EPIGENETICS AS A DEEP INTIMATE DIALOGUE BETWEEN HOST AND SYMBIONTS

frontiers Research Topics

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**

EPIGENETICS AS A DEEP INTIMATE DIALOGUE BETWEEN HOST AND SYMBIONTS

Topic Editors:
Ilaria Negri, Koiné - Environmental Consulting S.n.c., Italy
Eva Jablonka, Tel-Aviv University, Israel
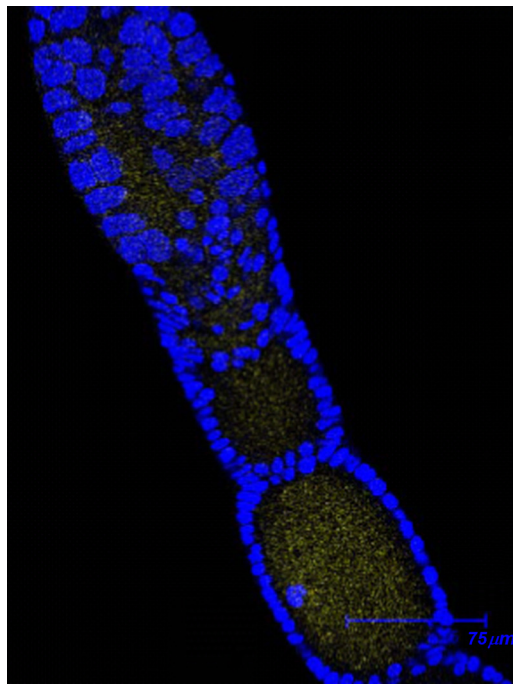
Fluorescent in situ hybridization of the endosymbiont *Wolbachia* (yellow dots) in the ovary of the insect *Zyginidia pullula*. Copyright © American Society for Microbiology, [Applied and Environmental Microbiology, 77, 2011, 1423-1435, 10.1128/AEM.02121-10]

Cover image by Fabio Sabatini

Symbiosis is an intimate relationship between different living entities and is widespread in virtually all organisms. It was critical for the origin and diversification of Eukaryotes and represents a major driving force in evolution. Indeed, symbiosis may support a wide range of biological processes, including those underlying the physiology, development, reproduction, health, behavior, ecology and evolution of the organisms involved in the relationship.

Although often confused with mutualism, when both organisms benefit from the association, symbiosis actually encompasses several and variable relationships. Among them is parasitism, when one organism benefits but the other is harmed, and commensalism, when one organism benefits and the other remains unaffected.

Even if many symbiotic lifestyles do exist in nature, in many cases the intimacy between the partners is so deep that the "symbiont" (sensu strictu) resides into the tissues and/or cells of the other partner. Since the partners frequently belong to different kingdoms, e.g. bacteria, fungi, protists and viruses living in association with animal and plant hosts, their shared "language" should be a basic and ancient form of communication able to effectively blur the boundaries between extremely different living entities.

In recent years studies on the role of epigenetics in shaping host-symbiont interactions have been flourishing. Epigenetic changes include, but are not limited to, DNA methylation, remodelling of chromatin structure through histone chemical modifications and RNA interference.
In this E-book we present a series of papers exploring the fascinating developmental and evolutionary relationship between symbionts and hosts, by focusing on the mediating epigenetic processes that enable the communication to be effective and robust at both the individual, the ecological and the evolutionary time scales. In particular, the papers consider the role of epigenetic factors and mechanisms in the interactions among different species, comprising the holobiont and host-parasite relationships. On the whole, since epigenetics is fast-acting and reversible, enabling dynamic developmental communication between hosts and symbionts at several different time scale, we argue that it could account for the enormous plasticity that characterizes the interactions between all the organisms living symbiotically on our planet.

Table of Contents

frontiers
in Genetics

CrossMark

Editorial: Epigenetics as a Deep Intimate Dialogue between Host and Symbionts

Ilaria Negri[1]* and Eva Jablonka[2]

[1] *Koiné - Environmental Consulting S.n.c., Parma, Italy,* [2] *The Cohn Institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Tel-Aviv, Israel*

The Editorial on the Research Topic

Epigenetics as a Deep Intimate Dialogue between Host and Symbionts

HOST-SYMBIONT EPIGENETIC CROSSTALK: A "*KOINÉ* LANGUAGE" THAT ENABLES COMMUNICATION BETWEEN DIFFERENT SPECIES

In ancient Greek, the term *koiné* refers to a common language, which spread in the Eastern Mediterranean region and the Near East following the conquests of Alexander the Great. In the 4th century BC, *koiné* quickly became a shared language that enabled people speaking different dialects to communicate with each other. Indeed, effective communication is always necessary for overcoming "barriers" between individuals and groups, tightening relationships, and enabling new ways of coping with a changing world. This is the case not only for interactions involving cultural relations. Different living entities communicate in many ways, with the most intimate and often the most long-lasting communication being the formation of mutually beneficial, obligatory symbiotic associations. As Lynn Margulis forcefully argued, symbiotic relations are found among virtually all living organisms, and were critical for the origin and diversification of the eukaryotes (Sagan, 1967; Margulis, 1971; Guerrero et al., 2013). Since the partners in symbiotic relationships frequently belong to different kingdoms, and the intimacy may be so deep that the Symbionts reside in the tissues and/or cells of the host, their shared "language" should be a basic—and ancient—form of communication. Such effective communication blurs the boundaries between different living entities, giving rise to a single biomolecular network, a "holobiont" with a "hologenome" (Zilber-Rosenberg and Rosenberg, 2008; Bordenstein and Theis, 2015), thus problematizing the conventional notion of individuality. The series of papers presented in this topic-issue explore the fascinating developmental and evolutionary relationship between Symbionts and hosts, by focusing on the mediating epigenetic processes that enable the communication to be effective and robust at both the individual, the ecological, and the evolutionary time scales.

One of the currently most researched cases illustrating the productive cross-talk between Symbionts and hosts is the symbiosis between mammals and their gut microbiome. As Gilbert stresses, the birth of a mammal is not merely the origin of a new distinct (traditional) individual, but the onset of a new community. Already *in utero*, the fetus interacts with a network of symbiotic

factors that is provided by the mother and, at birth, as it leaves the maternal symbiotic association system, it forms its own community which, although largely based on the legacy it gets from its mother, becomes rapidly adapted to its specific idiosyncratic conditions of life. Since the normal development and the thriving of the newborn depend on the symbiotic legacy it receives, it is the holobiont—the community of interacting species—that is the target of both developmental and natural selection. However, as Fridmann-Sirkis et al. illustrate, in extreme circumstances, when interactions fail, for example, because of acute stress to the microbiome, the host adapts by making adaptive changes in its epigenome, which can be inherited by subsequent generations. In more normal conditions, the maternal legacy provides crucial developmental resources, and is the basis on which flexible, context-sensitive, modifications of these resources, that allow rapid adaptation to the normally dynamic and fluctuating conditions in which the new individual lives, are constructed.

Since the microbiome reacts and evolutionarily adapts much faster than the host, most host's physiological adaptations are likely to be initiated by evolutionary changes to the constitution of the microbiome. Soen analyzes the manner in which the changes in the constitution of the microbiome lead to the modification of the communication process among the species comprising the microbiome, as well as between the microbiome and the host; and shows how these changes enable the developmental construction of a new dynamic, adaptive equilibrium, involving mutually beneficial epigenetic modifications in the host that can be epigenetically inherited between host generations. Hence, adaptive evolutionary modifications in the rapidly evolving microbiome, that occur at the time scale of host's individual life, are followed by the epigenetic inheritance of the host's physiological adaptations, thus enabling adaptive adjustments of the host at the ecological time scale (Soen; see also Moran and Sloan, 2015).

Although many symbiotic associations play essential roles in host development, physiology, and health, some associations may be harmful for the host. In these cases, the "selfish" Symbiont (often referred to as "parasite") must evade the host's immune responses, and this may occur through "misleading" the host, for example, by mimicking the host's signaling and control system for its own benefits. Some intracellular pathogenic bacteria, such as *Anaplasma phagocytophilum* a pathogen studied by Sinclair et al. encode eukaryotic-like proteins (e.g., the nucleomodulin AnkA) for subverting the host cells metabolism by recruiting chromatin-modifying enzymes or by altering the folding patterns of chromatin that bring distant regulatory regions together to coordinate control of transcriptional reprogramming. Other epigenetic factors are SET-domain containing proteins, which are known to modify chromatin structure in eukaryotic cells. Alvarez-Venegas shows that SET domain genes, which meditate host-symbiont epigenetic interactions, and that have been identified in several bacterial genomes on the basis of their similarity to the SET domains of eukaryotic histone

methyltransferases, allow pathogens to inhibit transcriptional activation of host defense genes. Indeed, there is plenty of evidence supporting the role of epigenetic mechanisms (e.g., DNA methylation, mechanisms that re-model chromatin structure through histone modifications, and mechanisms underlying RNA interference) both in pathogen plasticity and pathogen-induced alterations of the host (Gómez-Díaz et al., 2012). One such example is the molecular pathogenesis of Epstein–Barr virus (which induces diverse lymphoid and epithelial malignancies) discussed by Niller et al. that is accompanied by epigenetic alterations of both the viral and the host genomes.

Some "parasites" promote genetic exchange with the host genome. According to many studies, the exchange of human and viral genes must have happened repeatedly during evolution, and this can account for the high degree of homology between many viral and human genes. Niller et al. point to the example of the thymidylate synthase gene in the genome of *Herpesvirus saimiri*, which has 70% amino acids homology to that of the human gene, and Shapiro reviews and discusses the already massive and still growing evidence showing that such invasions into a host genome through virally-transmitted genetic elements (transposons, retrotransposons, and genomic proviruses) rewire the hosts' genomic networks, rearrange the host's genome and exert new types of selection pressures on it. The evolution of epigenetic silencing strategies, that prevent mobile DNA from destroying the integrity of the host genome and that have epigenetic heritable effects, is one such genomic adaptation. Indeed, it has been argued that one of the main functions of epigenetic mechanisms is a form of "genome immunity," promoting protection against foreign genetic elements. Sagy et al. suggest that this cellular immune function may explain the systematic failure of cloning some eukaryotic genomes in bacteria. For example, about 20% of the nematode *Caenorhabditis elegans* genome is not clonable in *Escherichia coli*, but can be cloned in yeast, a eukaryote. Interestingly, the bacterially unclonable sequences are enriched with repetitive DNA transposons and PIWI-interacting small RNAs (piRNAs). In worms and many other organisms (other animals and protists), piRNA-mediated RNA interference has a role in genome surveillance against foreign sequences. Sagy et al. therefore propose that piRNAs may act *in trans* to eliminate bacteria (*E. coli*), thus preventing its ability to be cloned within the bacterium. As they note, this role of piRNAs may have therapeutic applications, and could be used against pathogenic bacteria. Another potential epigenetic-medical intervention can be based on a genome-wide profiling of histone modifications combined with gene expression in the human malaria vector *Anopheles gambiae*. Targeting changes in histone modification and transcription of genes that regulate malaria transmission may provide, as Gómez-Díaz et al. point out, new type of malaria control. Deciphering the mechanisms underlying the molecular "dialogue" between hosts and symbionts therefore seems of major importance for many research areas including medicine, genetics, cell biology, zoology, microbiology, evolutionary biology, and ecology.

CONCLUDING REMARKS AND PERSPECTIVES

The papers published in this topic-issue, which focus on the role of epigenetic factors and mechanisms in the interactions among the different species comprising the holobiont, point to important research directions. Epigenetic mechanisms enable dynamic developmental communication, an epigenetic *koiné*, between hosts and symbionts at several different time scales. We are confident that the molecular-epigenetic technologies already available—and those that are rapidly being developed—will provide important insights into the evolution and development of the organisms on our planet, whose history and future are based on ongoing communication and interactions.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Bordenstein, S., and Theis, K. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* 13:e1002226. doi: 10.1371/journal.pbio.1002226

Gómez-Díaz, E., Jordà, M., Peinado, M. A., and Rivero, A. (2012). Epigenetics of host–pathogen interactions: the road ahead and the road behind. *PLoS Pathog.* 8:e1003007. doi: 10.1371/journal.ppat.1003007

Guerrero, R., Margulis, L., and Berlanga, M. (2013). Symbiogenesis: the holobiont as a unit of evolution. *Int. Microbiol.* 16, 133–143. doi: 10.2436/20.1501.01.188

Margulis, L. (1971). The origin of plant and animal cells. *Am. Sci.* 59, 230–235.

Moran, N. A., and Sloan, D. B. (2015). The hologenome concept: helpful or hollow? *PLoS Biol.* 13:e1002311. doi: 10.1371/journal.pbio.1002311

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* 14, 255–274. doi: 10.1016/0022-5193(67)90079-3

Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* 32, 723–735. doi: 10.1111/j.1574-6976.2008.00123.x

A holobiont birth narrative: the epigenetic transmission of the human microbiome

Scott F. Gilbert[1,2]*

[1] Department of Biology, Swarthmore College, Swarthmore, PA, USA
[2] Biotechnology Institute, University of Helsinki, Helsinki, Finland

This essay plans to explore, expand, and re-tell the human birth narrative. Usually, human birth narratives focus on the origins of a new individual, focusing on the mother and fetus. This essay discusses birth as the origin of a new community. For not only is the eukaryotic body being reproduced, but so also are the bodies of its symbiotic microbes and so is the set of relationships between these organic components. Several parts of the new narrative are surprising: (1) bacterial symbionts might cause some of the characteristics of pregnancy and prepare a symbiotic community for transfer; (2) the first bacterial colonizers of the mammalian organism my enter the fetus prior to the lysing of the amniotic membrane and birth; (3) the same signals that often cause immunological attack against a microbe may serve under these conditions to signal homeostatic stability between symbiont and host; and (4) the mother may actively provide substances that promote the growth and settlement of helpful bacteria. The birth of the holobiont exemplifies principles of co-evolution, co-development, niche construction, and scaffolding. Birth is nothing less than the passage from one set of symbiotic relationships to another.

Keywords: holobiont, birth, symbiosis, colonization, individuality

RETHINKING THE BIRTH NARRATIVE

This essay plans to explore, expand, and re-tell the human birth narrative. Usually, human birth narratives focus on the origins of a new *individual*, focusing on the heroic trevails of the mother or the amazing journey of fetus. I wish to discuss birth as the origin of a new *community*[1]. For not only is the eukaryotic body being reproduced, but so also are the bodies of its symbiotic microbes and so is the set of relationships between these organic components. Not only are the nuclear and mitochondrial genomes being transmitted, so are the genomes of the symbiotic community, whose microbial genes outnumber those of the eukaryotic component by over 100-fold (Human Microbiome Project Consortium, 2012; McFall-Ngai et al., 2013). Birth is nothing less than the passage from one set of symbiotic relationships to another. The holobionts (mother, fetus, infant) are preserved, but the components of these consortia have changed.

Mammalian birth is one of the great biological stories—the fundamental mammalian symbiosis of mother-and-fetus becoming another fundamental mammalian symbiosis, that of mother-and-child or mother-and-pup. In the classical story, the two partners are obviously the mother and the conceptus (zygote, embryo, fetus, child). Their interactions constitute the grand idea behind the obstetric subspecialty of Maternal-Fetal Medicine, which justifies its existence through the claim that the mother and fetus constitute a co-organized interacting whole, the "maternal-fetal unit." Indeed, they are a symbiotic unit, a system, where the treatment of one affects the physiology of the other. The mother

influences the fetus, providing it with nutrition, oxygen, antibodies, and hormones for its growth. The fetus reciprocally influences the mother, changing her blood circulation, immune responsiveness, and metabolism, while providing her with hormones to retain pregnancy. The physiology of the mother changes as the pregnancy continues, with both the mother and the fetus producing hormones and other growth factors to influence the other's survival and development.

But there is a third major player in this symbiotic mix: the mother's microbes. The pregnant mammal is herself a symbiotic community, a holobiont, composed of numerous species, most of them bacterial (Rosenberg et al., 2007; Gilbert et al., 2012). Nine out of every 10 of the mother's cells are microbial (Bäckhed et al., 2005; Ley et al., 2006) and metagenomic sequencing has shown that each human has entered into partnerships with over 150 species of bacteria (Qin et al., 2010). These bacteria are actively metabolizing nutrients, and the blood being received by the fetus has been substantially altered by the mother's microbes. About 30–35% of the metabolites in mammalian blood has a bacterial origin (Wikoff et al., 2009; McFall-Ngai et al., 2013). Different microbes metabolize dietary products differently, and different diets promote the population of different microbial communities within the mother (Turnbaugh et al., 2006, 2008; Frankenfeld et al., 2014). In mice, for instance, nearly all the blood-born serotonin is made by symbiotic bacteria (Wikoff et al., 2009). So the fetus is not free of a mother's symbiotic associations, even if it is thought that the fetus is sterile or persists in a sterile environment. Rather, the microbes of the mother are interacting with it.

---

[1] Always be wary of male scientists telling birth narratives.

So the fetus survives and develops in a network of symbiotic relationships provided by the pregnant mammal. When the infant/pup is born, these young animals will be going from one set of symbiotic associations to another. They will be leaving the symbiotic consortium of the mother and forming their own symbiotic consortium. They will not be "independent organisms," nor will they ever be. One is always a holobiont. But at birth, one has to pass through different symbiotic associations. Birth is the process of leaving of one symbiotic association system and forming another.

And remarkably, this transition and transmission is mediated by the maternal holobiont. This transition is not as stark as might be imagined. Rather, the mother is both actively and passively engaged in providing the symbiotic community that will persist over the life of the newborn. The standard scientific birth narrative has been that the fetus is growing in sterile conditions within the amnion, and that bacterial colonization of the fetus is made possible as soon as the amnion breaks and the fetus travels through the birth canal (see Funkhouser and Bordenstein, 2013 for a history and critique of this view). Then, according to this hypothesis, as the fetus traverses the cervix and vagina, the bacteria residing there can enter this new body, whose immune system is not sufficiently mature to attack them. Somehow, though, only certain bacteria can be let into this body, while other bacteria must be kept out. Moreover, only certain areas (gut, pores, mouth, etc.) will be allowed to retain persistent colonies, and these places will select for different types of bacteria.

I wish to critique this standard model more thoroughly, for recent studies find several more interactive and important roles for the mother. The symbiotic microbiome must be understood as constituting a third set of inherited genes. In addition to the nucleus and mitochondria, the symbiotic microbiome is passed from one generation to the next (see Moran, 2007; Douglas, 2010; Gilbert, 2011). This inheritance can be vertically in the germ line (as is often the case in invertebrates) or horizontally by infection (as is often seen in vertebrates). In some cases, both types of transmission are used. In mammals, where the germline is not seen to contain symbiotic bacteria, the newborn acquires its symbionts from its immediate environment. However, the mammalian fetus does not just leave the uterus and passively acquire a new set of symbionts. Rather, the mother actively passes the symbiotic baton to the developing fetus, and she doesn't relinquish control as rapidly and immediately as one might expect from the standard story. Indeed, the colonization of the body, along with the first breath that changes the circulatory system of the newborn, is possibly the most important biological aspect of birth, and the mother will be playing an active role in this process.

PREPARATIONS FOR DELIVERY

The pregnant mother (i.e., the maternal holobiont) changes dramatically during pregnancy, as the body undergoes hormonal, immunological, and metabolic changes. These include fat gain starting early in pregnancy, and insulin resistance later in pregnancy. These two metabolic conditions, which are often detrimental to men and to non-pregnant women, are thought to be beneficial during pregnancy. At this time, increased adiposity and increased insulin resistance are thought to support fetal growth and to prepare the mother for lactation (Di Cianni et al., 2003; Lain and Catalano, 2007; Nelson et al., 2010). As we will see, this preparation for lactation is critical for the handover of symbiotic community from the mother to the infant.

Employing 91 women and assuming that stool samples accurately reflect the intestinal microbiota, Koren et al. (2012) used polymerase chain reactions to show that the gut microbiome of pregnant women changed dramatically during pregnancy. These included women who took probiotics during pregnancy and who had used antibiotics during either the first or second trimester. In most of the women, the first-trimester gut bacteria were similar to that of the general non-pregnant population, but the third-trimester samples differed significantly. In a majority of women, the relative abundances of Proteobacteria and Actinobacteria increased substantially ($P = 0.0004$ and 0.003, respectively) during this time between 13 and 33 weeks of pregnancy. Moreover, the microbial community became more streamlined, with the diversity much reduced by the third trimester.

The importance of these bacteria to normal pregnancy was demonstrated by transferring the bacteria from the healthy first-trimester and third-trimester women into healthy female germ-free mice. The mice receiving the bacteria from the stools of first-trimester pregnant women remained normal. However, within 2 weeks, the healthy formerly germ-free mice that received the third-trimester bacteria had a pregnancy-like metabolic syndrome, complete with insulin insensitivity, excessive weight gain, and increased markers of inflammation.

These bacteria were derived from the gut. The vaginal bacterial community has also been analyzed (Aagaard et al., 2012; Romero et al., 2014) and was found to have a dynamic pattern during gestation, returning to an essentially non-pregnant state toward the end of pregnancy. *Lactobacillus* species, however, appear to be enriched during pregnancy. Several enriched Lactobacci species digest glycogen, and they produce an acidic environment that prevents pathogenic infection (O'Hanlon et al., 2011). One particular Lactobacillus, *L. johnsonii* is found in the gut and vagina. In the gut it is an important component of the upper digestive tract and is critical for the processing of bile salts (Pridmore et al., 2004). However, *L. johnsonii* also produces a bacteriotoxic compound, Lactacin F, which prevents the growth of particular bacterial pathogens (Abee et al., 1994). So the vaginal microbiota appear to be helping the mother ward off infections of the reproductive tract during pregnancy.

Thus, there is a dramatic remodeling of the gut and vaginal microbiological communities over the course of pregnancy. Although the mechanisms have not been delineated, it appears that the hormonal changes of the host are changing the population of microbes in the gut and vagina. These are the microbial populations that will be experienced by the late-stage fetus as it leaves the birth canal.

INITIAL COLONIZATION: THE SOONERS AND THE INITIATION OF LABOR

It has long been assumed (Tissier, 1900) that the fetus develops within a sterile environment, and that when the amnion bursts during labor, the colonization could begin. The first microbes would reach the fetus as it was being born. These would be the

resident microbes of the birth canal. Later, microbes from the mother's breast and skin would be in line for colonizing the newborn baby.

