blackwell), 1953.

Worth, S. E. (1997). Wittgenstein musical understanding. *Br. J. Aesthetics* 37, 158–167.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 April 2012; paper pending published: 16 May 2012; accepted: 13 June 2012; published online: 12 July 2012.

Citation: Baquero F and Moya A (2012) Intelligibility in microbial complex systems: Wittgenstein and the score of life. Front. Cell. Inf. Microbio. ***2****:88. doi: 10.3389/fcimb.2012.00088*

frontiers in
CELLULAR AND INFECTION MICROBIOLOGY

REVIEW ARTICLE
published: 08 August 2012
doi: 10.3389/fcimb.2012.00106

Darwin's goldmine is still open: variation and selection run the world

Patrick Forterre [1,2]*

[1] *Institut Pasteur, Paris, France*
[2] *Institut de Génétique et Microbiologie, Université Paris-Sud, CNRS UMR8621, Orsay Cedex, France*

Edited by:
Eugene V. Koonin, National Institutes of Health, USA

Reviewed by:
David O'Callaghan, INSERM, France
Abdu F. Azad, University of Maryland, USA

***Correspondence:**
Patrick Forterre, Institut Pasteur, 25 rue du Dr Roux, 75015, Paris, France.
e-mail: forterre@pasteur.fr
Institut de Génétique et Microbiologie, Bat409, Université Paris-Sud, Centre d'Orsay, 91405 Orsay Cedex, France.
e-mail: patrick.forterre@igmors.u-psud.fr

The scientific contribution of Darwin, still agonized in many religious circles, has now been recognized and celebrated by scientists from various disciplines. However, in recent years, several evolutionists have criticized Darwin as outdated, arguing that "Darwinism," assimilated to the "tree of life," cannot explain microbial evolution, or else was not operating in early life evolution. These critics either confuse "Darwinism" and old versions of "neo-Darwinism" or misunderstand the role of gene transfers in evolution. The core of Darwin explanation of evolution (variation/selection) remains necessary and sufficient to decipher the history of life. The enormous diversity of mechanisms underlying variations has been successfully interpreted by evolutionists in this framework and has considerably enriched the corpus of evolutionary biology without the necessity to kill the father. However, it remains for evolutionists to acknowledge interactions between cells and viruses (unknown for Darwin) as a major driving force in life evolution.

Keywords: evolutionary synthesis, variation, natural selection, lateral gene transfer, Darwinian threshold, viruses

INTRODUCTION

Darwin had to be defended in the XIX century against those who wished to maintain the concept of our innate highness: the universe had been created for us. The sin of Darwin was to force us to consider ourselves as a "normal" part of the biosphere, and worst, of the animal kingdom. Although Darwin himself, still influenced by biblical thinking, sometimes viewed human beings as the best product of evolution and "natural selection" as a kind of cosmic force (Richards, 2009), his ideas were "dangerous," because of the powerful explanatory power of the dyad variation plus natural selection and because a descend with modification does not imply progressive evolution (Gould, 1996). Darwin's ideas thus ruined the creationist credo, opening the living world to scientific exploration in the framework of a materialist agenda. Darwin's dangerous idea was defended and he was finally recognized and celebrated in the scientific community. In the last century, the original ideas of Darwin were completed (sometimes corrected) by the development of genetics (the *evolutionary synthesis* or *modern synthesis*) and later on by molecular biology. Recently, Addy Pross proposed a general theory of evolution, extending Darwin's theory to inanimate matter (Pross, 2011). He discusses how Darwin's principles can be deduced from more fundamental chemical principles that govern the evolution of complex chemical systems through imperfect replication and kinetic selection.

Darwin's ideas seem therefore to be alive and well. However, in recent years, whereas still being a devil for religious fanatics, Darwin became the target of heavy criticisms coming from part of the scientific community itself. Several genomists and molecular evolutionists have argued that genomic data have challenged Darwin's view of life (Bapteste et al., 2009; Dagan and Martin, 2009; Doolittle, 2009; Koonin, 2009a,b; Raoult, 2010). Notably, they have suggested that "Darwinism" is only valid for eukaryotes not for prokaryotes (assimilated to microbes) and proposed to replace the "Tree of life" (TOL, supposed to be the hallmark of Darwinism, but see Penny, 2011) by networks (or rhizome) to take into account gene flows between organisms (Bapteste et al., 2009; Dagan and Martin, 2009; Raoult, 2010). In a review title, Ford Doolittle wondered what "*the demise of Charles Darwin's tree of life hypothesis means for classification and the theory of evolution*" (Doolittle, 2009). Lamarck has been (once more) awoken to confront Darwin, as illustrated for instance by the title of another recent review paper: "*Is evolution Darwinian or/and Lamarckian?*" (Koonin and Wolf, 2009).

Carl Woese himself, one of the greatest biologists of the last century, has suggested replacing "Darwininan" evolution (driven by competition between individuals) by communal evolution (driven by exchange of experiences between individuals, via lateral gene transfers) for the early steps of life history, i.e., from the origin of life up to the formation of the three modern cellular domains (Archaea, Bacteria, and Eukarya) (Woese, 2002). He wrote: "*the time has come for biology to go beyond the doctrine of common descent*" and proposed the term "*Darwininan threshold*" to name the transition between communal and "Darwininan" evolution (Woese, 2002). Finally, gradualism and uniformitarianism, considered to be essential pillars of Darwin's view of life, are also (again) strongly attacked (Koonin, 2009a,b).

These views are the bedrock of this special issue with its provocative title "*Microbial genomics challenges Darwin.*" This is a sensitive topic, considering the renewal of creationist thinking in fundamentalist religious circles and the wide publicity given to these "non-orthodox views." This was best illustrated by the cover of the "*New Scientists*" issue published in January 2009 showing a tree of life superimposed with the sentence: "*Darwin was wrong.*" Although the existence of anti-Darwinists in the political arena is certainly not a reason to hide fierce debates between evolutionists over mechanisms and representations of evolution, one can regret to see the name of Darwin used as a foil in these debates. After all, nobody said: "*Mendel was wrong*" because his concept of the gene was quite different from what we know today (personal quote from Eduardo Rocha).

DARWIN AND/OR DARWINISM

The debate around Darwin and Darwinism is important for the future of our discipline since, as pointed out a few years ago by Bos (1999): "*progress in science is not only a matter of mere technology but of philosophy as well,*" "*progress therefore is reflected in terminology and in the definition of terms.*" In our case, the definition of terms has always been a complex and evolving story. Darwin was not Darwinist. Indeed, although we are biologists, there is no such thing as "biologism." Scientists are not born to produce doctrines but rational explanations supported by experiments (when possible) and open to criticisms, refutation, and/or modifications. To be consistent with this view, this assay will not be a defense of Darwinism, but of Darwin's core ideas, the couple variation/selection, because Darwinism, as a doctrine, evolved into many different ways, and it is all too easy to select or forge one of them, either to prove it right or wrong. For a recent comprehensive presentation of Darwin's conceptions (beyond the core ideas discussed here), I refer the reader to a recent review by David Penny (2010).

My aim of course is not to argue that Darwin himself was always right, or that we should come back to Darwin's initial views, since, living two centuries ago, he was by necessity ignorant of today biology and he was thinking in another intellectual framework (Richards, 2009). Darwin, originally born Christian and once student in theology, progressively changed his own credo in being confronted to geological and biological facts that could not be explained by the creation theory. Darwin thus finally adopted a materialistic view of the world, putting back human beings into Mother Nature. However, he was still influenced by the *Scala natura* concept of Aristotle (as many modern biologists still are) and his theory initially preserved nature's moral purpose (Richards, 2009). In fact, he thought that "*man is the one great object of nature*" (Darwin, 1987). However, whatsoever Darwin's limitations, I will argue that we have still more to gain standing on his shoulders than tripping him, especially when he cannot reply.

SELECTION, YES, BUT VARIATIONS FIRST

It is well known that Darwin was not the first to introduce the idea of evolution in biology (beside Lamarck, one of his predecessors was his own grandfather Erasmus) but it's Darwin and Wallace, who were the first to propose a mechanism for the origin of new species: variation followed by natural selection, leading (in Darwin's term) to: "*descend with modifications*," an expression much more important for Darwin himself than its tree depiction. Although selection does not make sense without variations, "Darwinism" is often reduced to "natural selection" (struggle for life, survival of the fittest) without reference to his emphasis on variation. This is of course because the nature of biological variations remained a complete mystery for Darwin and his contemporaries. In contrast, much was already known on the efficiency of processes such as artificial selection in agriculture and "breeding" and this was determinant for Darwin to formulate his ideas.