This would be the obvious way. However, there appear to be "sooners."[2] New evidence affirms that the first settlers of this newfound fetal territory are colonists from the mother's microbiome that gain entry into the developing fetus, bypassing the placental and amniotic barriers. Strangely, bacteria have been found in normal amniotic fluid, in the umbilical cord, and in infant's first bowel movements, the meconium (see Funkhouser and Bordenstein, 2013). This would indicate that the bacteria were already in the fetus before birth. To experimentally observe whether maternal bacteria could be transferred into the fetus, Jimenez et al. (2008) fed pregnant mice milk that had been inoculated with genetically labeled *Enterococcus faecum* bacteria. A day before the mice were to be naturally born, the researchers performed a Caesarean section on the mice, delivering them aseptically. They found that their first bowel movement not only contained bacteria, but that some of the bacteria had the transgenic label that could only have been received through the oral cavity or gut of their mothers.

However, while the vaginal microbiota might prevent pathogenic infection, and while the gut bacteria may induce a metabolic syndrome in the mother, these sources do not appear to be where the initial colonizers are coming from. Surprisingly, data suggest that the first colonizing bacteria arise from the mouth and then work their way into the fetus while it is still within the amnion. Molecular studies also indicated that the early colonization of human neonates appears to be accomplished by bacteria originating from the oral cavity (Palmer et al., 2007; Human Microbiome Project Consortium, 2012; Jost et al., 2012; Milisavljevic et al., 2013). According to sequencing data (Stout et al., 2013; Aagaard et al., 2014; Prince et al., 2014), the neonatal gut microbiota do not resemble the maternal vaginal or gut microbiota, but contain populations of bacteria derived from a placental source that stems from the oral cavity. Moreover, specific bacteria that are found normally or pathologically in the oral cavity (and not in the lower gut or vagina) have been isolated from human amniotic fluid (Ernest and Wasilauskas, 1985; Douvier et al., 1999; Bearfield, 2002; and Han et al., 2004). Bacteria were formerly thought to be found in placentae only in those mothers at risk for preterm labor. However, Stout et al. (2013) have questioned this idea by identifying intracellular bacteria in normal term and preterm placentae[3].

The mechanism by which oral bacteria can get to the placenta is not yet known. However, one possibility is that *dendritic*

cells of the oral cavity transport bacteria to lymphatic tissue in the placenta (see Donnet-Hughes et al., 2010; Funkhouser and Bordenstein, 2013). The oral mucosa contains numerous populations of dendritic cells, and these cells migrate through the blood and lymphatic vesicles to lymphoid tissues to mediate tolerance or immunogenicity (Hovav, 2014). When reaching the lymphoid tissues, the dendritic cells can diapedese across the endothelial cells into the lymphoid tissues (see de la Rosa et al., 2003; Johnson and Jackson, 2014). In many cases, they transport bacteria or other potential pathogens with them, and they present these microbial cells to the lymphocytes. The uterine decidua has a population of resident lymphocytes, and these cells are essential for normal implantation and the lack of rejection of the fetus (Blois et al., 2004; Juretic et al., 2004; Laskarin et al., 2007; Zarnani et al., 2008). So the oral cavity has a mucosa with associated dendritic cells, and the placenta has a lymphoid region capable of receiving dendritic cells. Recently, it was shown that dendritic cells carrying pathogens in them can migrate to the placenta and enable their parasitic passengers to infect the fetus. The transplacental passage of the intracellular *Toxoplasmosis*-like parasite *Neospora caninum* in mice appears to be facilitated by such dendritic cells. Inoculation of pregnant mice with dendritic cells infected with *Neospora* resulted in the migration of these dendritic cells to the placenta, the transmission of the parasite to the offspring, and often the resulting neonatal death (Collantes-Fernandez et al., 2012). There is therefore a pathway by which bacteria in the oral cavity can be transported to the placenta.

But the innate immune system of the fetus should prevent these bacteria from entering the fetal gut. The proteins that regulate immediate ("innate") immune responses against bacteria in the adult are the Toll-like receptors (TLRs). Interestingly, the activity of these receptors appears to be down-regulated in the fetus. The amniotic fluid, in addition to providing suspension and anti-dessication protection to the embryo, also contains large concentrations of Epidermal Growth Factor. This protein prevents the function of the TLRs. So while the early digestive tract is being bathed by amniotic fluid (the mouth and anus are open and exposed to amniotic fluid), the bacteria might be accepted as colonizers (Good et al., 2012). So there is a pathway through which bacteria in the oral cavity can become the first colonizers of the fetus, even before the amniotic membrane has lysed.

THE SECOND WAVE: THE COLONIZATION OF THE COLON

When John Donne famously wrote that "No man is an island," he was fully correct from the sociological perspective. But from the bacterial perspective, a man is a remarkable island, and the rules of island colonization hold for bacteria (Costello et al., 2012). Those bacteria who arrrive first get a wide choice of options, and they restrict the conditions for the next wave of settlers. Once the amnion has broken, the fetus is exposed to a wide variety of microbes, mostly from the gut and birth canal. Vaginal delivery exposes the fetus and newborn to the microbes of the mother's vagina and gut, a microbiome that has changed over the course of pregnancy (Tannock et al., 1990; Makino et al., 2011). These bacteria appear to be very important, as babies born through Caesarean section (i.e., not passing through the birth canal) have an altered bacterial colonization pattern early in life compared

---

[2]"Sooner" comes from an American slang for those European settlers who entered the Oklahoma Territories of the United States prior to the legal opening of the land.

[3]This has led Prince et al. (2014) to speculate that these bacteria may be involved in the aberrant timing of parturition. It had been thought that preterm births were caused by infections ascending from the vagina. However, recent studies (Madianos et al., 2013) shows that these bacteria are not representative of pathogenic vaginal microbes but are most likely a set of oral microbes. Moreover, periodontal pathogens might reach the fetus and cause preterm births or fetal illness. These, according to Prince and colleagues, would be those that had colonized the placenta.

with vaginally delivered babies (Ley et al., 2006; Makino et al., 2013). These gut and vaginal bacteria initiate new host-symbiont relationships, and they will have an important role in the health of the holobiont (Conroy et al., 2009; Le Huërou-Luron et al., 2010). Babies born vaginally have gut bacterial communities that resemble those of the maternal vagina (Matsumiya et al., 2002; Dominguez-Bello et al., 2010). Those babies born by Caesarean section receive many of their colonizers from the hospital environment and from the mother's skin (Martirosian et al., 1995; Dominguez-Bello et al., 2010). Mother's milk also supplies bacteria that will reside in the newborn's gut (Martín et al., 2003; Collado et al., 2009; Solís et al., 2010; Garrido et al., 2012).

It takes over a year for the babies born by C-section to have a similar bacterial profile, and during this time, they have a lower microbial diversity, delayed colonization of important microbes (such as *Bacteroides*) and reduced lymphocyte responses (Guarner and Malagelada, 2003; Jakobsson et al., 2014). The ability of the fetus to decide which bacteria stay and which ones must be excluded is still very much a mystery. The process of colonization appears to involve many of the same molecules that are usually used to attack bacteria. It seems that at the core of either acceptance or rejection is recognition, and the reaction (rejection or tolerance) depends on the context in which the fetus receives these microbes.

Immunology is the science of recognition in contexts. The same signal may call out for destruction or synthesis depending on how it is presented. Here, the host recognizes symbiotic bacteria apparently by the same sets of molecules usually used to attack bacteria; but the turns this recognition into acceptance rather than attack, and in this particular context, these bacteria are actually encouraged to settle into our guts (Chu and Mazmanian, 2013; Lee et al., 2013). This, along with the bacterial regulation of pregnancy weight gain and the entry of bacteria into the fetus prior to tocolysis, has been another surprise. The innate immune system had been thought to recognize bacteria by their pathogen-associated molecular patterns (PAMPs) by their pattern recognition receptors (PRRs). It now turns out that PAMPs are found in all microbes, including symbionts, and the agents recognized by the host's PRRs are now called "microbe-associated molecular patterns" (MAMPs). In mice (as well as in numerous other organisms, including hydra and *Drosophila*), the activation of PRRs in the newborn induces a homeostatic integration of host and symbiont. It appears that the composition of the early microbiome dictates whether the response of the cells is inflammatory or tolerant (See Chu and Mazmanian, 2013).

THE POWER OF POSITIVE SELECTION

Now that the newborn's body is being colonized by microbes originating from the gut, vaginal, and oral cavities of the mother, the problem becomes one of specificity: Which bacteria are going to remain and which are to be eliminated? This is a critical question, and one in which research is just beginning. However, some of the research is pointing out how important the mother is in determining which populations persist. One of the most important bacteria present in the mother's intestines is *Bifidobacteria*. These bacteria provide several services to the neonate (Garrido et al., 2012). First, they actively prevent the colonization of the gut by pathogenic bacteria and help induce and sustain the immune system. Second, they provide essential vitamins to the infant (Lievin et al., 2000; Schell et al., 2002; Fukuda et al., 2011). They also increase the tight junctions that are necessary for tightly linking the intestinal epithelial cells together (Chichlowski et al., 2012). Surprisingly, this tightening of intestinal epithelial binding may be essential for the cognitive health of the infant (Hsiao et al., 2013). Bifidobacteria is a good bacterium to have as one of the colonizers of the gut, and genetic evidence supports the idea that mammals and Bifidobacteria have a co-evolutionary history of helping each other for at least 200 million years.

These Bifidobacteria are encountered as the fetus squeezes through the birth canal. Makino et al. (2013) found that specific strains of *Bifidobacteria* become translocated from the mother's intestine to the newborn's gut. When delivered vaginally, monophyletic representatives of the woman's intestinal *Bifidobacteria* formed colonies in the newborn within 3 days of birth. This did not occur in those babies born by Caesarean section, where Bifidobacterial counts were lower after the first week (Makino et al., 2013).

And the mother supports the growth of the newborn's Bifidobacteria colonies. One of the most interesting components of a newborn's diet is mother's milk. And here came another surprise: some of the complex sugars found in human mothers' milk are not digestible by the infant. Rather, they serve as food for certain bacterial symbionts such as *Bifidobacteria* that help the infants' bodies develop (Sela et al., 2011; Zivkovic et al., 2011; Yoshida et al., 2012; Underwood et al., 2013).

The genome of one of the most common Bifidobacteria (*B. longum subspecies infantis*) has been sequenced and shown to contain a remarkable region of DNA—a series of genes linked together and dedicated to the intake and digestion of complex sugars found specifically in the earliest secretions of mother's milk. This small (43 kb) unit is not found in related bacteria that are not part of the gut microbiome, and these data indicate a remarkable co-evolution between this symbiotic bacterium and its host, suggesting that the host product (human milk) and the microbial genome enabling the bacteria to use this product have reciprocally formed each other (Sela et al., 2008). Thus, there would be a co-evolution of human and microbe for the purpose of gut colonization by this microbe, a colonization that would benefit both.

Not only do the sugars in mother's milk feed the "good guys," these oligosaccharides "prevent the pathogens from latching onto healthy cells, routing trouble-makers into a dirty diaper instead" (Bode, 2012; Manthey et al., 2014; Shugart, 2014). Mother's milk may also help symbiosis by instructing changes in the immune system through microRNAs in its lipid fraction (Munch et al., 2013).

Although each baby starts with a unique bacterial profile, within a year the types and proportions of bacteria have converged to the adult human profile that characterizes the human digestive tract (Palmer et al., 2007). Upon weaning and ingesting solid food, the bacterial population changes again, to the more adult form (Pantoja-Feliciano et al., 2013). Interestingly, the types of bacteria allowed to colonize depend on (1) the prevalence of a particular species of microbes in the environment;

(2) which microbes have already entered the gut; (3) the genetics of the digestive tract; and (4) the diet one receives (Nicholson et al., 2012; Pacheco et al., 2012; Kashyap et al., 2013).

HOLOBIONT PROCREATION

So we have re-told the birth narrative from a holobiont perspective. In doing so, several surprising hypotheses have emerged, ideas that had not been part of the traditional account of pregnancy:

- bacteria help regulate pregnancy
- bacteria can enter the fetus before the amnion breaks
- the activation of PRRs can mediate symbiosis as well as immune attack
- some material in mother's milk is for the bacteria, not the infant.

In this new narrative, we see "birth" as involving the reproduction of the holobiont. In other words, both the mammal and her persistent microbial populations have to be reproduced. Both niche construction and scaffolding are critical. *Niche construction* is defined as "the process whereby organisms, through their metabolism, their activities and their choices, modify their own and/or each other's niches" (Odling-Smee et al., 2003, p. 419). Our symbiotic bacteria employ such niche construction in specifying their environment by changing their host's development. The niches in which bacteria reside are to a large part generated by the bacteria, themselves. Some of the symbiotic microbes in the mouse intestine, for instance, induce gene expression in the gut epithelia not only to help the host, but to help themselves. The normal gut microbes, such as *Bacteroides*, induce gene expression in the Paneth cells of the intestine, instructing these cells to produce two compounds—angiogenin-4 and RegIII—that prevent the colonization of the intestine by *other* species of microbes. *Bacteroides, Escherichia coli*, and other symbiotic species are impervious to this compound, while several pathogenic Gram-positive bacteria (*Enterococcus faecalis* and *Listeria monocytogenes*) are wiped out by it (Hooper et al., 2003; Cash et al., 2006). These bacteria are enemies of Bacteroides and of the mammalian host. Thus, the microbial species is modifying its niche, causing its environment (i.e., the mammal) to change in such a way that they can better survive.

Scaffolds are "material environmental inputs with organizations that are sensitive and responsive to the developmental state of the developmental system being scaffolded" (Griesemer, 2014). A scaffold facilitates developmental processes that would be difficult or costly without it, and the scaffold is often temporary. Such scaffolds are critical and reciprocal parts of the holobiont's birth. First, we've seen that the bacteria are part of the scaffolding that allows human reproduction. If some of the "symptoms" of late pregnancy that support the fetus and its delivery are caused by bacteria, then the bacteria is part of the scaffolding of human reproduction. And if the milk sugars of the mother and the modified immune system of the newborn enable the successful reproduction of a particular set of bacteria (that will enable the completion of the developmental capacities of the newborn),

then humans are a critical scaffolding for the reproduction of the microbes.

The mother, through her hormones, anatomy, and milk production, is in large part responsible for the successful handing over of the fetus to a new set of symbionts. Going from the maternal environment to the outside world is not merely leaving a symbiotic support system and gaining "independence." There is no such thing as "independence." It's mutual dependency all the way down, and birth is the exchanging of one symbiotic system for another.

ACKNOWLEDGMENTS

REFERENCES

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. *Sci. Transl. Med.* 6, 237ra65. doi: 10.1126/scitranslmed.3008599

Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T. A., Coarfa, C., et al. (2012). A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PLoS ONE* 7:e36466. doi: 10.1371/journal.pone.0036466

Abee, T., Klaenhammer, T. R., and Letellier, L. (1994). Kinetic studies of the action of lactacin F, a bacteriocin produced by *Lactobacillus johnsonii* that forms poration complexes in the cytoplasmic membrane. *Appl. Environ. Microbiol.* 60, 1006–1013.

Bäckhed, F., Ley, R. E., Sonnenbury, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920. doi: 10.1126/science.1104816

Bearfield, C. (2002). Possible association between amniotic fluid micro-organism infection and microflora in the mouth. *Brit. J. Obst. Gynaecol.* 109, 527–533. doi: 10.1111/j.1471-0528.2002.01349.x

Blois, S. M., Alba Soto, C. D., Tometten, M., Klapp, B. F., Margni, R. A., and Arck, P. C. (2004). Lineage, maturity, and phenotype of uterine murine dendritic cells throughout gestation indicate a protective role in maintaining pregnancy. *Biol. Reprod.* 70, 1018–1023. doi: 10.1095/biolreprod.103.022640

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 22, 1147–1162. doi: 10.1093/glycob/cws074

Cash, H. L., Whitman, C. V., Benedict, C. L., and Hooper, L. V. (2006). Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 313, 1126–1130. doi: 10.1126/science.1127119

Chichlowski, M., De Lartigue, G., German, J. B., Raybould, H. E., and Mills, D. A. (2012). Bifidobacteria isolated from infants and cultured on human milk oligosaccharides affect intestinal epithelial function. *J. Pediatr. Gastroenterol. Nutr.* 55, 321–327. doi: 10.1097/MPG.0b013e31824fb899

Chu, H., and Mazmanian, S. K. (2013). Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat. Immunol.* 14, 668–675. doi: 10.1038/ni.2635

Collado, M. C., Delgado, S., Maldonado, A., and Rodriguez, J. M. (2009). Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett. Appl. Microbiol.* 48, 523–528. doi: 10.1111/j.1472-765X.2009.02567.x

Collantes-Fernandez, E., Arrighi, R. B., Alvarez-García, G., Weidner, J. M., Regidor-Cerrillo, J., Boothroyd, J. C., et al. (2012). Infected dendritic cells facilitate systemic dissemination and transplacental passage of the obligate intracellular parasite *Neospora caninum* in mice. *PLoS ONE* 7:e32123. doi: 10.1371/journal.pone.0032123

Conroy, M. E., Shi, H. N., and Walker, W. A. (2009). The long-term health effects of neonatal microbial flora. *Curr. Opin. Allergy Clin. Immunol.* 9, 197–201. doi: 10.1097/ACI.0b013e32832b3f1d

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J., and Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science* 336, 1255–1262. doi: 10.1126/science.1224203

de la Rosa, G., Longo, N., Rodríguez-Fernández, J. L., Puig-Kroger, A., Pineda, A., Corbí, A. L., et al. (2003). Migration of human blood dendritic cells across endothelial cell monolayers: adhesion molecules and chemokines involved in subset-specific transmigration. *J. Leukoc. Biol.* 73, 639–649. doi: 10.1189/jlb.1002516

Di Cianni, G., Miccoli, R., Volpe, L., Lencioni, C., and Del Prato, S. (2003). Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes Metab. Res. Rev.* 19, 259–270. doi: 10.1002/dmrr.390

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., et al. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11971–11975. doi: 10.1073/pnas.1002601107

Donnet-Hughes, A., Perez, P. F., Doré, J., Leclerc, M., Levenez, F., Benyacoub, J., et al. (2010). Potential role of the intestinal microbiota of the mother in neonatal immune education. *Proc. Nutr. Soc.* 69, 407–415. doi: 10.1017/S0029665110001898

Douglas, A. E. (2010). *The Symbiotic Habit.* Princeton, NJ: Princeton University Press.

Douvier, S., Neuwirth, C., Filipuzzi, L., and Kisterman, J. P. (1999). Chorioamnionitis with intact membranes caused by *Capnocytophaga sputigena. Eur. J. Obstet. Gynecol. Reprod. Biol.* 83, 109–112. doi: 10.1016/S0301-2115(98)00240-1

Ernest, J. M., and Wasilauskas, B. (1985). Capnocytophaga in the amniotic fluid of a woman in preterm labor with intact membranes. *Am. J. Obstet. Gynecol.* 153, 648–649. doi: 10.1016/S0002-9378(85)80251-9

Frankenfeld, C. L., Atkinson, C., Wähälä, K., and Lampe, J. W. (2014). Obesity prevalence in relation to gut microbial environments capable of producing equol or O-desmethylangolensin from the isoflavone daidzein. *Eur. J. Clin. Nutr.* 68, 526–530. doi: 10.1038/ejcn.2014.23

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., et al. (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469, 543–547. doi: 10.1038/nature 09646

Funkhouser, L. J., and Bordenstein, S. R. (2013). Mom knows best: the universality of maternal microbial transmission. *PLoS Biol.* 11:e1001631. doi: 10.1371/journal.pbio.1001631

Garrido, D., Barile, D., and Mills, D. A. (2012). A molecular basis for bifidobacterial enrichment in the infant gastrointestinal tract. *Adv. Nutr.* 3, 415S–421S. doi: 10.3945/an.111.001586

Gilbert, S. F. (2011). "Genetic and epigenetic sources of selectable variation," in *Transformations of Lamarckism: From Subtle Fluids to Molecular Biology*, eds E. Jablonka and S. Gissis (Cambridge: MIT Press), 283–293. doi: 10.7551/mit-press/9780262015141.003.0027

Gilbert, S. F., Sapp, J., and Tauber, A. I. (2012). A symbiotic view of life: we have never been individuals. *Q. Rev. Biol.* 87, 325–341. doi: 10.1086/668166

Good, M., Siggers, R. H., Sodhi, C. P., Afrazi, A., Alkhudari, F., Egan, C. E., et al. (2012). Amniotic fluid inhibits Toll-like receptor 4 signaling in the fetal and neonatal intestinal epithelium. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11330–11335. doi: 10.1073/pnas.1200856109

Griesemer, J. (2014). "Reproduction and scaffolded developmental processes: an integrated evolutionary perspective," in *Towards a Theory of Development*, eds A. Minelli and T. Pradeu (Oxford: Oxford University Press), 183–202.

Guarner, F., and Malagelada, J. R. (2003). Role of bacteria in experimental colitis. *Best Pract. Res. Clin. Gastroenterol.* 17, 793–804. doi: 10.1016/S1521-6918(03)00068-4

Han, Y., Redline, R., Li, M., and Yin, K. (2004). *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infect. Immun.* 72, 2272–2279. doi: 10.1128/IAI.72.4.2272-2279.2004

Hooper, L. V., Stappenbeck, T. S., Hong, C. V., and Gordon, J. I. (2003). Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat. Immunol.* 4, 269–273. doi: 10.1038/ni888

Hovav, A. H. (2014). Dendritic cells of the oral mucosa. *Mucosal Immunol.* 7, 27–37. doi: 10.1038/mi.2013.42

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., et al. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* 155, 1451–1463. doi: 10.1016/j.cell.2013.11.024

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234

Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., et al. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* 63, 559–566. doi: 10.1136/gutjnl-2012-303249

Jimenez, E., Marin, M. L., Martin, R., Odriozola, J. M., Olivares, M., Xaus, J., et al. (2008). Is meconium from healthy newborns actually sterile? *Res. Microbiol.* 159, 187–193. doi: 10.1016/j.resmic.2007.12.007

Johnson, L. A., and Jackson, D. G. (2014). Control of dendritic cell trafficking in lymphatics by chemokines. *Angiogenesis* 17, 335–345. doi: 10.1007/s10456-013-9407-0

Jost, T., Lacroix, C., Braegger, C. P., and Chassard, C. (2012). New insights in gut microbiota establishment in healthy breast fed neonates. *PLoS ONE* 7:e44595. doi: 10.1371/journal.pone.0044595

Juretic, K., Strbo, N., Crncic, T. B., Laskarin, G., and Rukavina, D. (2004). An insight into the dendritic cells at the maternal-fetal interface. *Am. J. Reprod. Immunol.* 52, 350–355. doi: 10.1111/j.1600-0897.2004.00232.x

Kashyap, P. C., Marcobal, A., Ursell, L. K., Smits, S. A., Sonnenburg, E. D., Costello, E. K., et al. (2013). Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17059–17064. doi: 10.1073/pnas.1306070110

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Bäckhed, H. K., et al. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150, 470–480. doi: 10.1016/j.cell.2012.07.008

Lain, K. Y., and Catalano, P. M. (2007). Metabolic changes in pregnancy. *Clin. Obstet. Gynecol.* 50, 938–948. doi: 10.1097/GRF.0b013e31815a5494

Laskarin, G., Kämmerer, U., Rukavina, D., Thomson, A. W., Fernandez, N., and Blois, S. M. (2007). Antigen-presenting cells and materno-fetal tolerance: an emerging role for dendritic cells. *Am. J. Reprod. Immunol.* 58, 255–267. doi: 10.1111/j.1600-0897.2007.00511.x

Lee, S. M., Donaldson, G. P., Mikulski, Z., Boyajian, S., Ley, K., and Mazmanian, S. K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* 501, 426–529. doi: 10.1038/nature 12447

Le Huërou-Luron, I., Blat, S., and Boudry, G. (2010). Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects. *Nutr. Res. Rev.* 23, 23–36. doi: 10.1017/S0954422410000065

Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi: 10.1016/j.cell.2006.02.017

Lievin, V., Peiffer, I., Hudault, S., Rochat, F., Brassart, D., Neeser, J. R., et al. (2000). Bifidobacterium strains from resident infant human gastrointestinal microflora exert antimicrobial activity. *Gut* 47, 646–652. doi: 10.1136/gut.47.5.646

Madianos, P. N., Bobetsis, Y. A., and Offenbacher, S. (2013). Adverse pregnancy outcomes (APOs) and periodontal disease: pathogenic mechanisms *J. Periodontol.* 84, S170–S180. doi: 10.1902/jop.2013.1340015

Makino, H., Kushiro, A., Ishikawa, E., Kubota, H., Gawad, A., Sakai, T., et al. (2013). Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota. *PLoS ONE* 8:e78331. doi: 10.1371/journal.pone.0078331

Makino, H., Kushiro, A., Ishikawa, E., Muylaert, D., Kubota, H., Sakai, T., et al. (2011). Transmission of intestinal *Bifidobacterium longum* subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. *Appl. Environ. Microbiol.* 77, 6788–6793. doi: 10.1128/AEM.05346-11

Manthey, C. F., Autran, C. A., Eckmann, L., and Bode, L. (2014). Human milk oligosaccharides protect against enteropathogenic *Escherichia coli* attachment *in vitro* and EPEC colonization in suckling mice. *J. Pediatr. Gastroenterol. Nutr.* 58, 167–170. doi: 10.1097/MPG.0000000000000172

Martín, R., Langa, S., Reviriego, C., Jiminez, E., Marin, M. L., Xaus, J., et al. (2003). Human milk is a source of lactic acid bacteria for the infant gut. *J. Pediatr.* 143, 754–758. doi: 10.1016/j.jpeds.2003.09.028

Martirosian, G., Kuipers, S., Verbrugh, H., van Belkum, A., and Meisel-Mikolajczyk, F. (1995). PCR ribotyping and arbitrarily primed PCR for typing strains of *Clostridium difficile* from a Polish maternity hospital. *J. Clin. Microbiol.* 33, 2016–2021.

Matsumiya, Y., Kato, N., Watanabe, K., and Kato, H. (2002). Molecular epidemiological study of vertical transmission of vaginal *Lactobacillus* species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction. *J. Infect. Chemother.* 8, 43–49. doi: 10.1007/s101560200005

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Loso, T., Douglas, A. E., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3229–3236. doi: 10.1073/pnas.1218525110

Milisavljevic, V., Garg, M., Vuletic, I., Miller, J. F., Kim, L., Cunningham, T. D., et al. (2013). Prospective assessment of the gastroesophageal microbiome in VLBW neonates. *BMC Pediatr.* 13:49. doi: 10.1186/1471-2431-13-49

Moran, N. A. (2007). Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8627–8633. doi: 10.1073/pnas.0611659104

Munch, E. M., Harris, R. A., Mohammad, M., Benham, A. L., Pejerrey, S. M., Showalter, L., et al. (2013). Transcriptome profiling of microRNA by Next-Gen deep sequencing reveals known and novel miRNA species in the lipid fraction of human breast milk. *PLoS ONE* 8:e50564. doi: 10.1371/journal.pone.0050564

Nelson, S. M., Matthews, P., and Poston, L. (2010). Maternal metabolism and obesity: modifiable determinants of pregnancy outcome. *Hum. Reprod. Update* 16, 255–275. doi: 10.1093/humupd/dmp050

Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., et al. (2012). Host-gut microbiota metabolic interactions. *Science* 336, 1262–1267. doi: 10.1126/science.1223813

Odling-Smee, F. J., Laland, K. N., and Feldman, M. W. (2003). *Niche Construction. The Neglected Process in Evolution. Monographs in Population Biology*, Vol. 37. Princeton, NJ: Princeton University Press.

O'Hanlon, D. E., Moench, T. R., and Cone, R. A. (2011). In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect. Dis.* 11:200. doi: 10.1186/1471-2334-11-200

Pacheco, A. R., Curtis, M. M., Ritchie, J. M., Munera, D., Waldor, M. K., Moreira, C. G., et al. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature* 492, 113–117. doi: 10.1038/nature11623

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., and Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol.* 5:e177. doi: 10.1371/journal.pbio.0050177

Pantoja-Feliciano, I. G., Clemente, J. C., Costello, E. K., Perez, M. E., Blaser, M. J., Knight, R., et al. (2013). Biphasic assembly of the murine intestinal microbiota during early development. *ISME J.* 7, 1112–1115. doi: 10.1038/ismej.2013.15

Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A. C., et al. (2004). The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2512–2517. doi: 10.1073/pnas.0307327101

Prince, A. L., Antony, K. M., Chu, D. M., and Aagaard, K. M. (2014). The microbiome, parturition, and timing of birth: more questions than answers. *J. Reprod. Immunol.* doi: 10.1016/j.jri.2014.03.006. [Epub ahead of print].

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Romero, R., Hassan, S. S., Gajer, P., Tarca, A. L., Fadrosh, D. W., Nikita, L., et al. (2014). The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* 2:4. doi: 10.1186/2049-2618-2-4

Rosenberg, E., Koren, O., Reshef, L.,Efrony, R. and Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease, and evolution. *Nat. Rev. Microbiol.* 5, 355–362. doi: 10.1038/nrmicro1635

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., et al. (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14422–14427. doi: 10.1073/pnas.212527599

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., et al. (2008). The genome sequence of *Bifidobacterium longum* subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18964–18969. doi: 10.1073/pnas.0809584105

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., et al. (2011). An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J. Biol. Chem.* 286, 11909–11918. doi: 10.1074/jbc.M110.193359

Shugart, J. (2014). *Mother Lode: Science News.* Available online at: https://www.sciencenews.org/article/mother-lode?mode=magazine&context=4527

Solís, G., de Los Reyes-Gavilan, C. G., Fernandez, N., Margolles, A., and Gueimonde, M. (2010). Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe* 16, 307–310. doi: 10.1016/j.anaerobe.2010.02.004

Stout, M. J., Conlon, B., Landeau, M., Lee, I., Bower, C., Zhao, Q., et al. (2013). Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *Am. J. Obste.t Gynecol.* 208, 226.e1–226.e7. doi: 10.1016/j.ajog.2013.01.018

Tannock, G. W., Fuller, R., and Pedersen, K. (1990). Lactobacillus succession in the piglet digestive tract demonstrated by plasmid profiling. *Appl. Environ. Microbiol.* 56, 1310–1316.

Tissier, H. (1900). *Recherches sur la Flore Intestinale des Nourrissons (e'tat Normal et Pathologique).* Paris: G. Carre and C. Naud.

Turnbaugh, P. J., Bäckhed, F., Fulton, L., and Gordon, J. I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* 3, 213–223. doi: 10.1016/j.chom.2008.02.015

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. doi: 10.1038/nature05414

Underwood, M. A., Kalanetra, K. M., Bokulich, N. A., Lewis, Z. T., Mirmiran, M., Tancredi, D. J., et al. (2013). A comparison of two probiotic strains of bifidobacteria in premature infants. *J. Pediatr.* 163, 1585–1591. doi: 10.1016/j.jpeds.2013.07.017

Wikoff, W. R., Anfora, A. T., Liu, J., Schultz, P. G., Lesley, S. A., Peters, E. C., et al. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3698–3703. doi: 10.1073/pnas.0812874106

Yoshida, E., Sakurama, H., Kiyohara, M., Nakajima, M., Kitaoka, M., Ashida, H., et al. (2012). Bifidobacterium longum subsp. infantis uses two different β-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology* 22, 361–368. doi: 10.1093/glycob/cwr116

Zarnani, A. H., Moazzeni, S. M., Shokri, F., Salehnia, M., Dokouhaki, P., Ghods, R., et al. (2008). Microenvironment of the feto-maternal interface protects the semiallogenic fetus through its immunomodulatory activity on dendritic cells. *Fertil. Steril.* 90, 781–788. doi: 10.1016/j.fertnstert.2007.01.102

Zivkovic, A. M., German, J. B., Lebrilla, C. B., and Mills, D. A. (2011). Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4653–4658. doi: 10.1073/pnas.1000083107

Delayed development induced by toxicity to the host can be inherited by a bacterial-dependent, transgenerational effect

Yael Fridmann-Sirkis[1†], Shay Stern[1†], Michael Elgart[1†], Matana Galili[1], Amit Zeisel[2], Noam Shental[3] and Yoav Soen[1]*

[1] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel
[2] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel
[3] Department of Computer Science, The Open University, Raanana, Israel

Commensal gut bacteria in many species including flies are integral part of their host, and are known to influence its development and homeostasis within generation. Here we report an unexpected impact of host–microbe interactions, which mediates multi-generational, non-Mendelian inheritance of a stress-induced phenotype. We have previously shown that exposure of fly larvae to G418 antibiotic induces transgenerationally heritable phenotypes, including a delay in larval development, gene induction in the gut and morphological changes. We now show that G418 selectively depletes commensal *Acetobacter* species and that this depletion explains the heritable delay, but not the inheritance of the other phenotypes. Notably, the inheritance of the delay was mediated by a surprising trans-generational effect. Specifically, bacterial removal from F1 embryos did not induce significant delay in F1 larvae, but nonetheless led to a considerable delay in F2. This effect maintains a delay induced by bacterial-independent G418 toxicity to the host. In line with these findings, reintroduction of isolated *Acetobacter* species prevented the inheritance of the delay. We further show that this prevention is partly mediated by vitamin B2 (Riboflavin) produced by these bacteria; exogenous Riboflavin led to partial prevention and inhibition of Riboflavin synthesis compromised the ability of the bacteria to prevent the inheritance. These results identify host–microbe interactions as a hitherto unrecognized factor capable of mediating non-Mendelian inheritance of a stress-induced phenotype.

Keywords: host–microbe interactions, non-Mendelian inheritance, *D. melanogaster*, epigenetics, stress responses

INTRODUCTION

Transgenerational epigenetic phenomena have been reported in a number of different animal and plant species (Cavalli and Paro, 1998; Morgan et al., 1999; Kaati et al., 2002; Sollars et al., 2003; Anway et al., 2005; Cropley et al., 2006; Molinier et al., 2006; Lumey et al., 2007; Xing et al., 2007; Greer et al., 2011; Rechavi et al., 2011; Seong et al., 2011; Crews et al., 2012; Stern et al., 2012). These events may be induced by the environment, and have been shown to involve a variety of host-intrinsic factors and molecular pathways (Jablonka and Raz, 2009), including chromatin modifiers (Cavalli and Paro, 1998; Greer et al., 2011; Seong et al., 2011), DNA methylation (Xing et al., 2007; Heijmans et al., 2008; Carone et al., 2010; Guerrero-Bosagna et al., 2010; Schmitz et al., 2011), RNAi machinery (Rechavi et al., 2011; Ashe et al., 2012; Buckley et al., 2012), and secreted signals (McCaffery et al., 1998). Involvement of other host intrinsic or extrinsic factors is also possible but has been largely overlooked. Of these, the commensal gut microbiome is a particularly attractive candidate. The gut microbiome is an integral part of the development and homeostasis of its host (Rosenberg and Zilber-Rosenberg, 2011; Charroux and Royet, 2012; Buchon et al., 2013), but nonetheless,

can be extensively modified by the environment. As with other environmental perturbations, disruption of the gut microbiome may lead to changes in the host which extend beyond one generation. However, potential multigenerational impacts of commensal gut bacteria on host development and physiology have not yet been investigated because typical studies assume that these bacteria affect each generation in a similar way.

Here, we examine multigenerational impacts of microbial changes using a model based on exposure of fly larvae to G418 toxicity in specific tissues. We have recently shown that this exposure can induce developmental phenotypes which persist for multiple generations in non-exposed offspring (Stern et al., 2012). In this system, we supplemented the larval food with G418 antibiotic and placed a resistance transgene fused to *GFP* (*neoGFP*) under the regulation of arbitrary, spatio-temporally restricted developmental promoters. This leads to toxic stress in tissues that are exposed to G418 but do not express sufficient levels of the "rescue," *neoGFP* gene. Exposure to G418 led to multiple phenotypes, including a delay in larval development, promoter-dependent induction of *neoGFP* expression and morphological changes in two promoter cases. Moreover, some of the induced

phenotypes persisted in a number of subsequent generations of non-exposed offspring (Stern et al., 2012). In particular, the delay in development and the induction of *neoGFP* expression were inherited at high penetrance and typically persisted for 3–10 generations without G418. One of the morphological phenotypes, wing abnormalities in the *Hsp70::neoGFP* case, was also heritable albeit at a much lower penetrance.

As has been previously shown, exposure of flies to an antibiotic (Chlortetracycline) can have a direct influence on the host tissue as well as an indirect effect mediated by an impact of the antibiotic on the commensal microbiome (Ridley et al., 2013). This rationale also applies to G418 which is an aminoglycoside which blocks polypeptide synthesis in both eukaryotic and prokaryotic cells. Thus, the above paradigm of G418-induced inheritance may provide a model for investigating potential contributions of host–microbe interactions to the inheritance of induced phenotypes in the host. We therefore, investigated the impact of G418 on the microbial composition and the resulting implications for induction and inheritance of responses in the host. We show that G418 leads to a selective depletion of commensal *Acetobacter* species. Removal of extracellular bacteria without exposure to G418 had an almost negligible effect on the first generation of bacterial-depleted larvae (F1), but nonetheless caused a considerable delay in larval development in the following generation (F2). The delay in offspring development following parental removal of gut bacteria was completely eliminated by re-introduction of a commensal *Acetobacter* species. Reintroduction of a commensal *Acetobacter* species also prevented the inheritance of the delay in development in offspring of G418-exposed flies. We further show that this prevention of the heritable delay is mediated in part by Riboflavin produced by the *Acetobacter* species.

These results show that environmental disruption of the gut microbiome can induce different effects in parents and offspring. They also uncover an unexpected scenario by which host–microbe interactions mediate the inheritance of delayed development in response to G418, namely: the delay in the parental generation is induced by a direct effect of G418 on the host tissue, but is maintained in non-exposed offspring by the transgenerational effect of *Acetobacter* depletion. Specifically, the *Acetobacter* depletion causes a modification in the parents which becomes phenotypic (delayed development) only in the offspring. We show that this transgenerational effect is responsible for the inheritance of the delay in development, but not for the inheritance of induced *neoGFP* expression and the inheritance of morphological changes.

The inter-generational difference between the rate of development in bacterial-depleted parents and offspring suggests that changes in the gut microbiome may influence the germline in the parents.

RESULTS

G418 SELECTIVELY DEPLETES *ACETOBACTER* SPECIES FROM THE LARVAL GUT

As a starting point for investigating the potential involvement of the gut microbiome in the response to G418, we tested if exposure to G418 modifies the composition of bacteria in the larva. We first analyzed the larval gut microbiome using an improved method of deep-sequencing of DNA coding for 16S ribosomal RNA (Amir et al., 2014). In line with recent findings (Wong et al., 2011), we identified in the gut of *hairy::neoGFP* larvae various *Acetobacter* and *Lactobacillus* spp. (**Supplementary Data Sheet 1**—Figure S1A, **Supplementary Data Sheet 2**). In addition to these extracellular species, we detected high abundance of the endosymbiont *Wolbachia*, which is known to exist in many lines of *D. melanogaster* (Bourtzis et al., 1996; Dobson et al., 1999; Veneti et al., 2003; McGraw and O'Neill, 2004) and is capable of manipulating various reproductive features of its host (Werren, 1997; Starr and Cline, 2002; Ikeya et al., 2009). The deep-sequencing analysis revealed strong reduction in the relative amount of *Acetobacter* spp. in G418-exposed larvae (**Supplementary Data Sheet 1**—Figure S1A). To validate this depletion, we developed a quantitative PCR-based assay capable of selectively measuring the total contents of *Acetobacter*, *Lactobacilus*, and *Wolbachia* spp. (**Supplementary Data Sheet 1**—Figures S1B–D). Measurement of the amounts of these three types of bacteria confirmed that G418 selectively depletes *Acetobacter* species (**Figure 1A**). Notably, the depletion of *Acetobacter* species persisted in non-exposed F2 offspring of G418 exposed flies (**Figure 1A**), indicating that the change in bacterial composition is itself heritable. Thus, G418 toxicity modifies the fly's microbiome by selectively depleting commensal *Acetobacter* species.

THE DELAY IN DEVELOPMENT IS INHERITED BY A BACTERIAL-DEPENDENT, TRANSGENERATIONAL EFFECT

We investigated whether the depletion of commensal bacteria might be responsible for the G418-induced phenotypes. We began by testing if removal of bacteria by egg dechorionation and sterilization (Brummel et al., 2004) can reproduce the delay observed in G418-treated larvae (Stern et al., 2012). We used the qPCR and bacterial growth assays to verify complete removal of extracellular bacteria (**Supplementary Data Sheet 1**–Figures S1E,F) and analyzed the influence of bacterial removal on pupation time. As previously demonstrated (Shin et al., 2011; Storelli et al., 2011), the effect of bacterial removal on pupation time depended on the food quality (**Supplementary Data Sheet 1**—Figure S2). Under the food conditions we have previously used (Stern et al., 2012), removal of bacteria led to a mild delay, which was considerably smaller than the delay caused by exposure to G418 (∼0.5 vs. 2–3 days, respectively; **Figure 1B**, **Supplementary Data Sheet 1**—Figure S3A). The lack of a considerable delay was consistent with previous work in which germ-free larvae developed in high quality diet (Storelli et al., 2011). Thus, depletion of extracellular bacteria did not reproduce the G418-induced delay in the first generation. Nevertheless, analysis of F2 *hairy::neoGFP* larval offspring of bacterial-depleted F1 parents revealed a very substantial delay, comparable to the delay observed in offspring of G418 exposed parents (**Figure 1C**). This effect of bacterial removal was also observed in three non-transgenic lines (**Figure 1D**). The difference between the rate of development in bacterial depleted parents and offspring depended on the diet, and was mostly evident within a particular range of diet quality (**Supplementary Data Sheet 1**—Figure S2). Within this range, removal of gut bacteria had a transgenerational
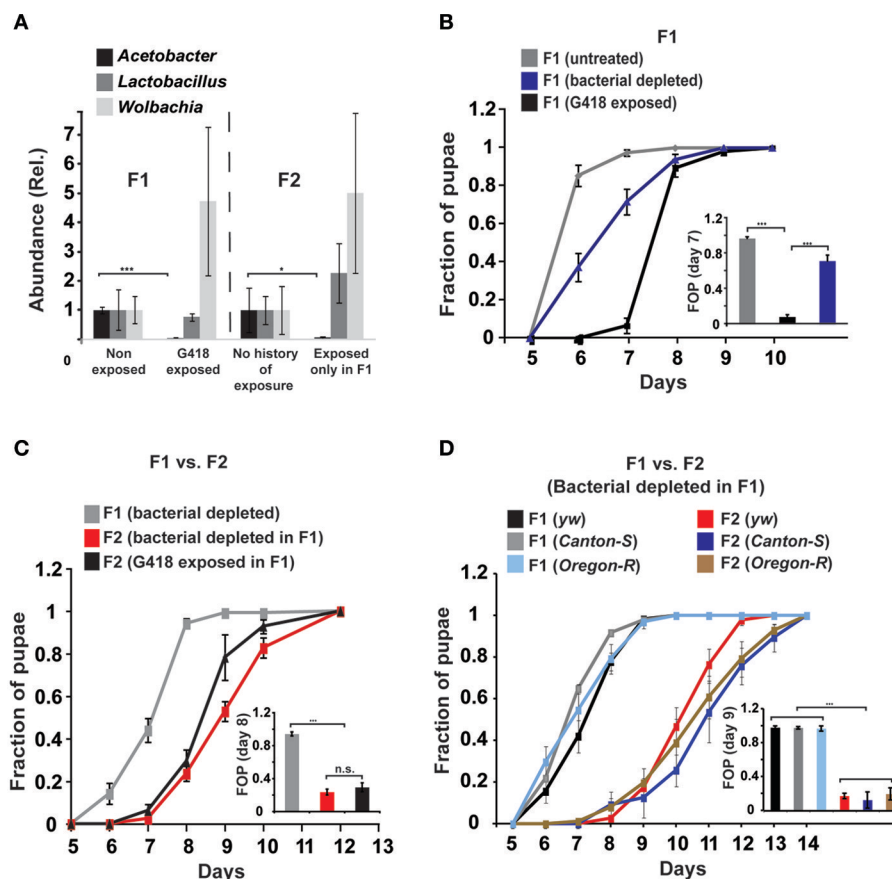
FIGURE 1 | Depletion of gut bacteria leads to a substantial delay in the development of the offspring, but not in the parental generation. (A) Selective and heritable depletion of *Acetobacter* spp. following exposure to G418. Shown are measurements corresponding to two consecutive generations. Abundance of each bacterial type is normalized to the measured abundance in larvae without a history of G418 exposure. Mean fold-change ± SE of type-specific 16S DNA sequences extracted from guts of third instar larvae (*n* > 7). **(B)** Effect of egg dechorionation and sterilization ("Bacterial depleted") on the kinetics of pupation of *hairy::neoGFP* larvae in the same generation. Note the small effect compared to the large delay following exposure to G418. Mean fraction of pupae ± SE in 7 vials. Inset: Statistical analysis of differences between fractions of pupae (FOP) in day 7. **(C)** Effect of egg dechorionation and sterilization in F1 on the kinetics of pupation in F1 and F2 *hairy::neoGFP* larvae. Note the large delay in F2 vs. F1 bacterial-depleted larvae, and the comparable kinetics of offspring of bacterial-depleted parents and offspring of G418-exposed parents. Inset: Statistical analysis of differences between fractions of pupae (FOP) in day 8. Mean fraction of pupae ± SE in at least three biological replicates. **(D)** Effect of egg dechorionation and sterilization in F1 on the kinetics of pupation in F1 and F2 in three non-transgenic fly lines (yw, Canton-S, and Oregon-R). Inset: Statistical analysis of differences between fractions of pupae (FOP) in day 9. *$p < 0.05$, ***$p < 0.001$ (Student's *t*-test).

effect that was not observed in the first generation of larvae lacking the bacteria. This transgenerational effect of bacterial depletion can account for the delay in development in non-exposed offspring of G418-exposed (and hence, *Acetobacter*-depleted) parents. As has been noted (Ridley et al., 2013), direct evidence can be obtained by antibiotic exposure of flies that had been depleted of their bacteria prior to the exposure. Applying this strategy in our system showed that removal of extracellular bacteria prior to G418 exposure led to significant reduction in the heritable delay (**Supplementary Data Sheet 1**—Figure S4A). This verifies the key contribution of bacterial depletion to the inheritance of delayed development in response to parental exposure to G418.