Importantly, focusing on natural selection helped the development of evolution as a new branch of biological sciences, the mechanism of natural selection being open to experiments *in situ* (in the fields) as well as in the laboratories, so that evolution became part of the mainstream biological research agenda. From the focus on selection emerged terms such as fitness, genetic drift, the introduction of statistics to "measure" evolution (making it a "true" science) and the creation of new disciplines, such as population genetics.

However, the contribution of Darwin cannot be rightly summarized by natural selection. Darwin also realized the importance of variations, as a prerequisite for evolution. The chapter 2 in "*On the origin of species*" entirely deals with variation, discussing varieties and sub varieties within species, whereas natural selection is discussed in chapters 3 and 4. This is not trivial. At the time of Darwin, biologists were still strongly influenced by current philosophical theories that focus on the essence of things (reminiscent of Plato's ideas). When considering a particular "species," zoologists or botanists were not fascinated, but rather annoyed, by the diversity of the individual members of this specie. They were looking for the ideal "type species" to describe species without having to mention all possible varieties. For religious scientists, they probably hoped in this way to reconstruct the first member of this species, the one directly created by god (varieties being less perfect by-products).

The great merit of Darwin was to change this perspective upside down. Instead to be confused by the diversity within species, he realized that this diversity is the essence of life, variations providing substrates for selection. As pointed out by Brüssow (2009) "*diversity is not an evolutionary accident, but the organizing principle in biology, without which evolution would not occur.*" The four years of Darwin exploration with the Beagle, far from academic life, were certainly critical in opening his eyes on this issue. In fact, there is no such thing as a species in the real world, except as concepts in our mind (and in books) but organisms and populations. There are myriads of individuals that are all different, even between members of the same "species" defined by any criteria. Darwin was the first to realize that this diversity was the key parameter allowing selection.

The historical focus of most evolutionists on selection, instead of variations, produces some confusion on the nature of selection. Darwin himself used to think of selection somehow as a kind of metaphysical force (Richards, 2009), and similarly, some evolutionists used to consider natural selection as the cause of evolution. As a consequence, each time a new mechanism of variation or any constraint in the mode of existence of organisms

is discovered, it is claimed that natural selection has been weakened (see for instance Table 1 in Pigliucci, 2009 in which natural selection is supposed to be *altered* or its *efficacy decreased* by phenomena such as *contingency, biological emergence or phenotypic plasticity in macroevolution*). However, in my opinion, this is quite misleading. Natural selection is not an "evolutionary force" but the necessary outcome of variation and multiplication. In particular, natural selection cannot be weakened by mechanisms that promote variations (such as epigenetic mechanisms or symbiogenesis), because these processes provide more substrates for selection.

Importantly, Darwin realized that natural selection is the inevitable consequence of the extraordinary multiplication power of living organisms. In that sense, microbial evolution does not challenge but vindicates Darwin, since the multiplication power of life is higher by several orders of magnitude in the microbial (and viral) world, making natural selection even more drastic in these realms. Despite the limited knowledge of his time, Darwin himself was in fact the first to consider that microbial evolution also involves natural selection (O'Malley, 2009).

Originally, Darwin mainly (but not exclusively) used to consider what we call now positive selection, we known today that variants can be also selected by chance (genetic drift) or strongly counter selected if they do not fit the basic life requirement of the organism (purifying or negative selection). This has been clearly observed by molecular biologists at the sequence level, as in the case of neutral evolution (Kimura, 1977). It is important to insist once more that any type of selection only makes sense because of variation. If conditions change, the successful variants will not be the same, but in any case, selection will operate as soon as variation exists.

THE NATURE OF VARIATIONS

The chapter five of "*On the origin of species*" is entirely devoted to the nature of variation (a tour de force, considering the state of the art in biology at his time). In contrast to a widely held assumption, Darwin did not think that variations occurred mainly by random processes (although he recognizes the existence of random variation). He was not opposed to the "inheritance of acquired characters" and agreed with the idea that "use or disuse" of a character led to its progressive gain (fixation) or loss. Two notions that are today associated to "Lamarckism" (see below). His most original idea was that important variations were slight changes induced (mysteriously) by the environment in the reproduction apparatus. Darwin was therefore "Lamarckian," although his focus on the reproduction apparatus can be interpreted as a premonition of the distinction made later on by Weissman between the soma and the germen. The traditional opposition between "Darwinism" and "Lamarckism," based on the idea that Darwin would favor random variations whereas Lamarck favored the inheritance of acquired characters is clearly wrong. The major difference between Lamarck and Darwin is that, for Lamarck, evolution came from concerted modifications triggered by an internal "vitalistic" forces (*le pouvoir de vie*), so that all individuals in the population experience similar changes to become more adapted to their environment. In that case, natural selection has no more *raison d'être*. For Darwin, acquired characters (even if triggered by the environment) were, first of all, individual acquisitions that should have survived the screen of selection.

The identification of DNA as the carrier of genetic information in the middle of the last century was a decisive blow for neo-Lamarckism. It was difficult to imagine how environmental changes could modify on purpose the sequence of DNA. Early molecular biologists assimilated variations to random mutations and assumed that environmental modifications cannot produce oriented-mutations. For them, selection (instead of variation) became the *Deus ex machina* who sorts out from the chaos of random mutations those making sense for the organism. The book "*Chance and Necessity*" by Jacques Monod perfectly illustrates the best achievement of this thinking (Monod, 1971). In this book, the dyad variation/selection is replaced by the dyad "chance/necessity" which is supposed to be more or less equivalent. It emphasizes that variations were assumed to be entirely the result of random processes, whereas the result of selection provides THE supposedly unique answer "necessary" to make the organism efficient in a given context.

We know today that molecular mechanisms of variations are much more diverse and complex than simple random punctual mutations. Molecular biologists have expended our concept of variations by revealing the importance of epigenetic systems, whereas cellular biologists have continues to reveal the importance of symbioses as a major form of variation. We also know that many answers are possible for a given situation, providing a much more complex history for life, introducing chance in the process (contingency). But none of these considerations challenge Darwin himself, even if they challenge successive historically dated versions of "Darwinism."

The great achievement of molecular biology has been to answer (still partially) to one of the most important question in biology: what are the mechanisms of variations? All discoveries of molecular biologists have vindicated Darwin by revealing the molecular mechanisms behind the multiplication and variations of living organisms. With genetic engineering, molecular biologists have finally got the possibility to produce by themselves artificial variations in the genetic material of organisms, making evolution a fully experimental discipline.

THE FALSE COME BACK OF LAMARCK

Molecular biologists are now out of fashion and spotlights focus on genomists and synthetic biologists. Possibly because Darwin's contribution was clearly recognized by the pioneers of molecular biology (now often accused of reductionism) it seems that genomists and some modern evolutionists look for another hero apparently fitting better with "holistic views" and "systemic biology." Lamarck (more precisely neo-Lamarckism) is again recruited in this crusade. Recently, when a novel mechanism of genome variation apparently triggered by the environment is discovered, it is often claimed that Lamarck was right and Darwin wrong. For example, it has been recently argued that the discovery of the clustered regularly interspaced short palindromic repeats (CRISPR) system in Bacteria and Archaea is Lamarckian, because these microbes can acquire in their genomes viral sequences that immunize them against future viral infections (Koonin and Wolf, 2009). This is interpreted as a hereditary

trait acquired from an environmental modification (the presence of a virus) and oriented by this modification (providing future resistance to THIS particular virus). The inserted viral sequence in the CRISPR locus will be maintained by positive selection if the virus is present (use) but will be lost after some time if the virus is no more encountered (disuses).

However, this interpretation is misleading. Indeed, the addition of new sequences to CRISPR loci would have been considered simply as another form of variation by Darwin. The CRISPR system itself emerged and has evolved through random variations and selection and still works that way. When a bacterial population encounters a particular virus, only part of the population is infected (randomly selected among non-lysogenic bacteria with the proper receptor for this virus). Most infected bacteria are killed, whereas some of them (again randomly selected) survive the infection either because of point mutations affecting host virus interactions or because they have successfully activated a CRISPR defense system. At the end, only a handful of survivors would have acquired new CRISPR sequences from the virus. Finally, these new sequences remain present in the genomes of the survivors only if these descendants are selected (*versus* those losing these sequences) by the continuous presence of this particular virus in the environment. The Lamarckian component of the CRISP system appears dependent of our subjectivity. We forget all selection steps that have modeled random variations into a mechanism that seems to have a purpose because it corresponds to an adaptation of Bacteria (or Achaea) to their environment (the presence of viruses).