Other phenotypes reported by Stern et al., were largely unaffected by the mere depletion of bacteria. Removal of extracellular bacteria by egg dechorionation without

exposure to G418 had no effect on the levels of *neoGFP* in the gut of the larvae (**Supplementary Data Sheet 1**—Figure S3B). Similarly, we did not observe any of the morphological influences of G418 exposure in bacterial depleted flies (**Supplementary Data Sheet 1**—Figures S3C,D). Genome-wide analysis of mRNA levels in the proventriculus of bacterial-depleted *hairy::neoGFP* larvae further revealed very small overlap with the response to G418 (only 8 and 6% of genes that were up- and down-regulated in G418 were similarly affected by bacterial removal; **Supplementary Data Sheet 1**—Figure S3E, **Supplementary Data Sheet 3**). As expected, the set of transcripts that were down regulated following dechorionation were enriched for genes involved in response to bacterium ($p < 10^{-4}$; **Supplementary Data Sheet 1**—Figure S2F, **Supplementary Data Sheet 4**), e.g., *Attacin*, and *Defensin* (**Supplementary Data Sheet 1**—Figure S3J). The set of

up-regulated transcripts, on the other hand, was strongly enriched with genes associated with phospholipase activity and polysaccharide metabolism ($p < 10^{-8}$ and $p < 10^{-4}$, respectively; **Supplementary Data Sheet 1—Figure S3F**, **Supplementary Data Sheet 4**). These categories were not enriched in the set of proventriculus genes which responded to G418 (**Supplementary Data Sheet 1—Figure S3F**, **Supplementary Data Sheet 4**). Moreover, unlike in G418-exposed larvae (Stern et al., 2012), the depletion of bacteria did not down-regulate Polycomb group (PcG) genes (**Supplementary Data Sheet 1—Figures S3G,H**) and did not induce the anti-detoxification, *GstD* genes (**Supplementary Data Sheet 1—Figure S3I**). Further analysis of offspring of bacterial-depleted flies revealed that the difference between parents and their offspring with respect to the delay in development is not observed with respect to the other heritable phenotypes. Specifically, depletion of bacteria in F1 did not lead to induction of *neoGFP* expression in the larval gut of the offspring (**Supplementary Data Sheet 1—Figures S4B,C**), or to wing deformations in *Hsp70::neoGFP* offspring of bacterial depleted parents (**Supplementary Data Sheet 1—Figure S4D**).

Altogether, this shows that the transgenerational effect of bacterial depletion can account for the inheritance of the delay in development, but not for the other transgenerationally heritable phenotypes of G418 exposure.

THE TRANSGENERATIONAL INHERITANCE OF THE DELAY IN DEVELOPMENT IS PREVENTED BY REINTRODUCTION OF THE DEPLETED *ACETOBACTER* SPECIES

To test if reintroduction of depleted bacteria can prevent the inheritance of the delay in development following parental exposure to G418, we isolated bacterial colonies from non-exposed flies. We then tested the ability of bacteria from these colonies to prevent the inheritance in non-exposed offspring of G418 exposed (F1) *hairy::neoGFP* parents. Supplementing the fly food with bacteria from a single colony (*Colony 1*, **Supplementary Data Sheet 1—Figure S5D**) sufficed to prevent the inheritance of the delay in development (**Figure 2A**). Analysis of the 16S sequence of the ribosomal RNA gene of *Colony 1* revealed similarity to 4 *Acetobacter* spp. (**Figure 2B**): *A. aceti* (99.5% similarity to *Colony 1*), *A. cibinongensis* (97.6% similarity), *A. pomorum* (96.6% similarity), and *A. tropicalis* (97.2% similarity). However, the non-commensal *Acetobacter* spp., *A. aceti*, *A. pomorum*, *A. tropicalis* and *A. cibinongensis* (all obtained from the DSMZ stock collection) could not prevent the inheritance of the delay in development (**Figures 2C,D**). Thus, the prevention of the inheritance of the delay following parental exposure to G418 required commensal *Acetobacter* species.

Notably, prevention of the inheritance of the delay was also observed in various natural, mating-dependent, or independent settings, in which exposed flies were infected with bacteria carried by non-exposed flies. Specifically, when G418 exposed males or females were crossed to naïve partners (i.e., flies lacking any history of exposure), the offspring larvae were not delayed in their development (**Figure 2E**). Additionally, when G418-exposed parents were inserted into a vial which temporarily contained naïve

flies (either males or virgin females) for 1 day prior to the insertion of the exposed parents, the offspring were no longer delayed (**Figure 2F**). Similar results were obtained when the vials were supplemented with PBS that was brought in contact with naïve flies, but not when this fluid was filtered to remove all the bacteria (**Supplementary Data Sheet 1—Figure S5A**). These findings show that the inheritance of the delay in development can be prevented by either natural or experimental exposures to bacteria from male or female flies with no history of exposure to G418.

While *Colony 1* supplementation always prevented the inheritance of the delay in offspring, it did not necessarily prevent the inheritance of induced *neoGFP* expression. Altogether, the average expression of *neoGFP* in the foregut of the offspring larvae across experiments was slightly reduced but this effect was not statistically significant (**Supplementary Data Sheet 1—Figure S5B**). Similar results were obtained in additional settings in which the food was exposed to micro-organisms carried by naïve flies (**Supplementary Data Sheet 1—Figure S5C**). Thus, gut bacteria had a significant preventive effect only on the inheritance of the delay in development.

To determine if reintroduction of commensal *Acetobacter* species can prevent the delay in development also in offspring of bacterial depleted flies that have not been exposed to G418, we added specific bacterial species to the food of the F1 larvae that hatched from dechorionated (and sterilized) eggs. Analysis of the rate of larval development in F2 showed that *Colony 1* completely rescued the delay in these offspring of bacterial-depleted flies (**Figure 2G**). As in the case of parental G418 exposure, non-commensal *A. aceti* had no effect on the delay in F2 (**Figure 2H**). Additionally, isolated bacteria (*Colony 7*) with 98.9% 16S rRNA similarity to *L. plantarum* (**Figure 2B**) and *L. plantarum* from another stock (Sharon et al., 2010), led to only partial rescue (**Figures 2G,H**). Thus, commensal *Acetobacter* species in the F1 parents are necessary for complete prevention of the delay in the F2 offspring. This indicates that lack of these *Acetobacter* species following exposure to G418 is sufficient to account for the heritable delay.

Altogether, the results show that G418 induces a delay in larval development by a direct stress to the host (**Figure 1B**, **Supplementary Data Sheet 1—Figure S3A**), but the delay is inherited in the non-exposed offspring because of a transgenerational effect of *Acetobacter* depletion in their parents (**Figures 1C,D**, **2A,B**). This indicates that the induction of the developmental delay and its transgenerational inheritance occur by different mechanisms.

To evaluate the generality of involvement of *Acetobacter* species in the inheritance of delayed development, we exposed non-transgenic, *yw* flies to other types of antibiotics, supplemented at sub lethal dosages. We established antibiotic-specific dosages that induce a 2–3-day delay in larval development (G418— 100 ug/ml, Puromycin—200 ug/ml, Ampicillin—100 ug/ml, and Ciprofloxacin—300 ng/ml). All these exposures to antibiotics induced developmental delay (**Figure 3A**), which except for the Puromycin case, persisted in subsequent generations without further exposure to the antibiotic (**Figure 3B**). Since Puromycin stood out as the only antibiotic which did not induce heritable delay, we suspected that it does not sufficiently deplete bacteria
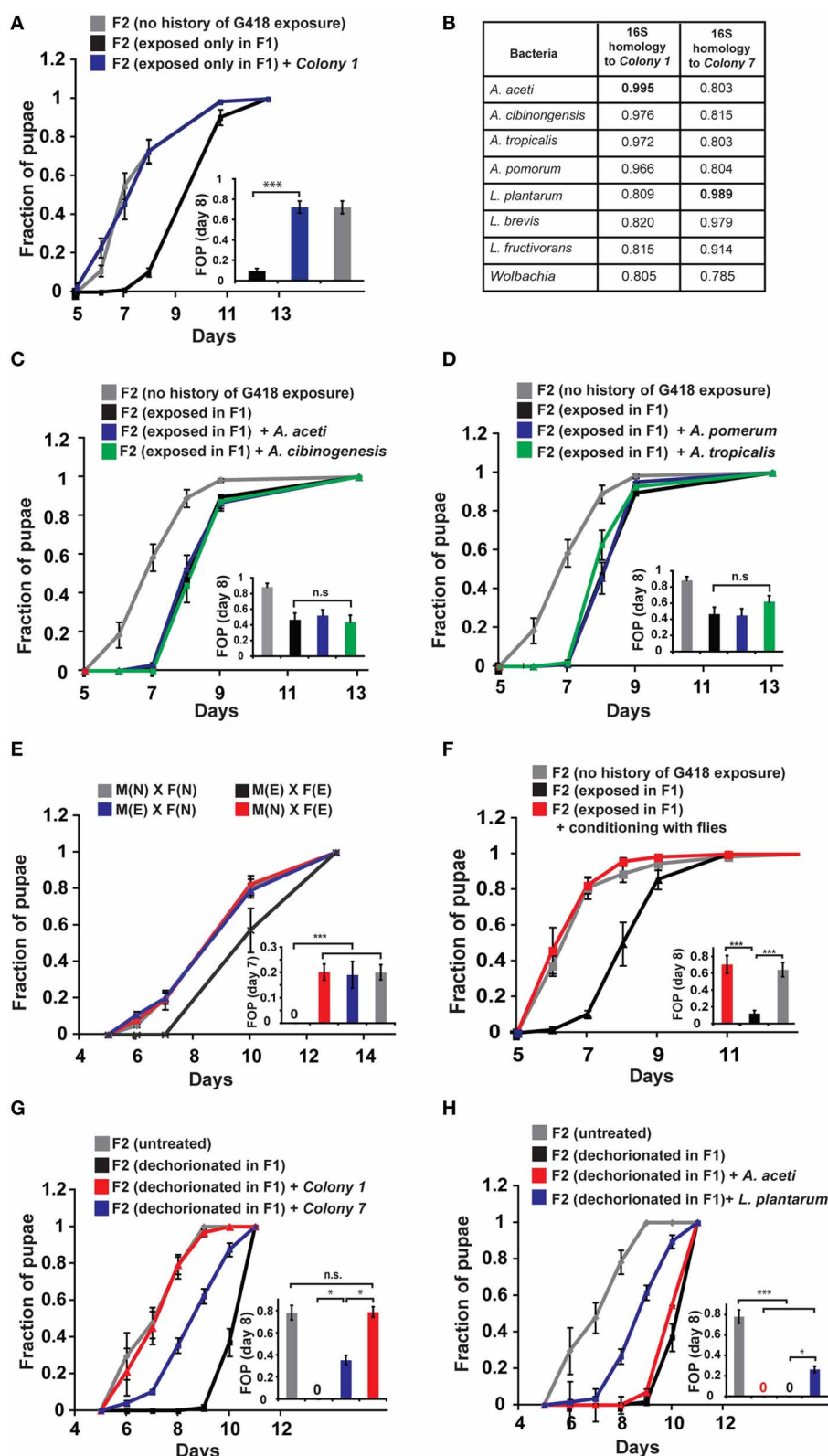
FIGURE 2 | Reintroduction of a commensal *Acetobacter* species prevents the inheritance of the delay in development. (A) Supplementing the food with bacteria from a single *Acetobacter* colony (*Colony 1*) isolated from non-exposed flies, prevented the inheritance of the delay in

non-exposed F2 offspring of G418-exposed parents. Mean fraction of pupae ± SE in 10 vials. All insets: Statistical analysis of differences between fractions of pupae (FOP) in the indicated day (7 or 8). **(B)** Analysis of 16S

(Continued)

FIGURE 2 | Continued

rRNA gene sequence of bacteria from two isolated colonies, *Colony 1* and *Colony 7*. Shown are computed similarities to various species of *Acetobacter*, *Lactobacillus*, and *Wolbachia*. **(C,D)** Supplementing fly vials with the non-commensal species, *A. aceti* or *A. cibinogenesis* **(C)**, and with *A. pomorum* or *A. tropicalis* **(D)**, did not prevent the inheritance of the delay. Mean fraction of pupae ± SE in 10 vials. **(E)** Kinetic curves of pupation of F2 larval offspring generated from crosses between G418-exposed and non-exposed *hairy::neoGFP* F1 males or virgin females (all four combinations are shown). In all cases, the F2 larvae were not exposed to G418. Delay in F2 was observed only when both parents were exposed to G418. Mean fraction of pupae ± SE in 6 vials. "M(E)"—exposed males "M(N)"—non-exposed males, "F(E)"—exposed females "F(N)"—non-exposed females. **(F)** Prevention of the inheritance of delayed development in vials that were temporarily exposed for 1 day (conditioning) to flies with no history of ancestral exposure to G418. Mean fraction of pupae ± SE in 5 vials. **(G)** Placing dechorionated and sterilized F1 eggs on food with bacteria from *Colony 1* prevented a delay in development of the F2 larvae. Bacteria from *Colony 7* led to only partial reduction in the delay. Mean fraction of pupae ± SE in at least 15 vials pooled from three biological replicates. **(H)** Same as **(G)** for food with *L. plantarum* from another stock (partial reduction) and non-commensal *A. aceti* (no effect). Mean fraction of pupae ± SE in at least 15 vials pooled from three biological replicates. *$p < 0.05$, ***$p < 0.001$ (Student's *t*-test).

which are capable of preventing the inheritance. To test this, we examined the response of *Acetobacter* spp. to exposure with the same concentration of antibiotic. We found that Ampicillin and Ciprofloxacin removed nearly all of the *Acetobacter* spp. from the larval gut and abolished the growth of *Colony 1* in suspension (**Figures 3C,D**). Exposure to Puromycin, on the other hand, spared a substantial fraction of these bacteria (~1/3 of the *Acetobacter* spp. content of untreated flies; **Figure 3C**). Additionally, Puromycin had almost no effect on *Colony 1* in suspension (**Figure 3D**). Thus, substantial *Acetobacter* content could not prevent the delay in Puromycin-exposed flies, but nonetheless prevented the inheritance of the delay in their non-exposed offspring. These results show that the transgenerational effect of *Acetobacter* depletion mediates the inheritance of delayed development in response to additional antibiotics, and that these findings are not limited to transgenic flies.

ACETOBACTER-DEPENDENT PREVENTION OF THE HERITABLE DELAY IS MEDIATED BY BACTERIAL RIBOFLAVIN (VITAMIN B2)

Since the bacteria may provide vitamins and nutrients to the larvae, we tested if vitamins could influence the heritability of the delay in development. In particular, we tested the effect of six vitamins that have been shown to be essential additives for germ-free flies reared on axenic food (Sang, 1956). Supplementing all six vitamins to the food prevented the inheritance of the delay in development (**Figure 4A**), but not the inheritance of the induced expression (**Supplementary Data Sheet 1**—Figure S6A). To test which of the six vitamins is essential for preventing the inheritance of the delay, we repeated the experiment with one vitamin excluded from each pool. Of the six vitamins tested, Riboflavin (vitamin B2) had the strongest effect on the inheritance. Pools lacking Riboflavin could not prevent the delay in the non-exposed offspring of G418-exposed parents (**Figure 4A**), indicating that Riboflavin is necessary for the activity of the full vitamin pool. Removal of Pyridoxine (vitamin B6) also compromised the activity of the pool (**Supplementary Data Sheet 1**—Figure S6B), while the remaining vitamins, Pantothenic acid, Thiamine, Nicotinic acid and Biotin, had no significant effect on the delay in offspring development (**Supplementary Data Sheet 1**—Figures S6C–F).

Supplementing the food with Riboflavin alone reduced, but did not completely eliminate the delay in development of the non-exposed offspring (**Figure 4B**). Co-supplementing Riboflavin together with Pyridoxine and Pantothenic acid increased the effect of Riboflavin alone, but only by a small amount (**Figure 4B**). To test if G418-depleted bacteria produce Riboflavin, we analyzed bacterial extracts of *Colony 1* by Mass Spectrometry. We identified a clear signature of Riboflavin in the extract of *Colony 1* (**Figure 4C**). Quantitative Liquid Chromatography Mass Spectrometry (q-LCMS) analysis further showed that the amount of Riboflavin produced by *Colony 1* is substantially larger than the amount produced by commensal Lactobacilli from *Colony 7* (**Figure 4D**). These results suggested that the ability of *Colony 1* to prevent the inheritance of the delay is mediated, at least in part, by production and secretion of Riboflavin. We tested this by supplementing the food with Roseoflavin (RoF), an inhibitor of bacterial Riboflavin synthesis (Lee et al., 2009). Riboflavin is the product of 5 genes encoded by a single polycistronic transcript that is subjected to negative feedback via a riboswitch mechanism; Riboflavin-derived Flavin mononucleotide, FMN, binds to riboswitch aptemers in the transcript, changing the conformation of the transcript into a translationally inactive form, thereby repressing the production of Riboflavin biosynthesis genes (**Figure 4E**). The FMN analog, Roseoflavin, inhibits Riboflavin production by binding to this riboswitch. We identified a range of Roseoflavin concentrations which did not significantly compromise the growth of *Colony 1* (**Supplementary Data Sheet 1**—Figure S6G). Nonetheless, supplementing the food with Roseoflavin within this range repressed the production of Riboflavin by 10-fold, as determined by quantitative LCMS (**Figure 4F**). Consistent with the hypothesized influence of Riboflavin, the inhibition of its synthesis by Roseoflavin eliminated the ability of *Colony 1* to prevent the delay in the offspring (**Figure 4G**). This outcome was completely reversed by co-supplementing the food with exogenous Riboflavin (**Figure 4G**), indicating that the effect of Roseoflavin was indeed mediated by its inhibitory influence on the synthesis of bacterial Riboflavin. Altogether, these results show that Riboflavin produced by a commensal *Acetobacter* species interferes with the inheritance of delayed development and inhibition of Riboflavin synthesis in the bacteria compromises the ability of the bacteria to prevent this inheritance.

DISCUSSION

In this work we investigated the potential involvement of the gut microbiome in the inheritance of environmentally-induced phenotypes. In particular, we tested if G418-mediated disruption of the intact microbiome could account for the inheritance of induced phenotypes in the host. We found that G418 leads

FIGURE 3 | Generality of involvement of *Acetobacter* species in the inheritance of delayed development. (A) Kinetic curves of pupation of wild-type F1 *yw* larvae exposed to sub-lethal dosages of four different antibiotics: G418 (100 ug/ml), Ampicillin (100 ug/ml), Puromycin (200 ug/ml), and Ciprofloxacin (300 ug/ml). Mean fraction of pupae ± SE in at least three biological replicates. Inset: Statistics of differences between fractions of pupae (FOP) in day 8. **(B)** Kinetic curves of pupation of non-exposed F2 offspring of F1 flies that were exposed to G418, Ampicillin, Puromycin and Ciprofloxacin. Mean fraction of pupae ± SE in at least 5 vials. Inset: Statistics of differences between fractions of pupae (FOP) in day 8. **(C)** Measured abundance of *Acetobacter* spp. in the gut of third instar larvae exposed to various antibiotics. Mean abundance normalized to non-exposed larvae ± SE in three biological replicates. **(D)** Same as **(C)** for population density of *Colony 1* in YPD media supplemented with these antibiotics. Mean OD_{600} ratio measured after 24 h and normalized to the ratio in YPD alone ± SE in four biological replicates. *$p < 0.05$, **$p < 0.005$ (Student's *t*-test).

based on host response to the modifications in the gut bacteria. In the simpler, and more expected scenario, disruption of the gut microbiome modifies host phenotypes within one generation and this response is roughly the same in the following generations (which inherit the changes in the microbiome). This scenario might be expected if the germline is not modified by the disruption of the microbiome. In this case, persistence of the microbial changes creates the phenotypes anew in each generation. In the second potential scenario, the disruption of the microbiome has a different influence on the parents and their offspring. This difference between parents and offspring could suggest an effect of microbial disruption on the parental germline.

Our bacterial removal experiments revealed a clear signature of the second scenario. This was demonstrated by the dramatic difference in pupation time between extracellular bacteria-free parents and their offspring. Our rescue experiments with specific bacterial species further showed that as long as the parents are depleted of their commensal *Acetobacter* spp., the development of their offspring is delayed compared to the development of the parents. Together with the selective depletion of *Acetobacter* spp. by G418, this inter-generational difference accounts for the inheritance of the delay in development in offspring of G418-exposed parents. This scenario of inheritance was unexpected because the induction and inheritance of the delay in development are mediated by different mechanisms. Specifically, we show that the delay in pupation time is induced in the parental generation by the direct influence of G418 on the host, whereas the delay in the offspring generation is caused by the transgenerational influence of *Acetobacter* depletion. Reaching this conclusion requires bacterial depletion and rescue without G418, thus enabling de-coupling of the influence of bacteria from the toxic influence of G418 (Ridley et al., 2013).