It is also often argued that horizontal gene transfers (HGT) are Lamarckian because transferred genes are provided by the environment (contact with another organisms) and lead to a better adaptation to this environment (Koonin and Wolf, 2009). However, evolution is not working that way. Let's consider a schematic scenario (too simplistic but just for the demonstration) of adaptation to different temperatures. Our scenario starts with a species of thermophilic organisms living in a hot environment (70°C) whose temperature corresponds to their optimal growth temperature (OGT). In that environment, these thermophiles should coexist with hyperthermophilic (OGT of 80°C) and moderate thermophiles (OGT of 60°C), because their growth curves in function of temperature would overlap. As a consequence, some thermophiles can gain randomly by HGT either advantageous features to live at lower temperature from moderate thermophiles, or advantageous features to live at higher temperature from hyperthermophiles. If the temperature of the environment changes, different members of the thermophilic species will be selected, depending if the climate is cooling or warming. There is nothing "Lamarckian" in this sketch, but it can be interpreted *a posteriori* as such, because it gives the false impression that HGT have facilitated *a priori* the adaptation of former thermophiles to their new biotope. In fact, HGT are no more "Lamarckian" than "Darwinian." For a given organism (either a virus or a cell) any type of HGT is simply a particular type of variation (Forterre, 2011b). The same can be said from the acquisition of an organelle via endosymbiosis, even if long term changes introduced by this particular variation can be tremendous (Maynard-Smith, 1991). On a theoretical ground, this type of variation does not differ from a single point mutation. Darwin would have been delighted to learn of HGT and endosymbiosis as powerful tricks for variation providing adaptation to the environment.

DARWIN AND THE TREE OF LIFE

In recent years, Darwinism has often been associated to the Tree of life (amazingly written in capital letters by its detractors, TOL). For that reason, I will use the abbreviation Tol thereafter (also to prevent any post-biblical interpretation, see Penny, 2011). However, the supposed love affair between Darwin and trees is restricted to the single figure of his book in "*On the origin of species.*" In fact, Darwin preferred the coral metaphor, in order to emphasize the existence of extinct lineages (unsuccessful variants). For Darwin, the usefulness of the tree metaphor was to illustrate the concept of descend with modification. Each node in his tree exhibits multiple small branches symbolizing variations, one of them producing further bifurcations, all others being dead-ends, symbolizing variations that were counter-selected in the evolutionary game. However, all evolutionists are aware that the actual connections are more complex than those depicted by such simple schematic tree. Some lineages can fuse (sexual eukaryotes in fact evolve through successive fusions of individuals) and several robust branches can emerge from a single node. In fact, the same occurs in actual trees in natural forests (with lianas connecting different branches of the same or different trees). The tree metaphor is thus quite good if one refers to real trees and not simplified versions, although simplified versions are still useful to depict graphically the process of evolution on paper or on computer screens!

It is often claimed that HGT fundamentally contradicts the tree concept (Dagan and Martin, 2006; Doolittle and Bapteste, 2007; Doolittle, 2009; Koonin, 2009a,b). Several authors thus have suggested to replace trees by webs (and TOL by WOL!), in which all genes/organisms are connected by links forming networks in the three dimensional space (Halary et al., 2010). It is also argued that organisms cannot be placed on a tree because they are essentially chimera, produced by fusion of different evolutionary lineages, much like a rhizome (Raoult, 2010). However, as already discussed by several authors, these views confuse species and gene (or genome) trees (Galtier and Daubin, 2008; Gribaldo and Brochier, 2009; Valas and Bourne, 2010). For instance Koonin correctly noticed in his paper entitled "*Darwininan evolution in the light of genomics*" that: *The genomes of all life forms are collections of genes with diverse evolutionary histories.*" Then he conclude surprisingly that: "*a corollary of this generalization is that the TOL concept must be substantially revised or abandoned because a single tree topology or even congruent topologies of trees for several highly conserved genes cannot possibly represent the history of all or even the majority of the genes*" (Koonin, 2009a,b). This clearly demonstrates that in this conclusion, Koonin assimilates the TOL to a tree of genes (or genomes). But for most evolutionists, any "trees of life," including the Tol should be trees of organisms (either mono or pluricellular), depicting the history of their relationships, from cell to cell or from individual to individual. The histories of genes, genomes and replicons are fascinating (especially for genomists and molecular biologists) but

they only make sense if we have access to the history of organisms (otherwise, this would mean that we reduce living organisms to their genomes). Fortunately, the history of any gene, with its duplication and HGT, is constrained by the history of organisms, explaining why accurate analyses of gene trees can sometimes reconstitute efficiently trees of organisms.

The cover of the New Scientist, with a tree superimposed by the sentence, "Darwin was wrong," has perfectly illustrated the violence of the attack against Darwin, based on the critic of the tree metaphor. For instance, Dagan and Martin (2006) derided the Tol based on universally conserved proteins as the tree of 1%, because it is based on the analysis of about 1% of genes (around 100) in average bacterial or archaeal genomes (Dagan and Martin, 2006). On the contrary, the fact that this tree, reconstructed from so few data, confirmed the tripartite division of cellular life, originally deduced from rRNA sequences comparison (the tree of a single gene, 0.01%!), is for me a triumph of reductionism in studying the history of life. It illustrates the power of comparative sequence analysis (and "tree-thinking") to reconstruct ancient history (despite all well-known difficulties in resolving ancient nodes, see discussion in Forterre and Philippe, 1999; Forterre, 2010b; Gribaldo et al., 2010).

Interestingly, using holistic approaches, Koonin and colleagues also confirmed the tree of 0.01%, since they conclude that a tree-like signal confirming the tripartite division of life can be recovered from the "*phylogenetic forest*" of gene trees (Puigbò et al., 2009, 2010). Many authors have indeed noticed that phylogenies based on universal proteins and those based on whole genome trees produced more or less congruent global history of life on our planet (at least recovering the three domains). This observation suggests in particular that HGTs have not been so extensive between domains (Wolf et al., 2002; Ciccarelli et al., 2006; Puigbò et al., 2009, 2010). HGT have been indeed extremely rare between domains and between lineages of the same domain for informational proteins such as ribosomal proteins or RNA polymerase subunits (for a case study, see Brochier et al., 2005). Careful phylogenetic analyses of these proteins produce well resolved trees of the archaeal domain that most likely reflect quite accurately the history of this lineage (Brochier et al., 2005; Brochier-Armanet et al., 2011). This means that tentative trees of organisms can be indeed recovered from the forest of gene trees.

The tree metaphor is not only valid for organisms, but it also works for cells, genes and genomes (or more precisely replicons). In other word, the "*net component of prokaryotic evolution*" (Puigbò et al., 2010) is also tree-like. As soon as an object divides by duplication, the history of that object has a tree-like structure. However, there is no reason why trees of organisms and trees of genes, genomes or replicons should be congruent. The **Figure 1** compares a tree of organism (A) and the underlying tree of a particular gene (with one loss and two HGT) (B). Combining the organism and gene trees produces a network (C and D). Importantly, the tree-like structures depicting the history of the organism (A) and the history of the gene (B) is not changed by the HGT (see also Poole, 2009). The structures of organismal trees are not even changed by more drastic variations such as endosymbiosis. The acquisition and enslaving of a cyanobacterium by ancestors of modern plants has not changed the tree-like structure of the eukaryotic domain, viridiplantae emerged as one branch within the eukaryotic tree. Similarly, the endosymbiosis of mitochondria has not changed the tree-like structure of the eukaryotic lineage, defined by the continuity of the cell membrane of the engulfing species (either an archaeon or a proto-eukaryote, depending of your favorite hypothesis, see Gribaldo et al., 2010).

Trees of organisms, genes, and replicons can be of course much more complicated than those depicted by **Figure 1**, to accomodate hybridation or symbiosis (especially in the case of eukaryotic species), or splitting, fusion, or recombination in the case of genes and replicons. These complications make difficult to represent all evolutionary processes by simple tree-like diagrams, except if one focuses on a particular type of biological or molecular entity, but they can always be interpreted as combination of trees. However, combining all organismal, genomic and replicon trees to get an exhaustive view of life history would produce a monstrous network (the WOL!) that would make sense only if we are able to

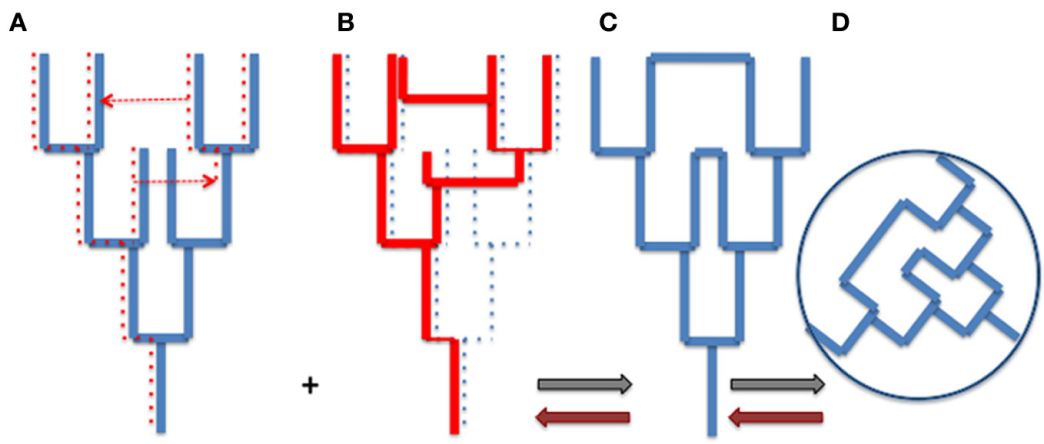


FIGURE 1 | From trees to webs and back. (A) an organism tree corresponding to its vertically inherited genes (blue), **(B)** an underlying gene tree [dotted lines in **(A)**] with one loss and two horizontal gene transfers [indicated by dotted arrows in **(A)**], **(C)** the network obtained by combining **(A)** and **(B)** (gray arrows), **(D)** the web corresponding to the unrooted network. Brown arrows indicate the path unveiling the organism and gene trees from the web.

deconstruct this network to identify the underlying trees and their evolutionary relationships (brown arrows in **Figure 1**).

At smaller scale, a useful example can be provided by the situation of an amoeba hosting endosymbiotic bacteria, which is infected by a mamavirus, itself infected by a sputnik (satellite virus) (Forterre, 2011b). In that case, we have four organisms living in the same cell. These organisms exchange genes and the cell can be viewed as a melting pot of different organisms (a holobiont). However, each organism maintains its individuality during evolution, the Amoeba remains a eukaryote, the bacterium a bacterium, and the two viruses remain viruses. We have four distinct evolutionary lineages (four putative trees) that need to be sorted out from each other to reflect the reality of organismal evolution. The co-evolution of these organisms in the same cells will of course provide the possibility of more variations induced by interactions between these organisms (such as HGT) and these variations will be selected both at the level of individual organisms and at the level of integrated cell.

Doolittle would possibly argue that views defended here are watered down versions of "Darwinism" (Doolittle, 2009). He suggests indeed that Darwin used fact tree metaphor as a major hypothesis to explain "*the hierarchical structure of tree-like classification.*" In such tree like classification, "*the characters which naturalists consider as showing true affinity between two or more species are those which have been inherited from a common parent.*" According to Doolittle, this hypothesis is incompatible with HGT. It is true that homologous characters common to two or more species may have been acquired by HGT (from an ancestral common parent anyway!) confusing the structure of phylogenetic classification if they are not recognized as such. But we also know from Henning that characters should not be used for natural classification based on the evolutionary history if they are not true synapomorphies, because they do not reflect true affinity (Hennig, 1966). The question is to identify HGT to purge phylogenomic data from sequences that were not vertically inherited and/or to use them as synapomorphies to confirm or identify monophyletic groups (Brochier et al., 2005; Huang and Gogarten, 2006).

THE WEB-LIKE STRUCTURE OF MICROBIAL EVOLUTION: AN OLD AND WRONG IDEA IN NEW CLOTHES

One of the greatest achievements of biology in the last century has been the inclusion of microbes in the tree of life, thanks to the rRNA tree and the dramatic discoveries of biochemists and molecular biologists, using *Escherichia coli* and its viruses as model systems. For a long time, it was unclear, both for microbiologists and for evolutionists working on animals and plants, if microbes and macrobes (organisms visible by naked eyes, see Forterre, 2008) could be unified into a single tree (although it was already clear for Darwin that microbes were subjected to variation and selection, see O'Malley, 2009). The discovery, in the middle of the last century, that bacteria and their viruses share with the rest of the biosphere the same macromolecules, the same genetic material, and the same genetic code, was a major breakthrough. However, it was not immediately obvious if microbial evolution could be studied in a meaningful way. The existence of HGT through transformation, transduction, or conjugation was recognized early on by pioneers of molecular biology (well before genomists) and it was widely believed that bacteria should be able to share their genes to such an extend that it was pointless to try reconstructing microbial evolution (for an historical account, see Sapp, 2005). Some authors even argued that bacterial "species" should be considered as "cell types" of a gigantic bacterial organism covering the planet (Sonea, 1971; Sonea and Paniset, 1976).

In one of their two seminal 1977 papers, Woese and co-workers remained us that "*the scattered classification*" of methanogenic bacteria *in the seventh edition of Bergey's Manual* (some being Gram positive, other Gram negative bacteria, some bacilliform, others coccoidal or filamentous) was *"rationalized in terms of reticulate evolution involving an appropriate plasmids*" (i.e., HGT of genes involved in methanogenesis) (Fox et al., 1977). Woese and co-workers however demonstrated experimentally, using the tools of molecular biology, that (1) methanogens form a coherent group of organisms very distant to other prokaryotes known at that time, and (2) they are not Bacteria but Archaea (formerly Archaebacteria) (Woese and Fox, 1977). These landmark papers opened the door to a comprehensive history of microorganisms. More recently, phylogenetic analyses have shown that genes involved in methanogenesis, although "operational," have not even been transferred between different archaea (Bapteste et al., 2005). These analyses have unveiled a complex history, with a single origin for methanogenesis, but several subsequent losses, leading to the formation of two paraphyletic groups of methanogens.

The example of methanogens again shows that it is possible to get meaningful information about the early history of life in our planet, thanks to tree thinking. By refuting the validity of this approach for studying microbial evolution, scientists who called themselves "microbalists" and derided those who refuse to criticize Darwin as "positivists" (Dagan and Martin, 2009) propose in fact to bring us back to the pre-Woesian era, when studying the history of microorganisms was considered to be a futile exercise. Let me consider that the recent proposal of a third major archaeal phylum, the Thaumarchaeota, by "positivist" is another positive outcome of "tree thinking" (Brochier-Armanet et al., 2008). This proposal, now accepted by the community of microbiologists (Spang et al., 2010) has opened new avenues in focusing the attention of evolutionists on this fascinating archaeal phylum. The discovery in this phylum of eukaryotic traits previously unknown in other archaea has raised new questions on their relationship with eukaryotes (Brochier-Armanet et al., 2011, 2012). Such questions can be only addressed meaningfully in a "tree-thinking" framework; they would vanish unfortunately in "web thinking."

Finally, recent post-genomic works on archaeal or bacterial speciation have emphasized that microbial speciation occur by mechanisms very similar to those occurring in macrobes, despite differences in HGT and sex (Cadillo-Quiroz et al., 2012; Shapiro et al., 2012). Amazingly, whereas Shapiro and co-workers illustrate their experimental work by a simple bifurcation tree showing the divergence of two populations (with progressive reduction of HGT between them), Papke and Gogarten (2012), in an accompanying comment, illustrate the same story by a reticulate

diagram (Web thinking) to emphasize these HGT. Surprisingly, they spoke of a "*startling anti-Darwninan outcome*" because extensive gene flow within the original population prevents to define a unique common ancestor containing all genes present in the two diverging populations! However, Darwin, who ignored the existence of genes, never discussed the genetic composition of ancestors of two populations. One again, there is here a clear confusion between genes and organisms.

THE DARWINIAN THRESHOLD AND THE VEIL OF COMMUNAL EVOLUTION

In the post-genomic era, positivists themselves are not immune to the influence of microbialists. Hence, whereas Woese's brilliant work has experimentally demonstrated that it was possible to decipher the history of microorganisms despite HGT, he reinstated himself later on HGT at the center of the universal tree, hiding the root itself under the veil of communal evolution (Woese, 2000, 2002). Carl Woese introduced the term "*Darwinian threshold*" to characterize the transition between the first period in life history, during which evolution has supposedly occurred mainly through HGT, and the second epoch (we are still living in) characterized by "Darwinian evolution" (formation and diversification of species) (Woese, 2002).

In the communal scenario proposed by Woese, primitive organisms (progenotes) living before the Darwinian threshold exchanged genes (thus characters) freely (and extensively) because they exhibited a loose modular fabric. As a consequence, different genetically encoded modules could be exchanged without damage for the organisms. Innovations occurred nearly simultaneously in the whole biosphere, without the possibility to form stable evolutionary lineages (species). Progenotes could not really compete and be selected since they were all quite similar at all stages of early evolution. Three distinct types of molecular biology finally "crystallized" to form the ancestor of each modern domain (the beginning of speciation) reducing dramatically the possibility of HGT between domains (at least for basic cellular informational processes). In a paper exploring the importance of HGT in the evolution of the genetic code toward universality and optimality, Woese and co-workers wrote that: "*evolution of the genetic code; translation, and cellular organization itself follows a dynamic whose mode is, if anything, Lamarckian*" Vetsigian et al., 2006).