The difference between generations with respect to pupation time, clearly indicate that the influence of bacteria in one generation can be different than the influence in the following generation. We hypothesize that this reflects an influence of microbial disruption on the germline. While bacterial-mediated manipulation of the germline has been previously noted for germline-residing endosymbionts (Bourtzis et al., 1996; Werren, 1997; Starr and Cline, 2002; Fast et al., 2011), it has not yet been shown for extracellular gut bacteria. This does not at all negate the diverse influence of gut bacteria on the development and physiology within generation. Indeed, we found that larvae which hatched from dechorionated and sterilized eggs are very sensitive to food quality and, upon reduction in quality, exhibited a considerable delay in development already in the first generation without bacteria (**Supplementary Data Sheet 1**—Figure S2). Previous studies of gut bacteria in flies further uncovered substantial involvement of the micrbiome in the fly development and homeostasis, including the regulation of larval growth (Shin et al., 2011; Storelli et al., 2011), immune activation (Charroux and Royet, 2012), lifespan determination (Brummel et al., 2004; Ben-Yosef et al., 2008), fecundity (Ben-Yosef et al., 2010), and even mating choice (Sharon et al., 2010). Many past studies, however, assumed that the effect of gut bacteria is the same in each generation. Accordingly, they examined the effect of bacterial removal in flies that had been kept without bacteria for
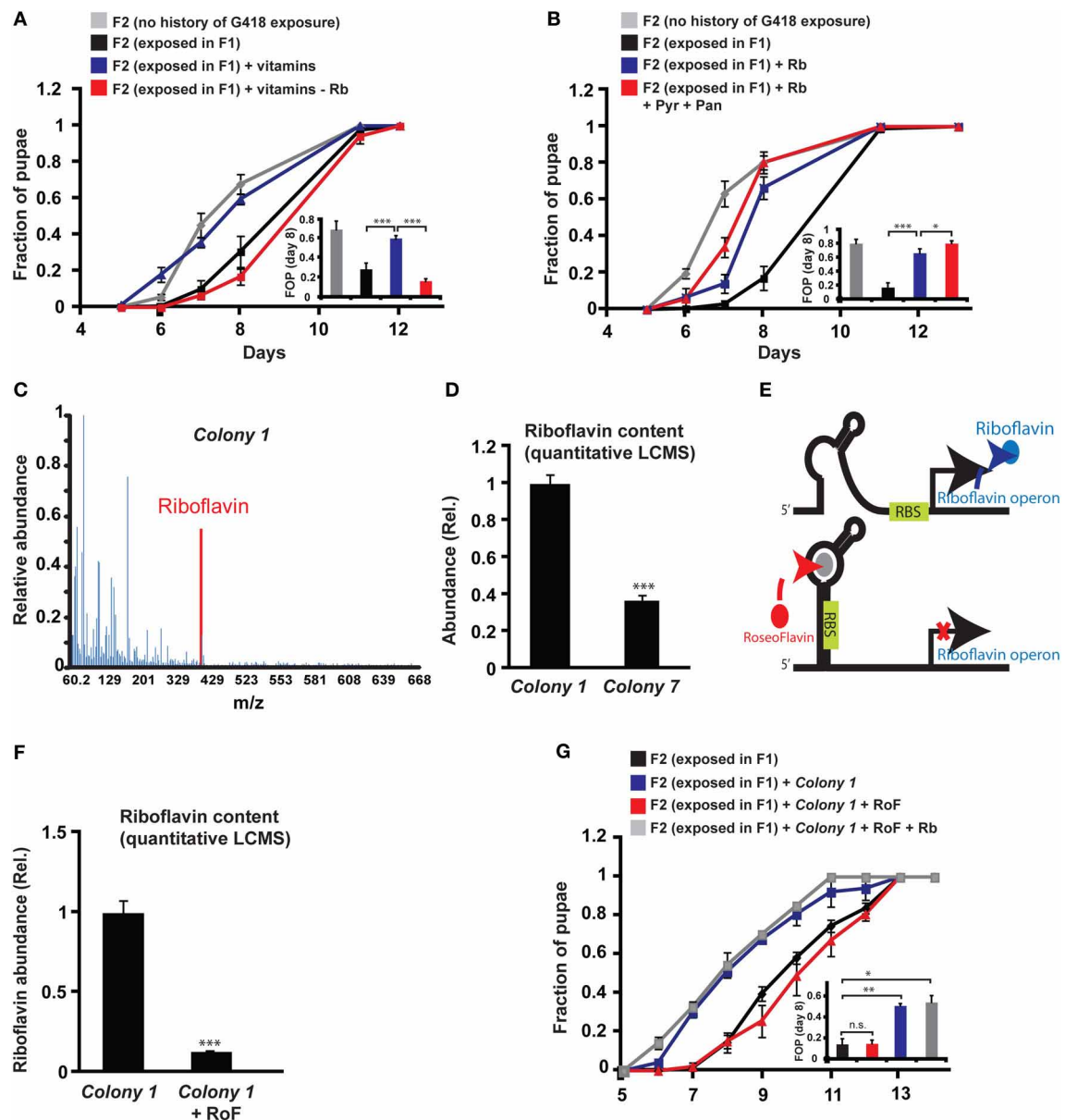
to selective changes in the microbial gut composition and that these changes persist in the next generation. This inheritance of microbial disruption raises two potential scenarios which may support inheritance of induced phenotypes. Both scenarios are

FIGURE 4 | Riboflavin produced by *Colony 1* is necessary for *Colony 1*-mediated prevention of the inheritance of the delay in development. (A) Effect of exogenous vitamins on the inheritance of delay in development. G418-exposed F1 *hairy::neoGFP* flies were allowed to lay eggs on G418-free food, supplemented with a full set of six vitamins (Riboflavin, Pyridoxine, Pantothenic acid, Thiamine, Nicotinic acid, Biotin) or with a set from which Riboflavin was excluded. Mean fraction of pupae ± SE in 10 vials. Inset: Statistical analysis of differences between fractions of pupae (FOP) in day 8. **(B)** Same as **(A)** with Riboflavin (Rb) alone (instead of the vitamin pool) or with Riboflavin together with Pyridoxine (Pyr) and Pantothenic acid (Pan). Mean fraction of pupae ± SE in 5 vials. Inset: Statistics of differences between

fractions of pupae (FOP) in day 8. **(C)** Mass spectrum of *Colony 1* extract indicating the presence of Riboflavin. Spectrum normalized to the highest peak. **(D)** Quantitative liquid chromatography Mass Spectrometry (q-LCMS) measurements of Riboflavin in extracts of *Colony 1* and *Colony 7*. Mean abundance relative to *Colony 1* ± SE in six biological replicates. **(E)** Schematic diagram of the current model of bacterial Riboflavin riboswitch and its inhibition by Roseoflavin. **(F)** Quantitative LCMS measurements of Riboflavin abundance in *Colony 1* with or without Roseoflavin (RoF). Mean abundance ± SE in three biological replicates. **(G)** Same as **(A)** with Roseoflavin (red) and Roseoflavin together with Riboflavin (blue). Mean fraction of pupae ± SE in five vials. *$p < 0.05$, **$p < 0.005$, ***$p < 0.001$ (Student's *t*-test).

multiple generations prior to the experimental analysis. Under these settings, the transgenerational effect of bacterial removal is influencing both parents and offspring thus, eliminating the ability to determine which part of the response reflects an influence

within generation and what part is due to the transgenerational effect. The ability to resolve the different scenarios is further reduced by treatments with antibiotics, which are often used to maintain bacteria-free conditions after dechorionation.

We have previously shown that G418 toxicity induces additional phenotypes beyond the delay in development. These include induced expression of *neoGFP* in gut regions of *hairy::neoGFP* and *drm::neoGFP* larvae, morphological abnormalities in the wings of *Hsp70::neoGFP* flies, reduction in fly size in G418-exposed *drm::neoGFP* larvae, down-regulation of a number of *Polycomb* group genes in the proventriculus of *hairy::neoGFP* larvae, and up-regulation of detoxification (*GstD*) genes in the same region (Stern et al., 2012). Here we show that removal of gut bacteria without exposure to G418 does not induce these other G418 reported phenotypes. This indicates that these phenotypes are caused by a direct effect of the toxic stress of G418 on the fly's tissues, and not by its effect on the fly's microbiome. Additionally, unlike the delay in pupation time, the other heritable phenotypes of G418 exposure (induced expression of *neoGFP* and wing abnormalities) were not observed in offspring of bacterial depleted parents. Thus, the inter-generational difference between parents and offspring did not extend to other heritable phenotypes reported in Stern et al. (2012). Altogether, this indicates that the inheritance of these phenotypes following parental exposure to G418 reflects additional mechanisms beyond those of host–microbe interactions. Recent work in mice also provided evidence for multiple mechanisms supporting the inheritance of different phenotypes in a given experimental setting (Padmanabhan et al., 2013). Having more than one mechanism of transgenerational inheritance in a single experimental setting (e.g., exposure to G418), and in different animals, suggests that non-Mendelian inheritance might be more prevalent than often assumed. Additionally, by uncovering the contribution of microbial disruption to the inheritance of a stress-induced phenotype, our findings expand the set of factors which could support transgenerational inheritance of induced responses.

The transgenerational effect of bacterial removal on pupation time was completely rescued by supplementing the parental food with bacteria from a commensal *Acetobacter* species (*Colony 1*). On the other hand, commensal *Lactobacillus* spp. (*Colony 7*, and *L. plantarum*) led to only partial rescue. The complete rescue by *Acetobacter* species indicates that their absence is sufficient for causing a delay in offspring generations. Additionally, this rescue of the transgenerational effect is consistent with the *Acetobacter*-mediated prevention of the developmental delay in offspring of G418-exposed parents. This prevention was observed in every experimental setting in which G418-free food was exposed to *Acetobacter* spp., including natural forms of contact with bacteria from non-exposed (male or female) parents. Altogether, these results indicate that depletion of *Acetobacter* spp. in both parents is necessary and sufficient for the inheritance of the delay in development. Since the bacterial composition, including *Acetobacter* content, may be influenced by a diverse set of environmental agents (and not just by antibiotics), the microbiome likely mediates multi-generational influences on growth kinetics in response to a variety of natural environmental settings.

The mechanistic basis of bacterial–host interactions in flies has been previously focused on the involvement of Insulin signaling (case of *A. pomorum* Shin et al., 2011) and the TOR pathway (case of *L. plantarum* Storelli et al., 2011). Here we identified Riboflavin as one of the mediators of *Acetobacter*-dependent prevention of the delay in larval development. We found that a commensal *Acetobacter* species produces relatively large amount of Riboflavin. Supplementing the food with exogenous Riboflavin reduced the delay in non-exposed larval offspring of G418-exposed parents. Conversely, co-supplementing the food with *Acetobacter* species and the Riboflavin inhibitor, Roseoflavin, compromised the production of Riboflavin by the *Acetobacter* species and inhibited its ability to prevent the delay in development. This inhibition was relieved by addition of exogenous Riboflavin. Collectively, these results indicate that Riboflavin is necessary for the *Acetobacter*-dependent prevention of the inheritance of the delay in development.

The contribution of Riboflavin to the prevention of the delay in offspring of bacterial depleted parents was also consistent with the smaller rescue ability of commensal *Lactobacillus* spp., *Colony 7*, and *L. plantarum*. Indeed, quantitative LCMS-based analysis of *Colony 7* revealed substantially lower amounts of Riboflavin compared to the commensal *Acetobacter* species (*Colony 1*).

The influence of Riboflavin is also consistent with previous studies which showed that vitamins affect the rate of development of germ-free larvae when the food conditions are also axenic (Tatum, 1939; Sang, 1956). In our settings of non-axenic rich diet, Riboflavin did not influence the growth of non-manipulated flies. Nonetheless, it expedited the development of non-exposed larval offspring of G418-exposed parents. This suggests that the history of exposure to G418 creates an additional requirement beyond that provided normally in the fly's diet.

Deciphering downstream mechanisms of Riboflavin action in host tissue is highly confounded by the broad influence of Riboflavin on many enzymes. Riboflavin is a source of FAD and FMN derivatives serving as cofactors in the oxidative metabolism of carbohydrate, amino acids, and fatty acids. These cofactors are involved in many functions and can have pleiotropic effects on development and physiology (reviewed in Powers, 2003). Similarly to other vitamin-derived cofactors, deficiency in Riboflavin supply (e.g., in the case of bacterial depletion) likely impacts multiple molecular pathways in host tissues. Future work will determine if the effect of Riboflavin deficiency following exposure to G418 is primarily mediated by a single host pathway or, alternatively, by a coordinated action of multiple mechanisms.

Non-Mendelian Transgenerational phenomena in multiple organisms can involve a variety of epigenetic mechanisms (reviewed in Jablonka and Raz, 2009; Daxinger and Whitelaw, 2012; Jablonka, 2012; Lim and Brunet, 2013). In this work we describe a previously unrecognized scenario of transgenerational inheritance mediated by microbial-based modifications in parents which, in turn, affect the development of their progeny. While we demonstrate this in flies, it might be more broadly applicable to many organisms which maintain complex interactions with their commensal bacteria. Indeed, commensal bacteria are an integral part of many, if not all other known animals. These bacteria can be viewed as a "distributed organ" which can be readily affected by the environment on rapid time scales (within and across several generations). The changes in the microbiome might be in species composition as well as in bacterial gene sequence (i.e., following bacterial adaptation to the stressful environment). These changes feedback on host development and physiology, and

may be responsible for diverse effects which could have been previously attributed to host-intrinsic factors. The heritability of the changes in the microbiome further provides potential infrastructure for influencing many generations (Bakula, 1969; Rosenberg and Zilber-Rosenberg, 2011). As such, the involvement of commensal bacteria in non-Mendelian, multi-generational responses to the environment might be considerably widespread. Thus, commensal bacteria are major mediators of gene-environment interactions with potential transgenerational and evolutionary implications that await further exploration. The diverse genetic tools available in *D. melanogaster*, together with the relatively low complexity of its microbiome and the ease with which this microbiome can be manipulated, make the fly a powerful model system for this exploration.

MATERIALS AND METHODS

DROSOPHILA STOCKS

drm-GAL4, hairy-GAL4, and Hsp70-GAL4 lines were obtained from the Bloomington Stock Center. yw line was obtained from the lab of Dr. Eli Arama (Weizmann Institute of Science), *Canton-S* and *Oregon-R* lines were obtained from the lab of Prof. Adi Salzberg (The technion). The UAS-*neoGFP* line was generated as described in Stern et al. (2012).

FOOD PREPARATION

Standard cornmeal food (Bloomington Stock Center recipe, http://flystocks.bio.indiana.edu/Fly_Work/media-recipes/molassesfood.htm) was heated to ~60°C and mixed (1:100 in volume) with antibiotics to reach the desired final concentration. For the experiments involving the transgeneic, *hairy::neoGFP*, *drm::neoGFP* and *Hsp70::neoGFP* lines we used G418 at final concentration of 400 μg/ml. For wild-type strains (*yw, Canton-S, Oregon-R*), final concentrations of antibiotics were as follows: G418 (100 μg/ml), Ampicillin (100 μg/ml), Puromycin (200 μg/ml), and Ciprofloxacin (300 μg/ml). The food was divided into standard 25 × 95 mm Drosophila vials (cat# 51–0500, Biologix, USA). Vials with 10 ml of food were left overnight at room temperature and stored at 4°C for up to 2 weeks prior to usage. For non-exposed conditions, the food was prepared in the same way but without the antibiotics.

MEASURING DURATION OF LARVAL DEVELOPMENT

Duration of larval development was measured by counting the number of pupae in each vial daily. The integrated number of pupae formed prior to each inspection time was normalized to the total number of pupae formed in the vial at the end of the experiment.

ISOLATION AND GROWTH OF COMMENSAL BACTERIA

Ten male flies were shaken in 1 ml of PBS buffer at room temperature for 30 min. 100 ml of this fluid were serially diluted, spread on YPD agar plates and colonies were grown at room temperature for 3 days. Several colonies were picked, underwent three additional rounds of isolation, grown overnight at 30°C in liquid YPD and stocked in 35% glycerol at −80°C.

TRANSGENERATIONAL EXPERIMENTS

Three males and two females were crossed and allowed to lay eggs for 3 days in vials with or without G418 (or other antibiotics).

Unless specifically indicated, all F1 experiments with transgenic flies were done using lines heterozygous for the *GAL4* driver and the UAS-*neoGFP* transgenes, as described in Stern et al. (2012). F1 flies developed from these eggs were collected after 19–20 days from the start of the experiment (4–7 day old adults) and the same number of males and females were crossed again in vials without antibiotics. For subsequent generations in experiments involving the transgenic lines, only fluorescent adults were crossed (i.e., flies carrying at least one *GAL4* and one UAS-*neoGFP* transgenes).

TRANSGENERATIONAL EXPERIMENTS WITH PRIOR REMOVAL OF BACTERIA FROM FLIES

Dechorionated and sterilized embryos were transferred to vials with or without G418. For analyses of F2 and F3 bacterial depleted flies, F1 flies developed from these eggs were collected after 19–20 days from the start of the experiment (4–7 day old adults) and the same number of males and females were crossed again in vials without G418.

STATISTICAL ANALYSES

All statistical tests were performed using the MATLAB software (MathWorks). One-sided student's *t*-test was used for evaluating the statistical significance of mean values. Analyses of functional enrichments in groups of up- and down-regulated genes in the microarray experiments were done using the DAVID web tool (http://david.abcc.ncifcrf.gov/, Huang da et al., 2009a,b).

DETECTION AND QUANTIFICATION OF *neoGFP* EXPRESSION

Quantification of *neoGFP* intensity in the proventriculus: For each proventriculus of an individual third instar larva, an image was taken using a fluorescent stereoscope (Leica MZ16F) with constant imaging parameters. Image analysis for measuring the induction of *neoGFP* in the proventriculi of *hairy::neoGFP* larvae was performed as follows: GFP intensity was measured in the anterior half of the proventriculus by computing the average pixel intensity of GFP. Then, for each image the background autofluorescence of the tissue (evaluated in the dark area in the middle of the proventriculus) was subtracted. Image analysis and computation was performed using a custom MATLAB (MathWorks) script.

Quantification of *neoGFP* intensity of expression in the midgut: images were taken from each larva as described above for the proventriculus tissue. Quantification of *neoGFP* levels in the midgut of *drm::neoGFP* larvae was performed by calculating the average pixels intensity in the region of the midgut with *GFP* intensity above background. Measurements from non-exposed larvae were performed in a midgut tissue of about the same size and location. Average intensity was corrected for autofluorescence in the midgut (primarily due to food inside the gut) by subtracting the average intensity measured in an adjacent (non-induced) midgut area.

DETECTION AND QUANTIFICATIONS OF PUPA AND ADULT PHENOTYPES

Images of adult males or females were taken using a fluorescent stereoscope (Leica MZ16F). Length of adult *drm::neoGFP* flies was calculated by measuring (using the NIS-Elements software, Nikon) the number of pixels along a line drawn from head to

genital of each fly. The number of pixels was then converted to millimeters.

Adult *Hsp70::neoGFP* flies with wing abnormalities were scored by eye.

CONDITIONING EXPERIMENTS

Conditioning experiments with flies were performed by insertion of 5 adult male flies into a regular fly vial. After 24 h, these flies were discarded and replaced by flies that were used for mating. Conditioning with wash fluid was performed by washing 10 adult flies in 1 ml of PBS for 30 min. 100 ml of the wash fluid was applied to the top of the fly food. After 24 h these vials were used for the mating experiments.

ANALYSIS OF 16S rRNA SEQUENCE OF *COLONY 1* AND *COLONY 7*

The 16S gene was PCR-amplified from bacterial DNA using universal primers 8F (AGAGTTTGATCCTGGCTCAG) and 1492R (GGTTACCTTGTTACGACTT) (Weisburg et al., 1991). The PCR product was cloned into PGMT plasmid, sequenced and compared to the 16S database (Green Genes database http://greengenes.lbl.gov/) using the BLAST algorithm. The resulted closest matches—*A. aceti* (AJ419840.1), *A. pomorum* (AJ001632.1), *A. tropicalis* (AJ419842.1), and *A. cibinongensis* (AB052711.1), along with the previously reported commensal bacteria *L. plantarum* (JQ411248.1), *L. brevis* (JQ236623.1), *L. fructivorans* (AB680532.1), and Wolbachia (DQ412083.1) were used for multiple alignment using the T-Coffee web resource.

BACTERIAL STRAINS FROM COMMERCIAL STOCKS

A. aceti, *A. tropicalis, A.pomorum, and A.cibinongensis* were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ, http://old.dsmz.de/identification/main.php?contentleft_id=2) and were grown in YPD or LB media.

SAMPLE PREPARATION FOR SCANNING EM

Bacteria were grown to mid-log in YPD medium at 37°C and were fixed with 2.5% gluteraldhyde and 2.5% PFA in 0.2 M cacodylate buffer. The cells were attached to PLL (poly-L-lysine) coated silica chip, dehydrated in an ethanol series (50, 70, 96, and 100%) and critical-point dried using BALTEC dryer. The dry samples were coated with gold-palladium for 4 min and visualized in a high-resolution Ultra 55 SEM (Zeiss).

SENSITIVITY OF *COLONY 1* TO DIFFERENT ANTIBIOTICS

A single colony from the *Colony 1* stock was isolated from YPD plates and grown to $OD_{600} = 1$ in YPD at 30°C with continuous shaking. 10 μl of this were inoculated into 2 ml of YPD, and different concentrations of antibiotics were added in triplicates. OD_{600} was measured after 24 h of incubation at 30°C. OD measurements were normalized to antibiotic-free sample.

BACTERIAL REMOVAL BY EGG DECHORIONATION

Flies were crossed and allowed to lay eggs for 2–4 h. Eggs were collected and dechorionated for 2 min in 2.7% sodium hypochlorite (2-fold diluted bleach), washed twice in 70% ethanol and then twice with sterile distilled water as previously described (Brummel et al., 2004).

BACTERIAL REINTRODUCTION EXPERIMENTS

For the bacterial reintroduction experiments we used bacteria from single colonies (either *Colony 1* or *Colony 7*) isolated from naive flies, a strain of commensal *L. plantarum* obtained from the Rosenberg lab (Tel-Aviv University), and bacteria from DSMZ collection (*A. aceti*, *A. pomorum*, *A. tropicalis*, and *A. cibinongensis*). The bacteria were grown (with shaking) overnight to an OD_{600} of 1–2 at 30°C in 3 ml YPD. Bacteria were diluted to OD 0.1 in YPD, and 100 μl were added to each vial before the transfer of untreated or dechorionated eggs to these vials.

For analysis of the F2 generation, 5–10 days old adult flies were collected from these vials, and 3 males and 2 females were crossed in a new (untreated) vial for 2 days.