Notably, the question of the root of the universal tree (a fundamental problem in biology) becomes futile in the communal scenario. Carl Woese wrote "*the universal tree has no root in the classical sense, the root is actually a Darwinian threshold*" (Woese, 2002). He simply suggested that Bacteria crossed first the Darwinian threshold, followed by Archaea and Eukarya (in that order) to fit with the traditional rooting in the bacterial branch (Woese, 2002) (ladder thinking, see Forterre, 2010b). A fundamental scientific question is thus abandoned (as in the case of fusion scenarios that roots *a priori* the tree between Archaea and Bacteria). We should renounce, for instance, to determine if traits shared by two of the three domains are ancestral or derived. In fact, we cannot anymore apply cladistic principles of evolutionary phylogenetic (Hennig, 1966) that have been so successful in deciphering the history of macrobes.

In fact, the concept of Darwinian threshold supposes that we abandon the powerful evolutionary mechanism discovered by Darwin (variations plus selection) as an explanation for the most difficult problem of all, how life originated and evolved from inanimate matter to the modern biosphere? This idea is at odd with the principle of continuity, a powerful weapon against creationism, as important now as it was at the time of Darwin (see below). In fact, in the communal scenario, we watered down the extraordinary multiplicative power of life (disruptive of communities), a phenomenon crucial for natural selection, to introduce instead a mysterious progressive force, reminiscent of the *pouvoir de vie* of Lamarck. Hence, Carl Woese wrote: "*a stage inevitably will be reached* (during communal evolution) *were some cellular entities become complex enough that their cell design start to be unique*" (Woese, 2000). This suggests indeed that life evolves inevitably toward complexity, but we do not know how and why.

The communal scenario also supposes that life originated and evolved up to the last common ancestors of the three domains in a very limited spatial environment to allow all cells in the evolving communal population to acquire rapidly any beneficial novelty by HGT. Koonin and Martin (2005) have indeed propose a scenario in which life evolved up to the emergence of Archaea and Bacteria in the confined environment of a single hydrothermal chimney. I think that such scenario is very unlikely. My guess is that (again considering the extraordinary multiplicative power of life) the whole planet was already covered by a biosphere at the time of the Last Universal Common Ancestor (LUCA) of modern cells. LUCA and its contemporaries were indeed already quite complex cellular organisms (Forterre, 2010b) with an already optimized genetic code (Vetsigian et al., 2006). However, there is no experimental data showing that extensive HGT would prevent speciation, even in a confined space. On the contrary, a recent study of archaeal populations living in a single hot terrestrial pool have shown that, even in such confined environment, microbial populations can diverge and form new species despite extensive HGT by acquiring specific traits that allow them to adapt to specific metabolic niches (Cadillo-Quiroz et al., 2012). The same should have occurred for ancient (pre-LUCA) microbial populations, except if we assume that no ecological differentiation take place at the time of LUCA, something very unlikely.

Fortunately, there is no reason to abandon Darwin's vision when we think of primordial evolution, even if HGTs were more prevalent during this period of life history (something *a priori* not obvious since HGT involves presently complex molecular mechanisms for DNA transfer from one cell to another). As previously mentioned, HGT cannot *per se* modify the nature of the evolutionary process (Forterre, 2012). Novelties can be transmitted by HGT but only fixed via the selection of individual. One could argue that HGT were so frequent before LUCA that a beneficial gene was rapidly transferred to the whole population, but the same would have been true for deleterious HGT! Parasites would have also invaded and disrupted such communal population (Poole, 2009). Again, only selection between a myriad of individuals could have make sense of these variations.

Positive selection was certainly at work, as a result of competition between parasites, predators, and preys (Forterre, 2005). Beside, neutral and purifying selection can have also play a major

role in building cellular complexity (Lukeš et al., 2011). In all cases, variation and selection were certainly the two pillars of life evolution in the pre-LUCA era as they are now. The formation of stable species might have indeed been prevented, but Darwinism cannot be assimilated to the formation of stable species, especially since, as previously discussed, the great merit of Darwin had been to focus on the variability of organisms and the lack of natural barrier between species, varieties and sub-varieties.

Even in the absence of "speciation," evolution of individual living entities before LUCA should be viewed in a tree-thinking framework. The evolution of replicators, protocells, and RNA-based cells produced myriads of trees in our distant past, and the best (or the lucky) replicators, protocells, and RNA-based cells were selected at each turning point of life history (the first organism with genomic RNA distinct from ribozyme RNA, the first organism with ribosomes producing encoded peptides, and the first organism with the modern genetic code, the first organism with DNA genomes). The winners were selected whereas many other types of organisms that once existed were eliminated (for instance organisms producing proteins—with possibly D amino-acids—using RNA-based machineries distinct from the ancestors of the ribosome). These selections were meaningful, thanks to variations that occurred in RNA replicons. Back in time, these complex chemical processes based on variation and selection probably originated from chemical kinetics selection that can favor the emergence of more complex replicators through imperfect replication (variation) and kinetic selection (Pross, 2011).

Why did Carl Woese propose the concepts of Darwininan threshold and communal evolution? One of his aims was probably to explain why the tempo of evolution was much higher at the time of LUCA (and before) than it became later on recently (Woese, 1998 and references therein). More precisely, he wonders why three different versions of universal proteins emerged in the relatively short period between the origin of proteins and LUCA, whereas they remained relatively similar within each domain until now, i.e., during a much longer period? This is an important question that Carl Woese raised early on, promoting the progenote concept (weak coupling between genotypes and phenotypes) as one possible explanation for the difference in evolutionary tempo before and after LUCA (Woese, 1998 and references therein). By introducing the Darwininan threshold concept, Woese now suggests that the tempo of evolution was faster before this threshold because the spread of characters by HGT accelerated the path of evolution. In his new model, three different versions of universal protein were established because the three domains crossed the "Darwininan threshold" at different times.

Explaining the change in evolutionary tempo that take place after the formation of the three domains is indeed an important issue. Hervé Phillipe and myself talked of: "*three dramatic evolutionary events*" that reduced this tempo at the onset of the three domains (Forterre and Philippe, 1999). I once suggested that these events were coupled to three independent transitions from RNA to DNA genomes that reduced the rate of genome evolution at the origin of each domain (Forterre, 2006a). Other hypotheses can be certainly put forward. However, I do not think that such events could be simply explained by a decrease in HGT frequency. In fact, we have no idea of the variation of HGT frequency during the history of life. Considering that HGT require today complex molecular machines for DNA transfer, it might be even possible that HGT were less frequent in ancient time than they are now. In fact, one could for instance argue that HGT were very rare at the time of LUCA, explaining why the three domains could indeed diverged so much, whereas more extensive HGT later on within domains homogenized molecular biology within but not between domains.

UNIFORMITARIANISM AND GRADUALISM

Darwin has insisted on gradualism (*natura non facit saltum*) because he had to fight against creation theories. He promoted the idea that evolution proceeds via the gradual accumulation of tiny modifications. This view is now another angle of attack against "Darwinism." In particular, genomics would have teached us that "*Natura facit saltum*" (see below) and that Darwin was wrong to insist on gradualism and uniformitarism in the biosphere (Koonin, 2009a,b). However, the history of science tells us that it does not make sense to claim that Darwin was wrong each time we discover a variation that is not so "gradual" after all. At the beginning of the XX century, "mutationism" was opposed to "Darwinism" because mutations were discrete, not gradual, modifications (a red eye becoming white) until population geneticists finally reconciled mutationism and Darwinism, showing that "*multiple Mendelian factors combined with random environmental effect to give apparently continuous variations*" (Barton et al., 2007). Today, it is claimed that HGT (once more), endosymbiosis, gene or genome duplications are not gradual too, introducing discontinuities in life history. For instance, Koonin (2009a,b) noticed that "*genome duplication is far from being an infinitesimal change.*" However, Darwin, who has no idea of the nature of variation, was speaking of ≪ tiny modifications ≫ at the phenotypic level. This debate in some way brings us back to the pre-modern synthesis era, instead of opening post-modernist avenues! Once again, if we look closely at the evolutionary process, I do not think that we need to forget gradualism so easily.

Let's take one example, mitochondrial endosymbiosis, *a priori* a dramatic discontinuity in the history of eukaryotes (and often viewed as such). The story started with a proto-eukaryote engulfing an alpha-proteobacterium. Interestingly, some modern eukaryotic cells still harbor endosymbiotic alpha-proteobacteria (Beninati et al., 2004; Park et al., 2009). Amazingly, in the tick *Ixodes ricinus*, the alpha-proteobacterium endosymbionts even parasitize mitochondria themselves (Beninati et al., 2004). The infected eukaryotic cells are not visibly different from their close relatives lacking endosymbionts. Similarly, the entrance of the alpha proteobacterium ancestor of mitochondria into an ancestor of modern eukaryotes most likely produced initially only minor phenotypic variations in the engulfing host. It's only the gradual accumulation of many (naturally selected) variations that transformed progressively the proto-eukaryote into the last common ancestor of modern eukaryotes. This evolutionary process was probably not cataclysmic but more likely take some time since, for instance, 19 eukaryotic specific proteins were added to the ancestral alpha-proteobacterium ribosome (and one bacterial protein lost) between the initial endosymbiosis at the origin of

the mitochondrion and the diversification of present-day eukaryotic supergroups (Desmond et al., 2011). **T**here are also many examples of drastic genome reduction in modern endosymbionts (e.g., Rickettsiae) that can be analyzed as the gradual accumulation of gene loss (variation), starting from an ancestor that was a free-living organisms and whose endosymbiotic descendants were gradually selected in a succession of discrete genome reductions steps (Andersson and Andersson, 1999; Renesto et al., 2005).

We should retain from Darwin's gradualist and uniformitarianist views that the basic mechanism explaining descent with modification (variation and selection) have been operating all along life history and, for instance, should serve as framework for origin of life scenarios (Chen et al., 2004). More generally, Darwin adopted the view of those geologists who realize that one should "*explain past changes to the earth in terms of mechanisms that could be studied in the present*" (Penny, 2011). It is essential to preserve this notion of continuity, even if life history went through phases involving very different types of organisms (RNA/peptide cells, RNA/protein cells, RNA/DNA/protein cells, RNA, and DNA viruses). We should imagine scenarios that explain the transition between these forms based on known biological mechanisms, connecting all these forms into convincing evolutionary stories and imagine how one form could have been transformed to another through small or large variations and fixed by selection (either neutral or positive).

EVOLUTIONARY SYNTHESIS EXTENDED TO .... VIRUSES

A recurrent theme in the evolutionary literature is that the "*evolutionary synthesis*" proposed in the middle of the last century should be completed or replaced by an "*extended evolutionary synthesis*" including in particular recent data obtained by studying metazoan development. The origin and evolution of multicellular metazoa is indeed an important aspect of life history, especially from a human perspective. However, macrobes concerned by these processes only represent a tiny fraction of the biosphere (Forterre, 2008). In my opinion, an updated evolutionary synthesis should *in priority* focus on including viruses in the history of life (Brüssow, 2009).

All cellular organisms from the three domains are infected by a plethora of viruses that co-evolved with their hosts and have dramatically altered their history (Prangishvili et al., 2006; Brüssow, 2009; Forterre and Prangishvili, 2009a). However, in the recent excellent textbook for students "*Evolution*" that exposes in parallel molecular biology and evolutionary biology (Barton et al., 2007) viruses are covered in 12 out of 782 pages! This is because viruses have been considered by most biologists to be simple by-products of cellular evolution. Hopefully, this situation is changing (Claverie, 2006; Forterre, 2006b; Brüssow, 2009; Forterre and Prangishvili, 2009b; Villarreal and Witzany, 2010). The qualitative and qualitative importance of viruses in all environments (from the Ocean to the human gut) has been recently realized, thanks to the development of viral ecology and to systematic studies of "viromes" (Rohwer and Thurber, 2009; Kristensen et al., 2010; Rohwer and Youle, 2011). Genomic studies have shown that cellular genomes are full of viral sequences or sequences evolutionary related to viral ones (probably derived from viruses) (Cortez et al., 2009; Feschotte and Clement, 2012). The study of viruses of microbes has made tremendous advances, as testified by the first international meeting devoted to them at the Institut Pasteur in 2010 (Koonin, 2010) and by the recent creation of the international society for the study of viruses of microbes. Viruses have been known for a long time as vehicles of cellular genes. However, viruses are first of all cradles of new genes (selected in viral genomes) and thus providers of new functions. The analysis of viromes has indeed revealed that viruses are the most important source of new genetic information in the biosphere (Rohwer and Youle, 2011 and references therein). This information is directly created during replication/recombination of viral genomes by gene duplication, recombination, insertion, frameshift, gene overlapping, retrotranscription, and so.

I recently introduced the concept of virocell (the infected cell producing virion and no more capable of classical cell division) to emphasize the intracellular creative phase of the virus life cycle (Forterre, 2010a, 2011b, 2012). Indeed, many evolutionists have difficulties to recognize the existence of *bona fide* viral genes. They reason as if all genes present in viral genome should have first originated in either archaeal, bacterial or eukaryotic genomes, before being transferred to viruses. Also, once a viral gene is integrated into a cellular genome, it is considered as a *bona fide* cellular gene in phylogenomic analyses, on the same footing as those inherited from cellular ancestors. For instance, genes present in "prophages" are often confused with "bacterial genes" in phylogenomic or metagenomic analyses. This has important consequences. In particular, introduction of foreign DNA coming from another cellular lineage (by transformation, conjugation or transduction) (true HGT) is not distinguished from introduction of foreign DNA via integration of viral DNA. However, these two types of HGT are completely different. In the first one (true HGT), viruses are not involved, or else eventually play the role of vehicles for cellular gene exchange (transduction), whereas in the second, they provide new genetic material of viral origin to the recipient cell. A prerequisite to understand so-called webs of life (recognizing underlying trees) would be to distinguish between these two types of HGT.

Since viruses and derived elements (plasmids, transposons and retrotransposons) mainly co-evolved with their hosts, HGT corresponding to viral integration does not usually blur the global phylogenetic signal present in cellular genomes (for cases studies, see Krupovic et al., 2010a,b; Soler et al., 2010) although they can produce a patchy distribution of characters that is difficult to interpret (**Figure 2A**). However, in some case, different viruses can introduce independently homologous viral proteins in different cellular lineages, a situation which will be usually interpreted wrongly as a real HGT between cells (**Figure 2B**). For instance, since the genomes of head and tailed viruses (Caudovirales) of Archaea and Bacteria sometimes encode homologous proteins (Krupovic et al., 2010a,b), a bacterial-like gene present in an archaeal genome might not testify for the transfer of a bacterial gene into this archaeon, but to the integration into the genome of this archaeon of an archaeovirus encoding a protein homologous to a protein encoded by a related bacteriovirus integrated into a bacterial genome. These confusing effects of integrated viral genes into cellular genomes should not be underestimated, considering that viral genes and evolutionary related elements

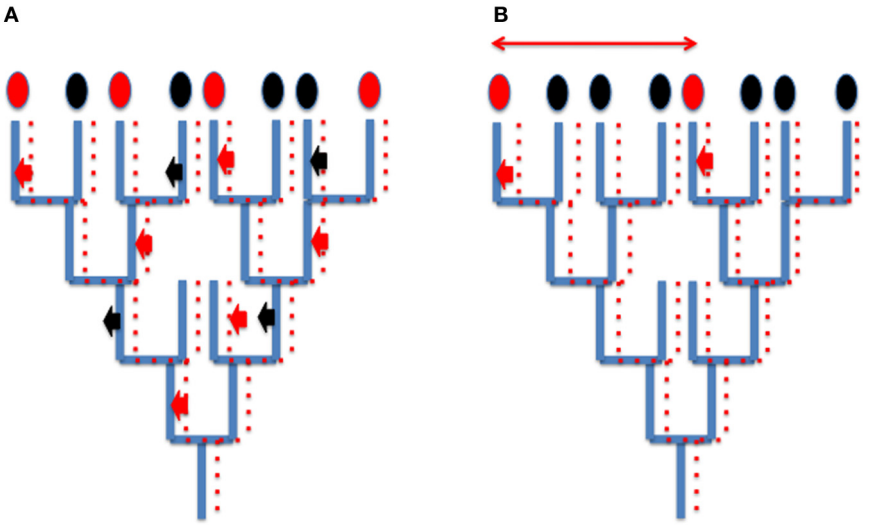


FIGURE 2 | How integration of viruses or related elements can confuse phylogenetic analyses? (A) Patchy phylogenetic distribution of viral genes in cellular genomes. A tree of organisms (blue lines) and a co-evolving viral (plasmid) lineages (dotted red lines). A viral (plasmid) gene is sometimes integrated (red arrow) sometimes loss (black arrows) from cellular genomes. The encoded viral proteins will appear as characters present (red ovals) or absent (black ovals) in cellular proteomes. Their use in whole genome tree construction will be misleading, grouping artificially organisms with common integrated viral (plasmid) genes. **(B)** Independent integration of viral genes encoding homologous proteins (small thick red arrows) mimicking horizontal gene transfer (thin red arrow) between two species.