SUPPLEMENTING FLY FOOD WITH VITAMINS AND ROSEOFLAVIN

Standard cornmeal food (Bloomington Stock Center recipe, http://flystocks.bio.indiana.edu/Fly_Work/media-recipes/molassesfood.htm) was heated to ∼60°C and mixed with Thiamine ($2 \cdot 10^{-3}$ mg/ml final), Riboflavin ($1 \cdot 10^{-2}$ mg/ml), Nicotinic Acid ($1.2 \cdot 10^{-2}$ mg/ml), Calcium Pantothenate ($1.6 \cdot 10^{-2}$ mg/ml), Pyridoxine ($2.5 \cdot 10^{-3}$ mg/ml), Biotin ($1.6 \cdot 10^{-4}$ mg/ml), Roseoflavin (100 μM), or combinations thereof dissolved in water. The mix was split to standard 25×95 mm *Drosophila* vials (cat# 51–0500, Biologix, USA). Vials with 10 ml of mixed food were left overnight at room temperature before storage at 4°C for up to 2 weeks prior to usage.

mRNA EXTRACTION AND ANALYSIS

Proventriculi were dissected from ∼50 *hairy::neoGFP* third instar larvae for each condition, and were kept during dissection in ice cold PBS. Total RNA was extracted and purified using the RNeasy MinElute Cleanup Kit (QIAGEN). Genome-wide mRNA expression levels in each sample were measured using Affymetrix drosophila 2.0 array using standard Affymetrix protocols. Arrays were normalized with the RMA algorithm using the Expression Console software (Affymetrix).

Results for specific genes were verified using real-time quantitative PCR as follows: mRNA was converted to cDNA using high-capacity Reverse Transcription kit (Ambion). Transcript levels were measured using real-time qPCR on a 7900HT Fast Real-Time PCR Machine using Power SYBR green PCR master mix (Applied Biosystems). Specific primers used in this study are listed below:

Gene	F primer	R primer
Act5C	CCCTCGTTCTTGGGAATGG	CGGTGTTGGCATACAGATCCT
Pc	AAATCATCCAAAAGCGCGTTA	CCGGTTCCCAGGTGTTGTAG
ph-p	AATTTTGGCCATGACCTCGAT	ACAGCGGTGCTTGTCACAGA
Scm	GTTTGCCCTGGAAGGAGATGT	TCATCCTTCATTCGCATTGG
pho	ACGGTCCTCGAGTCCATGTT	CACCGGTGTGAACCAACTGA

MASS SPECTROMETRY ANALYSIS

Mass spectra analysis of colony 1

Colony 1 was cultured overnight (with shaking) at 30°C in 100 ml YPD to OD_{600} between 1.5 and 2. Bacteria were

centrifuged at 4°C for 15 min at 4000 g and washed twice with 20 ml of water at the same conditions. The bacteria were re-suspended in 10 ml of water and sonicated on ice using microtip at 40% power (30 s ON followed by 50 s OFF for a total of 15 min) on a Sonics Vibra-Cell machine (Sonics and Materials, Inc.). The solution was then centrifuged (20,000 g) at 4°C for 20 min and the supernatant was passed through 0.22 micron filter. Pure vitamins [Riboflavin—Sigma-Aldrich R4500–25; Calcium D−(+)-pantothenate—Santa-Cruz SC-202515; Pyridoxine hydrochloride—Sigma-Aldrich P9755–25G] were reconstituted from powder to serve as references for the bacterial samples. All samples were analyzed on an Electron Spray machine in the Mass Spectrometry core facility unit at the Weizmann Institute of Science, Rehovoet, Israel.

Quantitative measurements of riboflavin content

Bacteria were cultured (with shaking) overnight to OD_{600} between 1 and 2 at 30°C in 20 ml YPD. For experiments with Riboflavin inhibition, Roseoflavin (0215425225, ENCO) was added to the YPD to a final concentration of 100 μM and the bacteria were allowed to grow. It was then diluted to OD_{600} 1, and 10 ml were centrifuged at 4°C for 15 min at 4000 g and washed twice with 10 ml of ice-cold DDW. The bacteria were re-suspended in 1 ml of methanol and sonicated on ice using "Bioruptor® Standard sonication device" on high setting (30 s ON followed by 30 s OFF for a total of 30 min). The solution was then centrifuged (20,000 g) at 4°C for 20 min and the supernatant was passed through 0.22 micron filter. It was completely evaporated with Nitrogen gas and re-suspended in 100 μl methanol. Pure vitamins (Riboflavin—R4500–25 and Thiamine hydrochloride—T4625–25G, Sigma-Aldrich) were reconstituted from powder to concentrations of 50 μg/ml to serve as references for the bacterial samples. Three biological repeats of each sample were subjected to LC-MS analysis.

Analysis was performed using an Ultra Performance Liquid Chromatography—Mass Spectrometry (UPLC)-MS/MS triple quadrupole instrument (Xevo TQ MS, Waters) equipped with an electrospray ion source and operated in positive ion mode. MassLynx and TargetLynx software (v.4.1, Waters) were applied for the acquisition and analysis of data. Chromatographic separation was done on a 100× 2.1-mm inside diameter, 1.7-μm UPLC BEH C18 column (Acquity, Waters) with mobile phases A (10 mM ammonium formate buffer, pH = 6.26) and B (acetonitrile) at a flow rate of 0.3 ml/min and column temperature 40°C. A gradient was as follows: 0–4 min linear gradient from 0 till 100% B, 4.0–4.5 min hold at 100% B, 4.5–5.0 min return to 0% B and equilibration at 0% B for 2 min. Injection volume of 5 μL was used.

For MS, argon was used as the collision gas with flow 0.22 ml/min. Cone voltage was 40 V, the capillary was set to 1.15 kV, source temperature—150°C, desolvation temperature—550°C, desolvation gas flow—800 L/min, cone gas flow—50 L/min. Riboflavin was detected using multiple reaction monitoring (MRM) applying the following parameters: transition 377.2 > 243.1 (collision energy—22 eV, was used for quantification), transition 377.2 > 198.1 (collision energy—36 eV).

BACTERIAL DNA-SEQ EXPERIMENTS

Ten guts of *hairy*::*neoGFP* third instar larvae that were developed in vials with or without G418, were dissected from each sample and pooled together. Bacterial DNA was extracted using the "chemagic DNA bacteria Kit" (Chemagen). The sequencing of the 16S rDNA was performed as described by Amir et al. (2014). Briefly, DNA was PCR amplified using six pairs of non-overalpping 16S primers. Each pair amplified relatively short (∼200 bp) region that is conserved among a subset of known bacteria. Altogether, the six regions amplify over 90% of the Greengenes database (version dated 2010), having an amplicon of approximately 1200 base pairs in total. Additionally, the primers included Illumina sequencing barcodes and therefore the library preparation stage was not necessary. PCR products were cleaned (Promega, Fitchburg, WI) and DNA concentration was measured using Qubit (Life Technologies). All samples were pooled in equimolar ratios and sequenced on a single lane of an Illumina HiSeq2000 sequencer using 100 nt paired-end reads. The number of reads per sample following quality filtering was about $20 \cdot 10^6$.

Microbial profiling based on the 16S sequences was performed using a novel framework which allows massively parallel sequencing (MPS) of a large genomic region, thus increasing the resulting phylogenetic resolution while reducing the number of false positively detected bacteria. Detailed description of the microbial reconstruction methodology is provided by Amir et al. (2014). In brief, we formulated a statistical model for generating the observed reads given a pre-defined database of about 450,000 known 16S sequences (Greengenes, version dated 2010), and solved an optimization problem whose output is the identity and frequency of each sequence. Each of the millions of reads measured by MPS, together with all "absent" reads that were not found, set constraints which enable "zooming in" on the correct species present in the mixture. Since reads are much shorter than the amplified region and contain errors, they are often shared by many bacterial sequences in the database. Nonetheless, each read provides evidence in support of the existence of the "correct" bacteria in a probabilistic way. The method integrates the statistical evidence from all reads to infer the frequency of each sequence in the database.

MEASURING BACTERIAL CONTENT BY PCR

For primer specificity tests, we used defined bacteria obtained from the DSMZ stock collection. Colonies were isolated and grown over-night in YPD medium. Wolbachia was omitted from these tests due to lack of culturing conditions. To measure the levels of bacteria in the gut, 7–10 third instar larvae were collected. The gut of the larvae was dissected and pooled. Bacterial genomic DNA was extracted using "chemagic DNA Bacteria" kit (Chemagen) and 5 ng of purified DNA was used per qPCR reaction. All reactions were performed in triplicates on a qPCR machine (Applied Biosystems 7900HT Fast Real-Time PCR System, Life Technologies Corporation) using SYBRGreen in 384 well-plates. The species-specific primers used are: aceto_rt_1_f (TAG TGG CGG ACG GGT GAG TA), aceto_rt_1_r (AAT CAA ACG CAG GCT CCT CC), lacto_rt_2_f (AGG TAA CGG CTC ACC ATG GC), lacto_rt_2_r (ATT CCC TAC TGC TGC CTC CC), wolb_rt_2_f (CAA TGG TGG CTA CAA TGG GC),

wolb_rt_2_r (GTA TTC ACC GTG GCG TGC TG). DNA content of the Drosophila *Actin* gene was used to normalize the bacterial content. *Actin* primers used: dros_rt_1_f (GGA AAC CAC GCA AAT TCT CAG T), dros_rt_1_r (CGA CAA CCA GAG CAG CAA CTT).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00027/abstract

Supplementary Data Sheet 1 | Supplementary Figures S1–S6 (including captions).

Supplementary Data Sheet 2 | Summary of 16S deep-sequencing results of gut samples of G418 exposed and non-exposed larvae. Samples were extracted from third instar *hairy::neoGFP* larvae. Data is represented as the normalized percent of reads mapped to a Greengenes Database annotation (for bacterial abundance >1%). The 16S ribosomal RNA sequence of each Greengenes annotation was associated with known bacteria by calculating the respective sequence homology using RDP classification (http://rdp.cme.msu.edu/tutorials/classifier/RDPtutorial_RDP-CLASSIFIER.html).

Supplementary Data Sheet 3 | Summary of microarray analyses of mRNA profiles in the proventriculus of bacterial depleted and untreated flies. mRNA was extracted from the proventriculi of *hairy::neoGFP* third instar larvae that hatched from dechorionated and sterilized eggs or from untreated eggs. Genome-wide profiles were analyzed using affymetrix microarrays. Data represented as log2 fold-change difference between bacterial depleted and untreated larvae in two biological replicates.

Supplementary Data Sheet 4 | Functional enrichments in sets of proventriculus genes which respond to removal of extracellular bacteria. Analyses of functional enrichments in groups of up- and down-regulated genes (>1.5-fold) performed using the DAVID web tool (http://david.abcc.ncifcrf.gov/).

REFERENCES

Amir, A., Zeisel, A., Zuk, O., Elgart, M., Stern, S., Shamir, O., et al. (2014). High resolution microbial community reconstruction by integrating short reads from multiple 16S rRNA regions. *Nucleic Acids Res.* 41, e205. doi: 10.1093/nar/gkt1070

Anway, M. D., Cupp, A. S., Uzumcu, M., and Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308, 1466–1469. doi: 10.1126/science.1108190

Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, M. P., Cording, A. C., et al. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 150, 88–99. doi: 10.1016/j.cell.2012.06.018

Bakula, M. (1969). The persistence of a microbial flora during postembryogenesis of *Drosophila melanogaster*. *J. Invertebr. Pathol.* 14, 365–374. doi: 10.1016/0022-2011(69)90163-3

Ben-Yosef, M., Aharon, Y., Jurkevitch, E., and Yuval, B. (2010). Give us the tools and we will do the job: symbiotic bacteria affect olive fly fitness in a diet-dependent fashion. *Proc. Biol. Sci.* 277, 1545–1552. doi: 10.1098/rspb.2009.2102

Ben-Yosef, M., Behar, A., Jurkevitch, E., and Yuval, B. (2008). Bacteria–diet interactions affect longevity in the medfly – *Ceratitis capitata*. *J. Appl. Entomol.* 132, 690–669. doi: 10.1111/j.1439-0418.2008.01330.x

Bourtzis, K., Nirgianaki, A., Markakis, G., and Savakis, C. (1996). Wolbachia infection and cytoplasmic incompatibility in Drosophila species. *Genetics* 144, 1063–1073.

Brummel, T., Ching, A., Seroude, L., Simon, A. F., and Benzer, S. (2004). Drosophila lifespan enhancement by exogenous bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12974–12979. doi: 10.1073/pnas.0405207101

Buchon, N., Broderick, N. A., and Lemaitre, B. (2013). Gut homeostasis in a microbial world: insights from *Drosophila melanogaster*. *Nat. Rev. Microbiol.* 11, 615–626. doi: 10.1038/nrmicro3074

Buckley, B. A., Burkhart, K. B., Gu, S. G., Spracklin, G., Kershner, A., Fritz, H., et al. (2012). A nuclear argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* 489, 447–451. doi: 10.1038/nature11352

Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., et al. (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 143, 1084–1096. doi: 10.1016/j.cell.2010.12.008

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93, 505–518. doi: 10.1016/S0092-8674(00)81181-2

Charroux, B., and Royet, J. (2012). Gut-microbiota interactions in non-mammals: what can we learn from Drosophila? *Semin. Immunol.* 24, 17–24. doi: 10.1016/j.smim.2011.11.003

Crews, D., Gillette, R., Scarpino, S. V., Manikkam, M., Savenkova, M. I., and Skinner, M. K. (2012). Epigenetic transgenerational inheritance of altered stress responses. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9143–9148. doi: 10.1073/pnas.1118514109

Cropley, J. E., Suter, C. M., Beckman, K. B., and Martin, D. I. K. (2006). Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17308–17312. doi: 10.1073/pnas.0607090103

Daxinger, L., and Whitelaw, E. (2012). Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat. Rev. Genet.* 13, 153–162. doi: 10.1038/nrg3188

Dobson, S. L., Bourtzis, K., Braig, H. R., Jones, B. F., Zhou, W., Rousset, F., et al. (1999). Wolbachia infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* 29, 153–160. doi: 10.1016/S0965-1748(98)00119-2

Fast, E. M., Toomey, M. E., Panaram, K., Desjardins, D., Kolaczyk, E. D., and Frydman, H. M. (2011). Wolbachia enhance Drosophila stem cell proliferation and target the germline stem cell niche. *Science* 334, 990–992. doi: 10.1126/science.1209609

Greer, E. L., Maures, T. J., Ucar, D., Hauswirth, A. G., Mancini, E., Lim, J. P., et al. (2011). Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* 479, 365–371. doi: 10.1038/nature10572

Guerrero-Bosagna, C., Settles, M., Lucker, B., and Skinner, M. K. (2010). Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. *PLoS ONE* 5:e13100. doi: 10.1371/journal.pone.0013100

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., et al. (2008). Persistent epigenetic differences associated with prenatal exposure

to famine in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17046–17049. doi: 10.1073/pnas.0806560105

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Ikeya, T., Broughton, S., Alic, N., Grandison, R., and Partridge, L. (2009). The endosymbiont Wolbachia increases insulin/IGF-like signalling in Drosophila. *Proc. Biol. Sci.* 276, 3799–3807. doi: 10.1098/rspb.2009.0778

Jablonka, E. (2012). Epigenetic inheritance and plasticity: the responsive germline. *Prog. Biophys. Mol. Biol.* 111, 99–107. doi: 10.1016/j.pbiomolbio.2012.08.014

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84, 131–176. doi: 10.1086/598822

Kaati, G., Bygren, L. O., and Edvinsson, S. (2002). Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur. J. Hum. Genet.* 10, 682–688. doi: 10.1038/sj.ejhg.5200859

Lee, E. R., Blount, K. F., and Breaker, R. R. (2009). Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol.* 6, 187–194. doi: 10.4161/rna.6.2.7727

Lim, J. P., and Brunet, A. (2013). Bridging the transgenerational gap with epigenetic memory. *Trends Genet.* 29, 176–186. doi: 10.1016/j.tig.2012.12.008

Lumey, L. H., Stein, A. D., Kahn, H. S., van der Pal-de Bruin, K. M., Blauw, G. J., Zybert, P. A., et al. (2007). Cohort profile: the dutch hunger winter families study. *Int. J. Epidemiol.* 36, 1196–1204. doi: 10.1093/ije/dym126

McCaffery, A. R., Simpson, S. J., Islam, M. S., and Roessingh, P. (1998). A gregarizing factor present in the egg pod foam of the desert locust *Schistocerca gregaria*. *J. Exp. Biol.* 201, 347–363.

McGraw, E. A., and O'Neill, S. L. (2004). *Wolbachia pipientis*: intracellular infection and pathogenesis in Drosophila. *Curr. Opin. Microbiol.* 7, 67–70. doi: 10.1016/j.mib.2003.12.003

Molinier, J., Ries, G., Zipfel, C., and Hohn, B. (2006). Transgeneration memory of stress in plants. *Nature* 442, 1046–1049. doi: 10.1038/nature05022

Morgan, H. D., Sutherland, H. G., Martin, D. I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318. doi: 10.1038/15490

Padmanabhan, N., Jia, D., Geary-Joo, C., Wu, X., Ferguson-Smith, A. C., Fung, E., et al. (2013). Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. *Cell* 155, 81–93. doi: 10.1016/j.cell.2013.09.002

Powers, H. J. (2003). Riboflavin (vitamin B-2) and health. *Am. J. Clin. Nutr.* 77, 1352–1360.

Rechavi, O., Minevich, G., and Hobert, O. (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* 147, 1248–1256. doi: 10.1016/j.cell.2011.10.042

Ridley, E. V., Wong, A. C., and Douglas, A. E. (2013). Microbe-dependent and nonspecific effects of procedures to eliminate the resident microbiota from *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 79, 3209–3214. doi: 10.1128/AEM.00206-13

Rosenberg, E., and Zilber-Rosenberg, I. (2011). Symbiosis and development: the hologenome concept. *Birth Defects Res. C Embryo Today* 93, 56–66. doi: 10.1002/bdrc.20196

Sang, J. (1956). The quantitative nutritional requirements of *Drosophila Melanogaster*. *J. Exp. Biol.* 33, 45–72.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., et al. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. *Science (New York, NY)* 334, 369–373. doi: 10.1126/science.1212959

Seong, K.-H., Li, D., Shimizu, H., Nakamura, R., and Ishii, S. (2011). Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell* 145, 1049–1061. doi: 10.1016/j.cell.2011.05.029

Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010). Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20051–20056. doi: 10.1073/pnas.1009906107

Shin, S. C., Kim, S. H., You, H., Kim, B., Kim, A. C., Lee, K. A., et al. (2011). Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334, 670–674. doi: 10.1126/science.1212782

Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M. D., and Ruden, D. M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat. Genet.* 33, 70–74. doi: 10.1038/ng1067

Starr, D. J., and Cline, T. W. (2002). A host parasite interaction rescues Drosophila oogenesis defects. *Nature* 418, 76–79. doi: 10.1038/nature00843

Stern, S., Fridmann-Sirkis, Y., Braun, E., and Soen, Y. (2012). Epigenetically heritable alteration of fly development in response to toxic challenge. *Cell Rep.* 1, 528–542. doi: 10.1016/j.celrep.2012.03.012

Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J., and Leulier, F. (2011). *Lactobacillus plantarum* promotes Drosophila systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab.* 14, 403–414. doi: 10.1016/j.cmet.2011.07.012

Tatum, E. L. (1939). Nutritional requirements of *Drosophila Melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 25, 490–497. doi: 10.1073/pnas.25.9.490

Veneti, Z., Clark, M. E., Zabalou, S., Karr, T. L., Savakis, C., and Bourtzis, K. (2003). Cytoplasmic incompatibility and sperm cyst infection in different Drosophila-Wolbachia associations. *Genetics* 164, 545–552.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703.

Werren, J. H. (1997). Biology of Wolbachia. *Annu. Rev. Entomol.* 42, 587–609. doi: 10.1146/annurev.ento.42.1.587

Wong, C. N. A., Ng, P., and Douglas, A. E. (2011). Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ. Microbiol.* 13, 1889–1900. doi: 10.1111/j.1462-2920.2011.02511.x

Xing, Y., Shi, S., Le, L., Lee, C. A., Silver-Morse, L., and Li, W. X. (2007). Evidence for transgenerational transmission of epigenetic tumor susceptibility in Drosophila. *PLoS Genet.* 3:e151. doi: 10.1371/journal.pgen.0030151

Environmental disruption of host–microbe co-adaptation as a potential driving force in evolution

Yoav Soen*

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel

The microbiome is known to have a profound effect on the development, physiology and health of its host. Whether and how it also contributes to evolutionary diversification of the host is, however, unclear. Here we hypothesize that disruption of the microbiome by new stressful environments interferes with host–microbe co-adaptation, contributes to host destabilization, and can drive irreversible changes in the host prior to its genetic adaptation. This hypothesis is based on three presumptions: (1) the microbiome consists of heritable partners which contribute to the stability (canalization) of host development and physiology in frequently encountered environments, (2) upon encountering a stressful new environment, the microbiome adapts much faster than the host, and (3) this differential response disrupts cooperation, contributes to host destabilization and promotes reciprocal changes in the host and its microbiome. This dynamic imbalance relaxes as the host and its microbiome establish a new equilibrium state in which they are adapted to one another and to the altered environment. Over long time in this new environment, the changes in the microbiome contribute to the canalization of the altered state. This scenario supports stability of the adapted patterns, while promoting variability which may be beneficial in new stressful conditions, thus allowing the organism to balance stability and flexibility based on contextual demand. Additionally, interaction between heritable microbial and epigenetic/physiological changes can promote new outcomes which persist over a wide range of timescales. A sufficiently persistent stress can further induce irreversible changes in the microbiome which may permanently alter the organism prior to genetic changes in the host. Epigenetic and microbial changes therefore provide a potential infrastructure for causal links between immediate responses to new environments and longer-term establishment of evolutionary adaptations.

Keywords: host–microbe symbiosis, genotype-to-phenotype transformations, canalization, developmental plasticity, epigenetics, dynamical systems, non-Mendelian inheritance, adaptation

HYPOTHESIS

THE MICROBIOME AS A 'DOUBLE AGENT' OF CANALIZATION AND DE-CANALIZATION

In his influential perspective on "Canalization and the inheritance of acquired characters" (Waddington, 1942), Conrad Waddington argued that evolution of developing organisms under natural selection tends toward canalization of the developmental process. That is to say, the sensitivity of developmental patterns to environmental and genetic perturbations is reduced over time in the environment to which the organism becomes adapted. A similar idea was developed independently by (Schmalhausen, 1949). Owing to this hypothesized tendency toward canalization, the patterns of development in the wild type organism are presumed to be more stable than the patterns in mutants (Waddington, 1942). This allows cryptic variation to be accumulated without phenotypic consequences. This variation can be unmasked by a sufficiently disruptive environmental or genetic perturbation (de-canalizing event) which breaks robustness and leads to phenotypic consequences reflecting the previously hidden variability (Flatt, 2005). Although canalization is most likely a systemic property, few specific molecular mechanisms have

been proposed to support canalization [reviewed in (Salathia and Queitsch, 2007)]. These include buffering of genetic variability by the Hsp90 chaperone (Rutherford and Lindquist, 1998; Queitsch et al., 2002; Milton et al., 2006; Samakovli et al., 2007; Sgro et al., 2010), tissue-specific maintenance of active and repressed state of developmental genes by the Polycomb (and by analogy also the trithorax) system (Lee et al., 2005; Sawarkar and Paro, 2010; Stern et al., 2012), stabilizing negative feedback by microRNAs (Ronshaugen et al., 2005; Hornstein and Shomron, 2006; Li et al., 2009; Wu et al., 2009), and piwi-mediated silencing of transposon activity (Specchia et al., 2010; Gangaraju et al., 2011) or of existing variation (Gangaraju et al., 2011).