(plasmids) represent a significant proportion of genomes in the three domains of life (Cortez et al., 2009; Feschotte and Clement, 2012 and references therein).

Viral integration can also mimic gene duplication. For instance many genes supposed to be paralogues (having originated by gene duplication in cellular genomes) might be homologous viral genes that have been introduced several times independently in the same cellular genome (for a case study, see the multiple integration of viral/plasmidic MCM helicases in Methanococcales, Krupovic et al., 2010a,b). Hence, it is unclear if multiple RNA polymerases, DNA polymerases, or MCM subunits in modern eukaryotic genomes originated by gene duplications (the common view) or multiple integrations of viral proteins (Forterre, 2006a).

I will argue here that confusion between cellular and viral genes partly explain the difficulties that many molecular evolutionists have to understand that the "web component" of gene trees (especially microbial ones), does not challenge Darwin but challenges the traditional view that confuse genes of viral and cellular origin.

VIRUSES AS MEDIATORS OF BIOLOGICAL EVOLUTION

Molecular biologists have now shown that viruses and related elements have played a major role in the origin of variations. Their integration into cellular genomes can inactivate cellular genes or promote various forms of genome recombination. Besides, when a viral genome becomes inserted into a cellular genome in regulatory regions, it can promote either activation or inactivation of neighboring genes, modifying the pattern of gene expression (de Parseval and Heidmann, 2005; Feschotte and Clement, 2012). These modifications can be drastic, especially if they touch genes controlling complex regulatory networks. Beside integration, viruses can live in symbiosis (carrier state) with cellular organisms (Ryan, 2007; Villarreal, 2007) providing another major route for the creation of diversity.

Importantly, the integration of viral genomes or the presence of viral symbionts brings at once new genes (hence, possibly new functions) into the cell. Many viral proteins encoded by these genes have been previously selected to interact with cellular proteins and manipulate cellular functions for the virus benefit. These proteins can now be recruited by the cells for their own purpose (exaptation) and help the cells to adapt to viruses but also to many other aspects of their environment (for instance when a bacteria recruited viral toxins to fight their eukaryotic predators). Since viral genomes replicate more often and are quantitatively more abundant than cellular genomes, it is possible that *in fine*, most cellular proteins originated first in the viral world (more precisely in virocells, see Forterre, 2010a, 2011b, 2012) and were only transferred later on into cellular lineages. One can conclude from all these considerations that interaction between viruses and cells has been probably (and still is) a (the) major source of variation (and novelties) in life history.

Darwin wondered about the multiplicative power of life, he would have been fascinated by the incredible multiplicative power of viruses. The huge number of infectious viral particles present in the biosphere has imposed a dramatic selection pressure (natural selection in grand scale) to natural populations all along life history. This has now been clearly established for modern marine viruses that fundamentally "*manipulate*" their environment, controlling the structure of microbial populations (Rohwer and Thurber, 2009 and references therein). Similarly, retroviruses and derived genetic elements have imposed a dramatic selection

pressure all along eukaryotic evolution, as testified by the huge number of endogenous retroviruses and derived elements now integrated into animals or plant genomes (Brosius, 2003; de Parseval and Heidmann, 2005; Feschotte and Clement, 2012). In fact, the major problem faced by any cellular population is how to adapt to their viral environment.

Considering the impact of viruses on both selection and variation, the two pillars of Darwin's core idea, the conflict between viruses and cells has been (and still is) probably the main engine of biological evolution (Forterre and Prangishvili, 2009b). In particular, the arm race between viruses and cells could partly explain the apparent tendency of life evolution toward complexity. This arm race has been probably a major source of novelties in the living world, much like arm races between tribes, cities and states have been a major factor of novelties in human history. More generally, the existence of parasites has been certainly a constant of life history (in agreement with uniformitarianism). As theoretically shown by Penny and co-workers in the case of the conflict between phagotrophs and their preys, "*there was no garden of Eden*" at the time of LUCA or before (De Nooijer et al., 2009), i.e., a "communal" word without parasites has never existed. The conflict between proto-viruses and RNA cells and later on between viruses and cells was probably a major evolutionary force at several critical steps in the history of life. This possibility has been explored to explain for instance the origin of DNA (Forterre, 2002) the origin of cell wall (Jalasvuori and Bamford, 2008), or else the emergence of unique eukaryotic features, such as the nuclear membrane, the telomere or else the odd mRNA capping structures (Forterre, 2011a and references therein).

It has been often claimed by anti-Darwinists, that the simple process of random mutations is not powerful enough to have produced the complexity and diversity of modern life forms. This is probably partly because they never consider in their reasoning the role that viruses have played in shaping cellular evolution. Unfortunately, Darwin was unaware of the existence of viruses, he would have been thrilled by the powerful tools for biological evolution hidden before our eyes, all these myriad of tiny "Darwinists" working days and nights to slowly but constantly change the face of the planet by promoting variation and selection.

CONCLUSION

At the dawn of the XXI century, some biologists apparently dream to bypass Darwin. For me, this is hopeless, except if we reduce Darwin to some *ad hoc* version of ≪ Darwinism ≫ or if we consider Darwin as one of our contemporaries, and not a scientist of the XIX century. With the dyad variation/selection, Darwin has provided us with key concepts that are necessary and sufficient to understand the logic of evolution, a goldmine that is still open. All the striking discoveries made in biology during the last 150 years have been extensions of these concepts and recent discoveries in microbial evolution and post-genomic studies are not different. We cannot bypass Darwin, but we can go beyond Darwin by the continuous exploration of the biosphere and the many particular mechanisms of life evolution. These mechanisms are much more diverse and sometimes complex than those imagined at the turn of the last century and new unexpected mechanisms certainly remained to be discovered. Of course, if we plan to reconstruct the history of life itself, especially those of ancient life, we are face to immense difficulties. However, we should not try to escape these difficulties by replacing trees by networks. In the meantime, we should go back to the fields to complete our inventory of microbes and their viruses, and be grateful to Darwin, who teaches us to look nature with open eyes beyond the veil of ideologies.

ACKNOWLEDGMENTS

I thank Drs David Penny, Céline Brochier-Armanet, and David Prangishvili for critical reading of first versions of this manuscript and Didier Raoult who invited me to contribute to this essay despite presently different views on Darwin and Darwinism.

REFERENCES

Andersson, J. O., and Andersson, S. G. (1999). Genome degradation is an ongoing process in Rickettsia. *Mol. Biol. Evol.* 16, 1178–1191.

Bapteste, E., Brochier, C., and Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* 1, 353–363.

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupré, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4, 34.

Barton, N. H., Briggs, D. E. G., Eisen, J., Goldstein, D. B., and Patel, N. H. (2007). *Evolution*. New York, NY: Cold Spring Harbour Laboratory Press.

Beninati, T., Lo, N., Sacchi, L., Genchi, C., Noda, H., and Bandi, C. (2004). A novel alpha-Proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl. Environ. Microbiol.* 70, 2596–2602.

Bos, L. (1999). Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 354, 675–685.

Brochier, C., Forterre, P., and Gribaldo, S. (2005). An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. *BMC Evol. Biol.* 5, 36.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). *Mesophilic Crenarchaeota*: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252.

Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* 14, 274–281.

Brochier-Armanet, C., Gribaldo, S., and Forterre, P. (2012). Spotlight on Thaumarchaeota. *ISME J.* 6, 227–230.

Brosius, J. (2003). The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118, 99–116.

Brüssow, H. (2009). The not so universal tree of life or the place of viruses in the living world. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2263–2274.

Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., Krause, D. J., and Whitaker, R. J. (2012). Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol.* 10:e1001265. doi: 10.1371/journal.pbio.1001265

Chen, I. A., Roberts, R. W., and Szostak, J. W. (2004). The emergence of competition between model protocells. *Science* 305, 1474–1476.

Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.

Claverie, J. M. (2006). Viruses take center stage in cellular evolution. *Genome Biol.* 7, 110.

Cortez, D., Forterre, P., and Gribaldo, S. (2009). A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol.* 10, R65.

Dagan, T., and Martin, W. (2006). The tree of one percent. *Genome Biol.* 7, 118.

Dagan, T., and Martin, W. (2009). Getting a better picture of microbial evolution en route to a network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2187–2196.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1987). Charles Darwin's Notebooks, 1836–1844: Geology, Transmutation of Species, Metaphysical Enquiries, eds P. H. Barrett, P. J. Gautrey, S. Herbert, D. Kohn, and S. Smith. Ithaca: Cambridge University Press.