It is quite possible, though, that at least part of the noticeable contribution of these molecular pathways to canalization is due to their particularly wide range of interactions with other developmental genes and processes. In other words, these pathways are so deeply embedded into many regulatory feedbacks, that their disruption (by the environment or by a genetic change) can lead to significant impacts on developmental and physiological phenotypes. If correct, this rationale should hold more broadly for
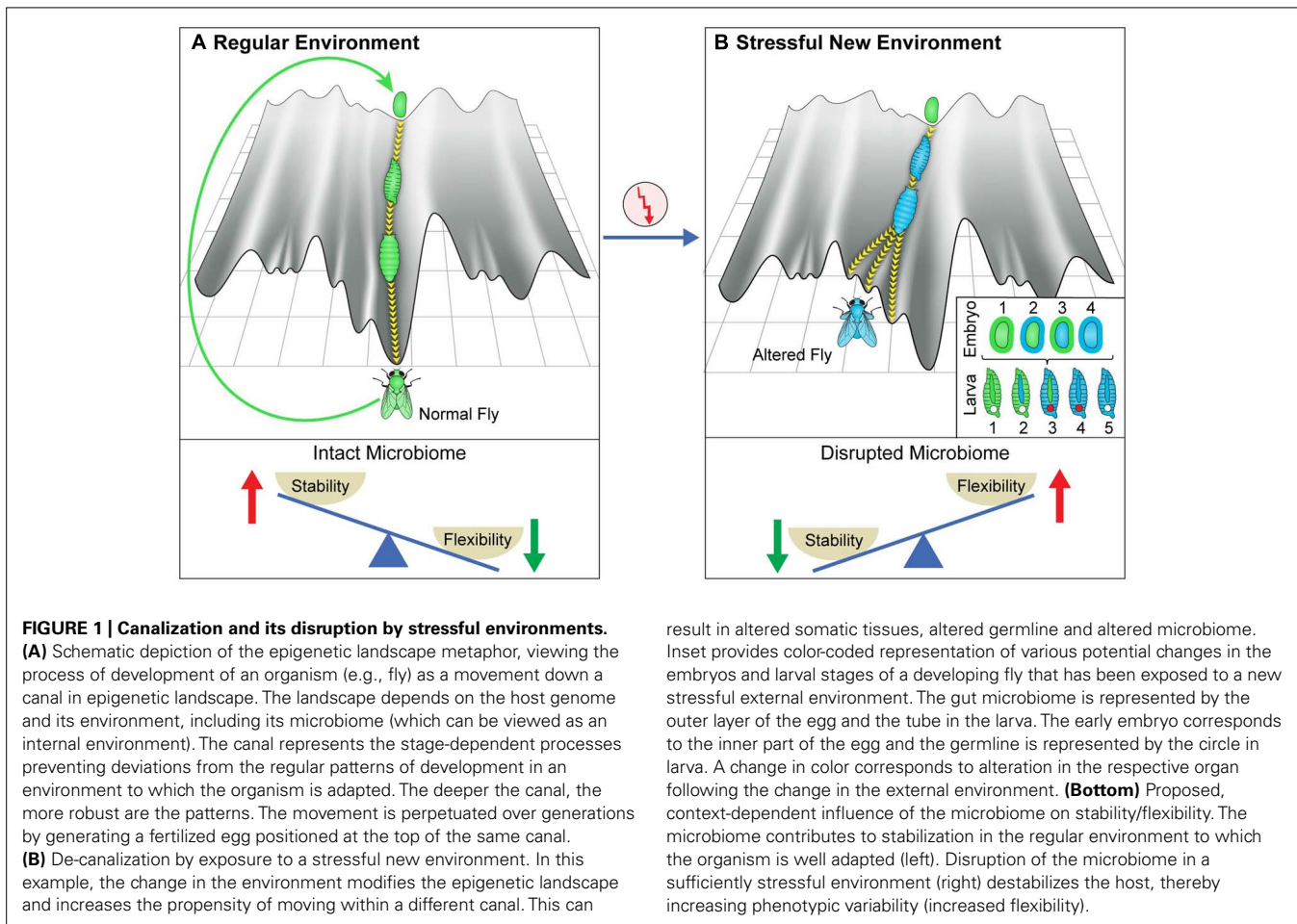
other intrinsic and extrinsic factors, which were co-opted to a particularly wide range of processes. One candidate factor is the microbiome which has co-evolved with its host for many generations. Similarly to host-intrinsic molecules and pathways that become involved in multiple functions, co-evolution of the microbiome with its host likely leads to integration of the microbiome into diverse host functions (Backhed et al., 2005; Rosenberg et al., 2009). This integration may be further widened and deepened by the ability of bacteria to produce many factors which could interact and cooperate with host processes. As part of this cooperation, the microbiome can provide the host with metabolites and signaling molecules (Backhed et al., 2004; Arora and Sharma, 2011; Shin et al., 2011; Flint et al., 2012; LeBlanc et al., 2013), assist in harvesting these molecules (Yamanaka et al., 1977; Brune and Friedrich, 2000; Backhed et al., 2005; Wang et al., 2011), synthesize organic compounds that are passed on to the host (Cavanaugh et al., 2006; Dubilier et al., 2008), regulate the immune system [reviewed in (Round and Mazmanian, 2009; Kranich et al., 2011)], reduce the propensity for disease (Markle et al., 2013), protect against invaders [Fukuda et al., 2011; Wang et al., 2013; which may be outcompeted and displaced by the microbiome (Kamada et al., 2012)], affect lifespan (Brummel et al., 2004; Behar et al., 2008) and fecundity (Ben-Yosef et al., 2010), influence mating preference (Sharon et al., 2010a), and likely perform many other developmental stage-dependent or independent functions yet to be discovered. Such evolved cooperation can result in wide, reciprocal interdependencies between the host and its microbiome. Owing to this broad dependency of the host on its microbial partners, disruption of the microbiome can compromise normal host development and homeostasis (Littman and Pamer, 2011; Martin et al., 2013). Conversely, the intact microbiome could contribute to the stability of host processes. This presumption is supported by recent work in a variety of organisms which have begun to uncover increasing numbers of host pathways and processes that could be affected by disruption of the microbiome (Brummel et al., 2004; Sharon et al., 2010b; Shin et al., 2011; Storelli et al., 2011), including modulation of epigenetic gene control [Takahashi, 2014; e.g., methylation of Toll-like receptors in intestine cells (Takahashi et al., 2011)]. As in the case of mutations (Waddington, 1942), a disrupted microbiome does not necessarily de-stabilizes every phenotype in every stressful scenario. However, it is expected to promote significantly more cases of compromised phenotypic robustness compared to cases of increased robustness.

When the host and its microbiome co-evolve in a stable, or a frequently occurring environment, they become adapted not only to this environment, but also to one another. We refer to this state as coordinated adaptation. Deviations from this co-adapted state (in response to environmental, genetic, or epigenetic modifications) could create stressful conditions promoting reciprocal changes in the host and its microbiome. Stated differently, modifications in the previously adapted microbiome can change the internal environment of the host. This change in the internal environment could promote response in the host which, in turn, may affect the conditions for survival and growth of the bacteria, and so on. Coordinated adaptation is re-established if and when the overall responses reduce the stress to the host and its microbiome to a level no longer sufficient for driving further changes. This is analogous

to the Le Chatelier's principle in chemistry (Atkins and de Paula, 2002), namely: if a system in chemical equilibrium is subjected to a disturbance, it tends to change in a way that opposes this disturbance. Thus, the establishment of a coordinated adaptation of the host and its microbiome can be viewed as an extension of the Le Chatelier's principle, from chemical systems to self-organization of living cells and organisms.

Establishing a new co-adapted state involves interactions between the microbiome and the physiological and epigenetic state of the host. Important aspects of this interaction can be heuristically illustrated using the epigenetic landscape metaphor (Waddington, 1957). This metaphor views the development of an organism as movement down a canal (**Figure 1**, top left), starting with the fertilized egg and moving toward the adult form. Developed adults produce a fertilized egg positioned at the top of the same canal, thus starting the process over again. Here, the walls of the canal represent stabilizing (canalizing) processes, acting at every stage to prevent deviations from the normal patterns of development (termed by Waddington as homeorhesis to denote the constancy of flow along a determined trajectory). Still, the canal and the entire landscape depend on the host genome and its environment (Waddington, 1957), including the microbiome which can be viewed as an "internal" environmental factor. It is important to note, however, that the use of a 2-dimensional surface in this metaphor is an over simplification which may not properly represent attractors and flows in a high dimensional space corresponding to a developing organism. Moreover, because of potential influences of (past and present) environments and the dependence of these influences on the actual trajectory occupied by the organism, the details of the landscape are more dynamic and less pre-defined than often perceived.

The robust patterns of development in frequently encountered environments (to which the organism is adapted) correspond to movement in a particularly deep canal (although the depth of the canal is likely inversely correlated with the plasticity of the developmental process and may therefore be shallower during early specification of cells and body segments and possibly also during metamorphosis in which a given tissue type is transformed into another). At any given time in a frequently encountered environment, the microbiome "assists" the establishment of the adapted patterns and is therefore, an agent of canalization. However, since the epigenetic landscape can be influenced by the environment, a sufficiently strong environmental stress can modify the landscape to an extent sufficient to de-canalize the process of development OFF the "regular canal" and ON to a "side canal" (**Figure 1**, top right). The change in landscape can reflect a direct effect of the external environment on the host as well as an indirect effect through environmental disruption of the microbiome. Indeed, severe disruption of the microbiome could compromise the normal function of many host processes, including those which contribute to robustness. This would tend to destabilize (de-canalize) the adapted patterns, leading to dysbiosis and increased phenotypic variability (wider sampling of the available epigenetic landscape). While such microbial disruption is not expected in frequently occurring environments to which the organism is well adapted, it can be readily occur under severe stressful environments in which the organism is considerably

FIGURE 1 | Canalization and its disruption by stressful environments.
(A) Schematic depiction of the epigenetic landscape metaphor, viewing the process of development of an organism (e.g., fly) as a movement down a canal in epigenetic landscape. The landscape depends on the host genome and its environment, including its microbiome (which can be viewed as an internal environment). The canal represents the stage-dependent processes preventing deviations from the regular patterns of development in an environment to which the organism is adapted. The deeper the canal, the more robust are the patterns. The movement is perpetuated over generations by generating a fertilized egg positioned at the top of the same canal.
(B) De-canalization by exposure to a stressful new environment. In this example, the change in the environment modifies the epigenetic landscape and increases the propensity of moving within a different canal. This can

result in altered somatic tissues, altered germline and altered microbiome. Inset provides color-coded representation of various potential changes in the embryos and larval stages of a developing fly that has been exposed to a new stressful external environment. The gut microbiome is represented by the outer layer of the egg and the tube in the larva. The early embryo corresponds to the inner part of the egg and the germline is represented by the circle in larva. A change in color corresponds to alteration in the respective organ following the change in the external environment. **(Bottom)** Proposed, context-dependent influence of the microbiome on stability/flexibility. The microbiome contributes to stabilization in the regular environment to which the organism is well adapted (left). Disruption of the microbiome in a sufficiently stressful environment (right) destabilizes the host, thereby increasing phenotypic variability (increased flexibility).
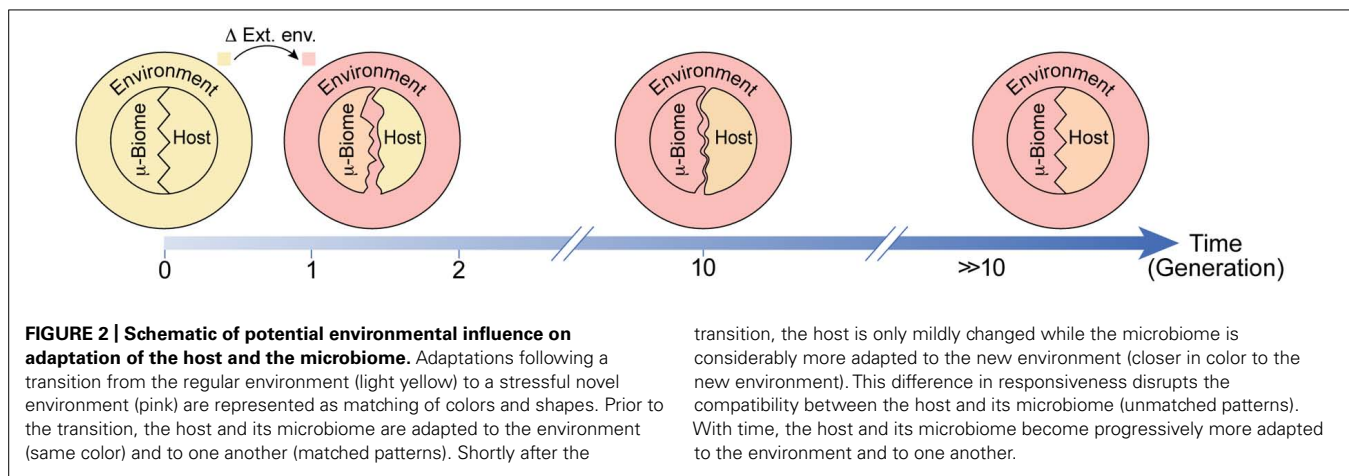
maladapted. Thus, the microbiome can contribute to stability in frequently encountered environments, while promoting variability in sufficiently stressful new environments (**Figure 1**, bottom panels).

SPECIES-SPECIFIC RATES OF ADAPTATION CREATE HOST–MICROBE INCOMPATIBILITIES WHICH DISRUPT CO-ADAPTATION AND PROMOTE RECIPROCAL CHANGES

Although the host and its microbiome form a partnership of organisms which propagate as a community, members of this partnership exhibit species-specific differences in the rate, type, and magnitude of responses to the environment (Zilber-Rosenberg and Rosenberg, 2008; Rosenberg and Zilber-Rosenberg, 2011). Owing to differences in generation time, population size, and developmental constraints, the microbiome can adapt to new environments on time scales that are much faster than those of genetic adaptation in the host (Zilber-Rosenberg and Rosenberg, 2008). For example, the doubling time of bacteria is on the order of 1 h whereas the generation time of a human host is about 20 years. Thus, within a single human generation, bacterial population can be propagated over 10^5 generations. Combining this large number of generations with very large population sizes [$>10^{13}$ bacterial cells in human (Backhed et al., 2005)] provides a very substantial potential for bacterial adaptation and diversification within

a single generation of the host. Notably, modifications in the microbiome include changes in gene sequence (and/or function) within existing bacteria, and changes in species composition of the microbial community. This community often comprises many different species, whose relative abundance depends on the environment. As such, a stressful environment for the bacteria leads to species-specific selection which alters the composition and genetic structure of the community. This selection can be independent of whether or not the stress is also compromising the survival of the host.

While the bacteria can undergo substantial adaptation to new environments already within a single or few generations of the host, the latter remains largely genetically adapted to the previous environment (**Figure 2**). This can compromise the adaptation of the host to the new environment as well as its compatibility with the microbiome, leading to several distinct effects: first, disruption of the microbiome can destabilize the host, thereby enhancing its phenotypic responsiveness to the new environment (potentially resulting in increased occurrence, magnitude, and diversity of the phenotypic response). Additionally, the change in the microbiome is itself an "internal" environmental input which feeds back on the physiological and epigenetic response of the host (**Figure 3**). For example, the microbiome can respond to a new stress by producing (or elevating) a substance that is toxic to the host. This substance

FIGURE 2 | Schematic of potential environmental influence on adaptation of the host and the microbiome. Adaptations following a transition from the regular environment (light yellow) to a stressful novel environment (pink) are represented as matching of colors and shapes. Prior to the transition, the host and its microbiome are adapted to the environment (same color) and to one another (matched patterns). Shortly after the transition, the host is only mildly changed while the microbiome is considerably more adapted to the new environment (closer in color to the new environment). This difference in responsiveness disrupts the compatibility between the host and its microbiome (unmatched patterns). With time, the host and its microbiome become progressively more adapted to the environment and to one another.

adds to the external stress and can modify the type and magnitude of the response in the host even if the external environment is back to normal (**Figure 3B**). This is because the microbiome may still remain altered and disruptive to the host. Thus, the response of the host to microbial-mediated inputs can persist as long as the host–microbe incompatibility is large enough to modify the host.

TRANSGENERATIONAL IMPLICATIONS OF THE DISRUPTION OF COORDINATED ADAPTATION

The presumed contribution of the microbiome to the stabilization of host development and physiology is analogous to that of other factors (e.g., Hsp90) which can confer robustness to the organism in frequently occurring environments. However, unlike many other factors, the change in composition and/or genetic sequence of resident bacteria might itself be heritable, thereby providing potential infrastructure for multi-generational influences on the host (Zilber-Rosenberg and Rosenberg, 2008). Indeed, the microbiome in many organisms has been shown to be transmitted across generations either vertically or horizontally (Bright and Bulgheresi, 2010). Vertical transmission refers to transfer within the zygote, while horizontal transmission is based on various forms of pickup of micro-organisms from the surroundings, followed by their re-colonization within the host. These mechanisms of transmission normally facilitate inheritance of the intact (co-adapted) microbiome. However, the same mechanisms could support inheritance of environmentally modified microbiome. In line with this possibility, we have recently found that environmentally induced changes in the composition of gut bacteria in flies can be inherited (Fridmann-Sirkis et al., 2014). Similar findings are expected whenever the microbiome is altered to a degree which no longer supports full recovery within one generation. In this case, the offspring inherit altered microbiome which may or may not have been partially restored. The extent of inherited changes depends on the details of the external environment (type and strength of stress, its onset and duration, proximity to unaffected individuals, etc.), the internal conditions inside the host and the dynamics within the community of microbial species.