De Nooijer, S., Holland, B. R., and Penny, D. (2009). The emergence of predators in early life: there was no Garden of Eden. *PLoS ONE* 4:e5507. doi: 10.1371/journal.pone.0005507

de Parseval, N., and Heidmann, T. (2005). Human endogenous retroviruses: from infectious elements to human genes. *Cytogenet. Genome Res.* 110, 318–332.

Desmond, E., Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011). On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Curr. Opin. Microbiol.* 14, 274–281.

Doolittle, W. F. (2009). The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2221–2228.

Doolittle, W. F., and Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2043–2049.

Feschotte, C., and Clement, G. (2012). Endogeneous retroviruses: insights into virl evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296.

Forterre, P. (2002). The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* 5, 525–532.

Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world, a story of viruses and cells. *Biochimie* 87, 783–803.

Forterre, P. (2006a). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3669–3674.

Forterre, P. (2006b). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5–16.

Forterre, P. (2008). In a world of microbes, where should microbiology stand? *Res. Microbiol.* 159, 74–80.

Forterre, P. (2010a). Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53, 362–378.

Forterre, P. (2010b). "The universal tree of life and the Last Universal Cellular Ancestor (LUCA): revolution and counter-revolutions," in *Evolutionary Genomics and Systems Biology*, ed G. Caetano-Anollés (Wiley-Blackwell), 43–62.

Forterre, P. (2011a). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res. Microbiol.* 162, 77–91.

Forterre, P. (2011b). Manipulation of cellular syntheses and the nature of viruses: the virocell concept. *C. R. Chimie* 14, 392–399.

Forterre, P. (2012). Virocell Concept, The. *eLS*. http://www.els.net. doi: 10.1002/9780470015902.a0023264

Forterre, P., and Philippe, H. (1999). Where is the root of the universal tree of life? *Bioessays* 21, 871–879.

Forterre, P., and Prangishvili, D. (2009a). The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann. N.Y. Acad. Sci.* 1178, 65–77.

Forterre, P., and Prangishvili, D. (2009b). The origin of viruses. *Res. Microbiol.* 160, 466–472.

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., and Woese, C. R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U.S.A.* 74, 4537–4541.

Galtier, N., and Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 4023–4029.

Gould, S. J. (1996). *Full House: the Spread of Excellence from Plato to Darwin*. New York, NY: Three Rivers Press.

Gribaldo, S., and Brochier, C. (2009). Phylogeny of prokaryotes: does it exist and why should we care? *Res. Microbiol.* 160, 513–521.

Gribaldo, S., Daubin, V., Forterre, P., Poole, A., and Brochier-Armanet, C. (2010). The origin of eukaryotes and their evolutionary relationship with Archaea: have we reached a phylogenomic impasse? *Nat. Rev. Microbiol.* 8, 743–752.

Halary, S., Leigh, J. W., Cheaib, B., Lopez, P., and Bapteste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 127–132.

Hennig, W. (1966). *Phylogenetic systematics*, eds D. Davis and R. Zangerl. Urbana: University of Illinois Press.

Huang, J., and Gogarten, J. P. (2006). Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* 22, 361–366.

Jalasvuori, M., and Bamford, J. K. (2008). Structural co-evolution of viruses and cells in the primordial world. *Orig. Life Evol. Biosph.* 38, 165–181.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267, 275–276.

Koonin, E. V. (2009a). Towards a postmodern synthesis of evolutionary biology. *Cell Cycle* 8, 799–800.

Koonin, E. V. (2009b). The origin at 150, is a new evolutionary synthesis in sight? *Trends Genet.* 25, 473–475.

Koonin, E. V. (2010). The wonder world of microbial viruses. *Expert Rev. Anti Infect. Ther.* 8, 1097–1099.

Koonin, E. V., and Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends Genet.* 21, 647–654.

Koonin, E. V., and Wolf, Y. I. (2009). Is evolution Darwinian or/and Lamarckian? *Biol. Direct* 4, 42.

Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2010). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19.

Krupovic, M., Forterre, P., and Bamford, D. H. (2010a). Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J. Mol. Biol.* 397, 144–160.

Krupovic, M., Gribaldo, S., Bamford, D. H., and Forterre, P. (2010b). The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* 27, 2716–2732.

Lukeš, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F., and Gray, M. W. (2011). How a neutral evolutionary ratchet can build cellular complexity? *IUBMB Life* 63, 528–537.

Maynard-Smith, J. (1991). "A Darwinian view of symbiosis," in *Symbiosis as Source of Evolutionary Innovation*, eds L. Margulis and R. Fester. (Cambridge, MA: MIT Press), 26–39.

Monod, J. (1971). *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology by Jacques Monod*. New York, NY: Vintage.

O'Malley, M. A. (2009). What did Darwin say about microbes, and how did microbiology respond? Trends. Microbiol. 17, 341–347.

Papke, R. T., and Gogarten, J. P. (2012). Ecology. How bacterial lineages emerge. *Science* 336, 45–46.

Park, M., Kim, M. S., Lee, K. M., Hwang, S. Y., Ahn, T. I. (2009). Characterization of a cryptic plasmid from an alpha-proteobacterial endosymbiont of *Amoeba proteus*. *Plasmid* 61, 78–87.

Penny, D. (2010). 20 points on the structure and testability of Darwin's theory. *Biol. Int.* 47, 11–35.

Penny, D. (2011). Darwin's theory of descent with modification, versus the biblical tree of life. *PLoS Biol.* 9:e1001096. doi: 10.1371/journal.pbio.1001096

Pigliucci, M. (2009). An extended synthesis for evolutionary biology. *Ann. N.Y. Acad. Sci.* 1168, 218–228.

Poole, A. M. (2009). Horizontal gene transfer and the earliest stages of the evolution of life. *Res. Microbiol.* 160, 473–480.

Prangishvili, D., Forterre, P., and Garrett, R. A. (2006). Viruses of the Archaea: a unifying view. *Nat. Rev. Microbiol.* 4, 837–848.

Puigbò, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J. Biol.* 8, 59.

Puigbò, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* 2, 745–756.

Pross, A. (2011). Toward a general theory of evolution: extending Darwinian theory to inanimate matter. *J. Sys. Chem.* 2, 1–14.

Raoult, D. (2010). The post-Darwinist rhizome of life. *Lancet* 375, 104–105.

Renesto, P., Ogata, H., Audic, S., Claverie, J. M., and Raoult, D. (2005). Some lessons from Rickettsia genomics. *FEMS Microbiol. Rev.* 29, 99–117.

Richards, R. J. (2009). Darwin's place in the history of thought: a reevaluation.*Proc. Natl. Acad. Sci. U.S.A.* 106(Suppl. 1), 10056–10060.

Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* 459, 207–212.

Rohwer, F., and Youle, M. (2011). Consider something viral in your search. *Nat. Rev. Microbiol.* 9, 308–309.

Ryan, R. F. (2007). Viruses as symbionts. *Symbiosis* 44, 11–21.

Sapp, J. (2005). The prokaryote-eukaryote dichotomy: meanings and mythology. *Microbiol. Mol. Biol. Rev.* 69, 292–305.

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., Polz, M. F., and Alm, E. J. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51.

Soler, N., Marguet, E., Cortez, D., Desnoues, N., Keller, J., van Tilbeurgh, H., Sezonov, G., and Forterre, P. (2010). Two novel families of plasmids from hyperthermophilic archaea encoding new families of replication proteins. *Nucleic Acids Res.* 38, 5088–5104.

Sonea, S. (1971). A tentative unifying view of bacteria. *Rev. Can. Biol.* 30, 239–244.

Sonea, S., and Paniset, M., (1976). Towards a new bacteriology. *Rev. Can. Biol.* 35, 103–167.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D. A., Wagner, M., and Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol.* 18, 331–340.

Valas, R. E., and Bourne, P. E. (2010). Save the tree of life or get lost in the woods. *Biol. Direct* 5, 44.

Vetsigian, K., Woese, C. R., and Goldenfeld, N. (2006). Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10696–10701.

Villarreal, L. P. (2007). Virus-host symbiosis mediated by persistence. *Symbiosis* 44, 1–9.

Villarreal, L. P., and Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* 262, 698–710.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271.

Woese, C. R. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6854–6859.

Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8392–8396.

Woese, C. R. (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8742–8747.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. (2002). Genome trees and the tree of life. *Trends Genet.* 18, 472–479.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 May 2012; accepted: 21 July 2012; published online: 08 August 2012.
*Citation: Forterre P (2012) Darwin's goldmine is still open: variation and selection run the world. Front. Cell. Inf. Microbio. **2**:106. doi: 10.3389/fcimb.2012.00106*