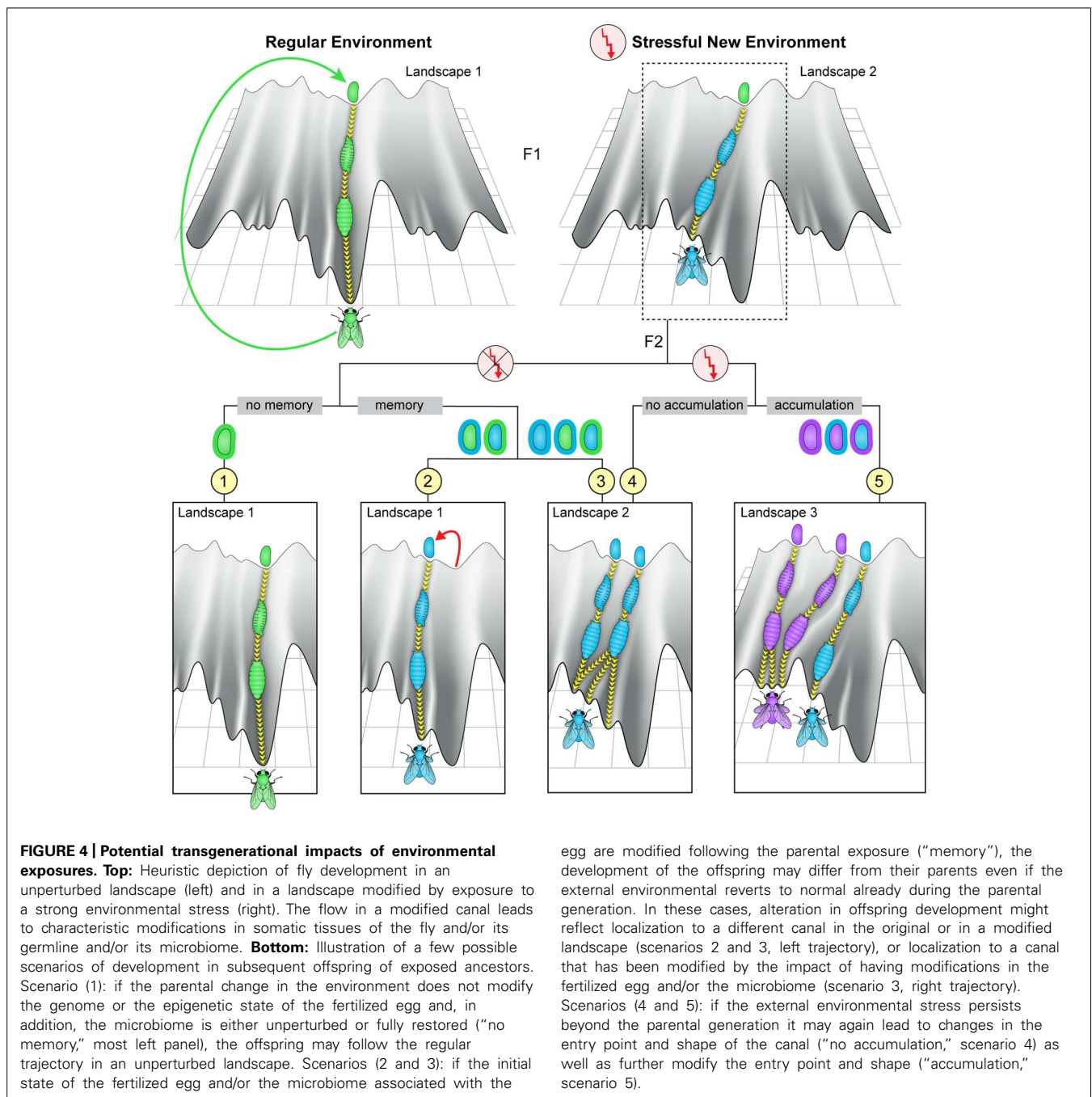Modifications in the inherited microbiome, can change the initial conditions of offspring development. Depending on the structure and dynamics of the epigenetic landscape, this could correspond to a change in trajectory leading to a different outcome in the adult (**Figure 4**, top right). The altered adults could again produce modified embryos, thus enabling multigenerational changes in movements in epigenetic space. These multi-generational changes may involve different scenarios which could support non-Mendelian inheritance. Heuristic representation of some of these scenarios is depicted in the bottom panels of **Figure 4** (scenarios 2–5). For example, a change in the microbiome in one generation can potentially modify the host germline, leading to altered embryonic development of the offspring (viewed either as an altered initial point in the landscape or as movement in a modified landscape). Alternatively, persistent change in the microbiome can alter later stages of development, leading again to a modified offspring adult with a potentially modified microbiome. Evidence supporting potential influence of resident bacteria on the germline in insects has been reported for the endosymbiont *Wolbachia* (Werren, 1997; Stouthamer et al., 1999; Starr and Cline, 2002; Negri et al., 2006, 2009). More recent evidence in flies (*Drosophila melanogaster*) provided indirect evidence for influence of gut microbiome on the germ line (Fridmann-Sirkis et al., 2014). Specifically, removal of gut bacteria in one generation led to a substantial delay in larval development starting in the following generation. The phenotypic difference between parents and offspring following the depletion of gut bacteria (in parents) suggests an influence of these bacteria on the parental germline. Such a transgenerational effect of bacterial depletion was further shown to be responsible for the inheritance of a delay in larval development following parental exposure to G418 antibiotic (Fridmann-Sirkis et al., 2014). In this case, the delay in the first generation of larvae is caused by a direct toxicity of G418 in the host tissue, but is maintained in the non-exposed offspring by the transgenerational impact of depleted gut *Acetobacter* species in the G418-exposed parents. This unexpected situation is just one example of potentially diverse circumstances in which disruption of the microbiome could contribute to non-Mendelian transfer of influences. Thus, regardless of whether the effects of microbial changes are restricted to the soma or not, inheritance of a modified microbiome provides substantial infrastructure for multi-generational persistence of non-genetic changes in the host.
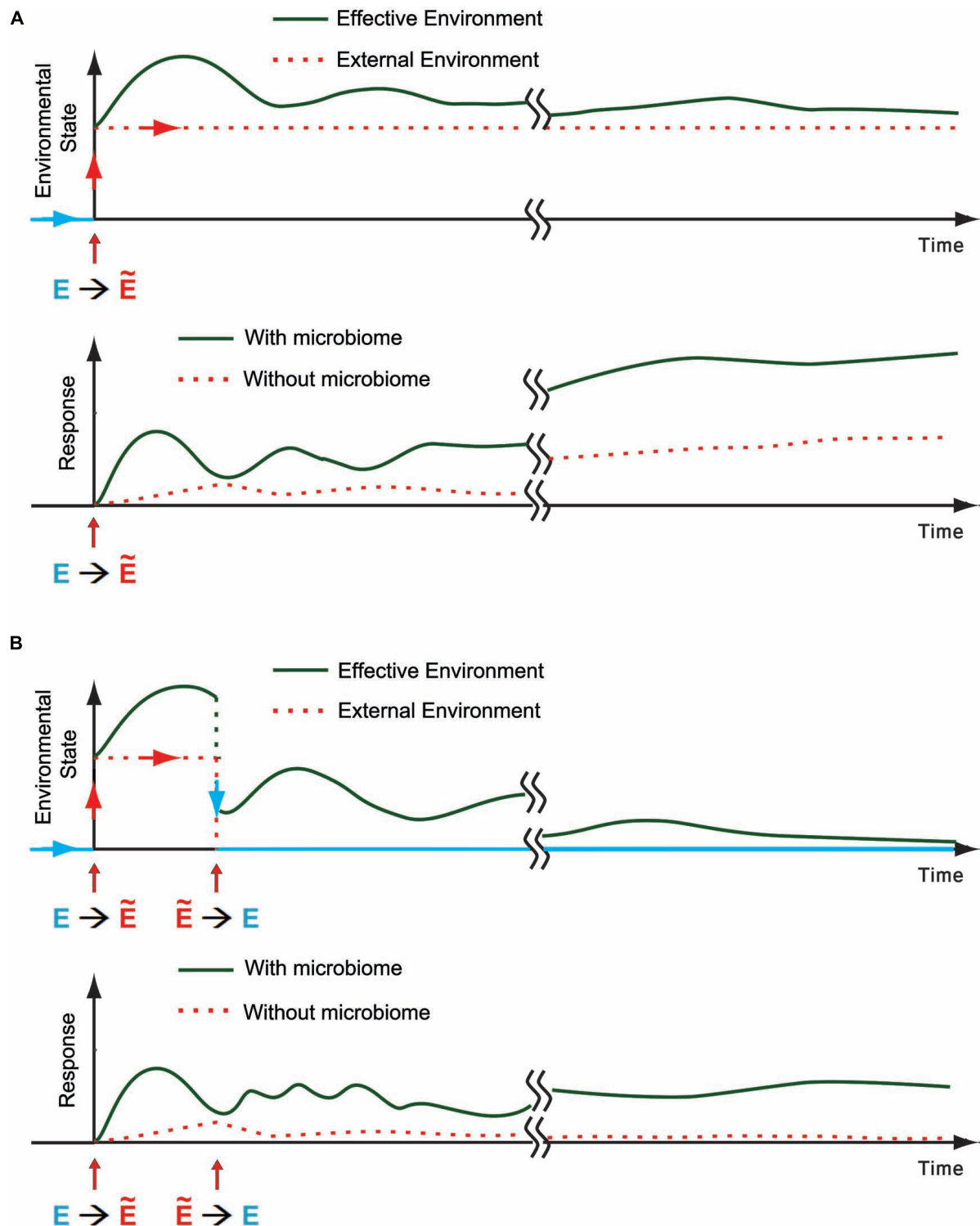
FIGURE 3 | **The microbiome as an environmental component for the host. (A) Top:** Schematic representation of a transition from the regular external environment toward a new stressful environmental state (E→Ẽ). The effective environmental state (green) is the sum of all the environmental influences on the host, taking into accounts the external conditions and internal inputs from the microbiome. An example of one possible trace of effective environmental state is shown in green.

Bottom: Schematic representation of a potential host response with and without contribution of microbial disruption by the external environment (green and dotted red, respectively). Disruption of the microbiome can increase the response of the host, leading to a potentially faster and more pronounced response which is also more diverse and affects more individuals. **(B)** Same as **(A)** when the environment reverts back (Ẽ→E) within generation.

FIGURE 4 | Potential transgenerational impacts of environmental exposures. Top: Heuristic depiction of fly development in an unperturbed landscape (left) and in a landscape modified by exposure to a strong environmental stress (right). The flow in a modified canal leads to characteristic modifications in somatic tissues of the fly and/or its germline and/or its microbiome. **Bottom:** Illustration of a few possible scenarios of development in subsequent offspring of exposed ancestors. Scenario (1): if the parental change in the environment does not modify the genome or the epigenetic state of the fertilized egg and, in addition, the microbiome is either unperturbed or fully restored ("no memory," most left panel), the offspring may follow the regular trajectory in an unperturbed landscape. Scenarios (2 and 3): if the initial state of the fertilized egg and/or the microbiome associated with the

egg are modified following the parental exposure ("memory"), the development of the offspring may differ from their parents even if the external environmental reverts to normal already during the parental generation. In these cases, alteration in offspring development might reflect localization to a different canal in the original or in a modified landscape (scenarios 2 and 3, left trajectory), or localization to a canal that has been modified by the impact of having modifications in the fertilized egg and/or the microbiome (scenario 3, right trajectory). Scenarios (4 and 5): if the external environmental stress persists beyond the parental generation it may again lead to changes in the entry point and shape of the canal ("no accumulation," scenario 4) as well as further modify the entry point and shape ("accumulation," scenario 5).

This persistence depends on the strength and duration of the environmental stress, the degree of incompatibility between the host and the microbiome, and their inherent plasticity, or adaptability. The manner in which these factors are combined has a significant effect on the timescale of establishing a new equilibrium state.

TIMESCALES OF INTERACTIONS BETWEEN STRESS-INDUCED EPIGENETIC AND MICROBIAL CHANGES

Current understanding of dynamical systems suggests that interaction between stress-induced disruption of the microbiome and

modifications in the epigenetic state provide potential for a rich variety of dynamical outcomes. Indeed, coupling between relatively simple non-linear systems can generate diverse modes of collective or non-collective dynamics, depending on the intrinsic parameters of the uncoupled systems, the type and strength of the coupling, the type and magnitude of the non-linearity, and the initial conditions (Kuramoto, 1984; Glass and MacKey, 1988; Badola et al., 1991; Otsuka, 1991; González-Miranda, 2004). For example, specific choices of couplings between two non-linear oscillators with sufficiently similar intrinsic frequencies can lead to synchronization of these oscillators, doublings of their intrinsic

or collective periods, chaotic dynamics (of either one or both) and destruction of the oscillatory behavior (Sarkar et al., 2013). It is plausible that analogous wealth of dynamics could be generated by interaction between the microbiome and the epigenetic state of the host (regarded here in an inclusive sense which covers various types of non-genetic host-intrinsic changes). This interaction can be viewed as coupling between two non-linear systems, each modulating the activity of the other. Unlike the slowly varying genome sequence of the host, epigenetic changes in the host can occur on a range of timescales which overlaps with timescales associated with microbial changes. This provides potential for inter-modulations which prolong the duration of changes expected in a microbiome-free host. Encountering such scenarios under extreme forms of stress is quite plausible and may have occurred in a recent example of transgenerational inheritance following parental exposure to toxic stress (Stern et al., 2012). This inheritance was shown to involve host-intrinsic (Stern et al., 2014) and microbial-mediated mechanisms (Fridmann-Sirkis et al., 2014), and typically persisted for 3–10 generations. It is possible that interactions between the host-intrinsic and microbial-mediated modifications extended the duration of inheritance beyond the duration that would have been observed without an interaction.

Depending on the duration of change in the external environment, one can further distinguish three different regimes:

Stressful external conditions which persist for only one, or a small number of generations

This regime of stress may initiate reciprocal modifications in the microbiome and the host. For most stressful environments, the modifications in the host would be primarily non-genetic (e.g., physiologic, metabolic, epigenetic etc.'). The microbiome, on the other hand, may change in gene sequence and species composition (in addition to epigenetic changes). Owing to the heritability of some of these changes [and potentially also heritability of epigenetic changes in the host (Jablonka and Raz, 2009; Jablonka, 2013; Lim and Brunet, 2013)], return to the normal environment will not necessarily lead to immediate reversion of the host back to its previous state (**Figure 4**, panels 2 and 3). This is because reversal of the environment does not necessarily eliminate host–microbe incompatibilities (and/or persistent host-intrinsic epigenetic changes) that have been induced during the external stress period. The induced host–microbe incompatibility promotes cross talk between the microbiome and the physiologic/epigenetic state of the host, eventually leading to establishment of a new co-adapted state in which the host and its microbiome become re-adapted to one another. The process of relaxation toward an altered state could span multiple generations of the host, but might be counteracted by re-exposure to non-modified microbiome (e.g., by infection-mediated transfer from individuals which were not exposed to the new environmental conditions).

Persistence of the stressful new environment over timescales much longer than required to reach a new host–microbe equilibrium state

This scenario likely promotes more extensive and longer lasting changes in the host and the microbiome. After a few generations in this stable environment, the microbiome will likely become irreversibly modified. The modified microbiome will continue to change with its host until they become sufficiently adapted to the altered environment and to one another. If, in addition, the stressful environment influences a large geographic region, it will substantially reduce the chances of encountering non-exposed individuals, thus reducing the chances of infection-mediated reversion to the original microbiome. It is also possible that the microbiome niche will be taken over by modified bacteria that are abundant enough to prevent re-colonization of the original bacteria. Additionally, the host may become sufficiently modified so as to favor the altered microbiome over the original microbiome.

Persistence of external stress over a time period comparable to that required for co-adaptation
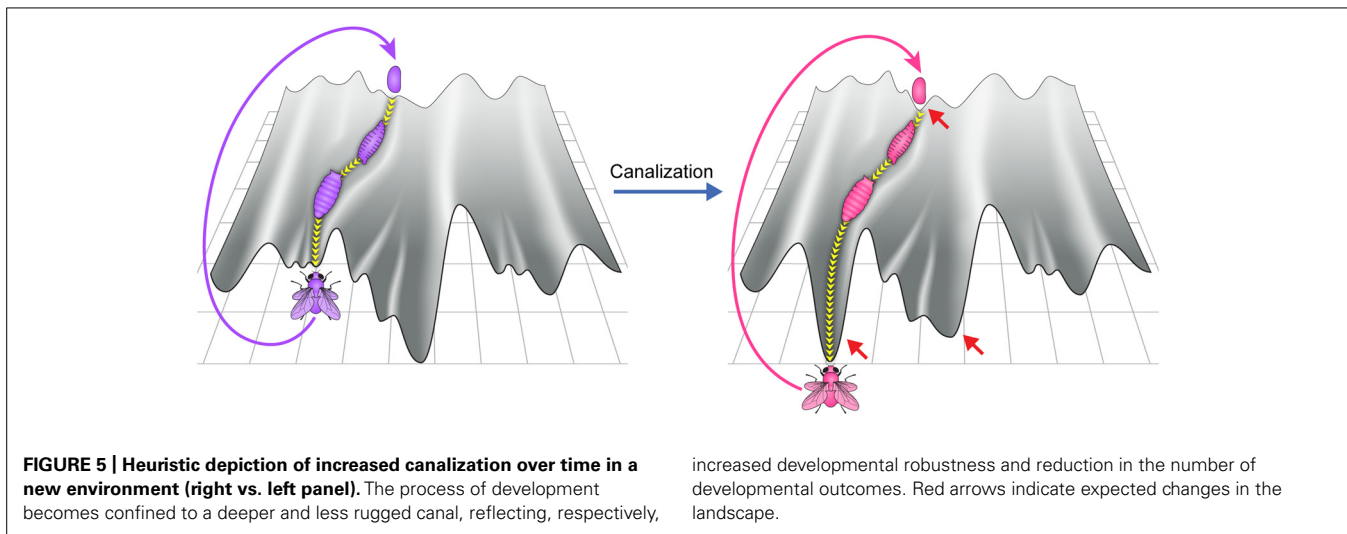
In this intermediate case, equilibrium of the host and the microbiome can be re-challenged by the reversion of the environment, thus prolonging the establishment of a new stable state. Since the timescales of environmental change might indeed be comparable to those of epigenetic and microbial changes, this scenario may not be extremely rare. Additionally, the likelihood of encountering this scenario may increase when the organism is capable of modifying its external environment (niche construction).

Taken together, the above regimes support phenotypic diversification and/or adaptation over timescales ranging from few to many generations. This diversification may become irreversible prior to genetic changes and adaptation in the host. As such, it may provide important bridge between the physiological timescales of one generation and the much longer timescales of genetic adaptation and phylogenetic diversification.

According to the original canalization hypothesis (Waddington, 1942; Schmalhausen, 1949), a new co-adapted state of the host and its microbiome is expected to become progressively more canalized (stabilized) over time (**Figure 5**). The principles and mechanisms leading to progressive stabilization are not entirely clear. It was suggested that the trend toward stabilization is implemented by increasing the connectivity of the underlying regulatory network, which is in turn supported by a tendency of natural selection to favor stable development (Siegal and Bergman, 2002). An increase in regulatory complexity can be achieved by different means, including by wider integration (co-option) of the microbiome into host processes. Increasing integration may reflect an entropic effect which is expected from having many more potential scenarios of integration compared with no integration. Among these scenarios would be host–microbiome integrations which either increase or reduce the stability of development. However, if changes which increase the stability tend to be more persistent, the entropic force of increasing co-option is expected to promote progressively more extensive host–microbiome interactions that are biased toward increased stability.

DIFFERENT UNITS OF SELECTION AND THEIR IMPLICATIONS FOR ADAPTATION AND EVOLUTION

Much of the above considerations stems from having more than one unit of selection. First, the host and its microbes are subjected to selection as a cooperating community. This is the most relevant

FIGURE 5 | Heuristic depiction of increased canalization over time in a new environment (right vs. left panel). The process of development becomes confined to a deeper and less rugged canal, reflecting, respectively, increased developmental robustness and reduction in the number of developmental outcomes. Red arrows indicate expected changes in the landscape.

unit of selection for the host. On the other hand, each of the resident bacterial species is subjected to its own selection conditions within the host. The outcomes of these individual selections feed into the higher level of host (or community) selection.

This interaction between different layers of selection has far reaching implications which are not accounted for in current models of population genetics (Jaenike, 2012). The latter often consider changes in allele frequencies in the host, ignoring the potential contribution of changes in bacterial allele frequencies to the selection of the host. These changes in the microbiome could affect the propagation of the host, either by modifying it or by changing its internal environment. Owing to these changes, the host phenotypes and epigenetic state may be stably altered prior to the emergence or fixation of adaptive genetic changes in its own genome.

Selection of bacterial species within the host may also promote maladaptive changes in the host, thus providing potential for conflicts between the microbiome and the host. On longer timescales, however, these conflicts are subjected to selection at the higher, community level, which could then promote successful communities, and hence, successful hosts (Zilber-Rosenberg and Rosenberg, 2008). A somewhat similar, host-intrinsic scenario involving multiple layers of selection is provided by somatic mutations which lead to generation of tumor cells. These cells undergo changes promoting their own propagation on the expense of the host. However, unlike the microbiome, these tumors are not inherited, and do not evolve to cooperate with their host.

What is then the added evolutionary value of having these two layers of selection? Here, one can immediately recognize a few important considerations: first, the microbiome offers a substantial gene and cellular pool which can evolve to support reproduction of the community under new environmental conditions (Zilber-Rosenberg and Rosenberg, 2008). For example, functions and products of host cells may be provided as a service by resident bacteria (Backhed et al., 2005). This includes extreme cases in which endosymbiotic bacteria synthesize organic compounds (e.g., by chemosynthesis) and provide nutrition to certain gutless flatworms (Ott, 1982). The feasibility of such remarkable

collaborations is indicative of the vast potential of the microbiome to generate "creative" solutions. Important components of this potential are the relative speed in which these solutions can emerge and the separation from the host. Having these rapid changes in entities distinct from the host allows the community to dramatically expand the range of genetic and cellular variation while paying significantly reduced costs of maladaptive changes. Indeed, a comparable extent of genetic and cellular changes in the host itself, will likely lead to failed development, or alternatively, tumorigenesis and death. In other words, variation of the microbiome is likely a safer option in much the same way as would therapy with bacteria as opposed to host-intrinsic gene therapy. Thus, the microbiome offers extended potential for rapid, flexible, and safer exploration of new solutions and functions which benefit the host. This potential is particularly important when the host encounters new problems which can be not eliminated fast enough by a change in the host, but may be alleviated by rapid adaptation of the microbiome. Thus, the microbiome may provide a developing organism with a substantial capability to adapt to novel stressful conditions. It is often assumed that the demand for such capability is not very strong because of the extremely low likelihood of encountering a truly novel environment. However, this rationale may not necessarily invalidate the demand for adaptive capacity because the organism may also face new stressful conditions in regular environments. Indeed, many changes in the genome, epigenome, or the microbiome can create stressful conditions. Owing to the immensely large space of possible changes at these levels, it is not be realistic to assume that a large fraction of all possible scenarios have been frequently encountered during the evolutionary history of the organism. In this case, the organism is not expected to be equipped with an efficient genetic program for coping with all these stressful scenarios. Thus, the demand for at least some ability to adapt to new conditions within one or few generations can be quite high. Future work will determine whether and how epigenetics and symbiosis provide the organism with such adaptive plasticity.

DIRECTIONS FOR FUTURE STUDIES

The hypothesized involvement of the microbiome in regulating stability and flexibility based on contextual demand generates clear predictions which could be used to assess the validity of this hypothesis. In particular, the hypothesis asserts that: (1) intact microbiome tends to suppress environmentally- and genetically-induced phenotypic variation (compared to modified microbiome), and (2) severe stressful conditions tend to compromise this buffering by disrupting the composition and/or genetic and epigenetic makeup of the intact microbiome. These predictions can be tested by combining experimental manipulation of the microbiome with environmental and genetic perturbations. For example, the buffering capacity of the intact bicrobiome can be tested by comparing phenotypic variation in germ-free animals with variation in animals containing intact bacteria. Since the hypothesized buffering refers only to a statistical tendency to stabilize the phenotypes, the evaluation must be based on a sufficiently large set of experimental setups, involving different phenotypes in a variety of organisms subjected to different conditions (e.g., malnutrition, exposure to environmental stressors, use of genetically modified animals etc.). If the hypothesis is correct, we expect a positive correlation between lack of bacteria and the extent of deviation from regular developmental phenotypes. Positive correlation is also expected for milder manipulations of the microbiome, but the statistical validation in this case likely requires averaging across larger sets of conditions and phenotypes. It is also important to note that compromised (or complete lack of) microbiome in otherwise good conditions might not be disruptive enough to modify particular phenotypes in the host. Still, if a compromised microbiome reduces buffering capacity in the host, it should increase the likelihood that additional environmental or genetic insults will induce phenotypic variation. We therefore expect that the positive effect of bacterial disruption on the induction of phenotypic variability would tend to increase as a function of the overall amount of stress. The second part of the hypothesis (environmental suppression of phenotypic buffering by bacteria), can be tested by exposing developing organisms to different types of severe environmental stress and analyzing the ability of the stress to modify the overall composition of the microbiome as well as the properties of its individual bacterial species. Here again, the evaluation should include a variety of stress paradigms which do not involve stressors that have been specifically selected for targeting bacteria.

To further evaluate the hypothesized contribution of the microbiome to stable diversification and adaptation prior to genetic changes in their host, it would be instrumental to establish experimental setups in which the host is lacking defined functionalities. Combining these with multi-generational analysis of the type and stability of changes in the host and its microbiome should provide direct assessment of the validity of this hypothesis.

While the above evaluations should ultimately include a variety of different organisms, it may be useful to focus on organisms enabling rapid and rigorous assessment. One example is the fruit fly, *D. melanogaster*, which contains a relatively small number of (culturable) bacterial species (Wong et al., 2011). Additionally, the gut microbiome of the fly can be easily manipulated without relying on anti-microbial agents that could also have a direct impact on host tissues (Ridley et al., 2013). Avoiding these direct impacts is crucial for creating defined setups enabling rigorous evaluation of the influence of microbial changes on the host (Ridley et al., 2013; Fridmann-Sirkis et al., 2014). It is also important to devote attention to the potential impact of bacterial manipulations on the germline. Such impact can generate phenotypic differences between parents and offspring (Fridmann-Sirkis et al., 2014) thereby invalidating certain comparisons between non-matched generations. Avoiding these potentially confounding effects is achieved in flies by dissolving and sterilizing the chorionic shell of fertilized eggs (without antibiotics or propagation under germ-free conditions for an uncharacterized number of generations). Larvae that hatch from these dechorionated eggs are devoid of gut bacteria and can be re-exposed to defined compositions of bacterial species (e.g., bacteria that were isolated from intact flies with or without additional manipulations). These useful features synergize with the obvious benefits of having a large knowledgebase, powerful genetic tools and generation time which is compatible with practical analysis over about 10 generations.

The outcome of these (and similar) studies may shed important light on two critical questions in gene regulation and evolution, namely: (1) How the balance between stability and flexibility of developing organisms can be regulated by the environment based on contextual demand, and (2) Whether and how epigenetic- and symbiotic-based responses to new stressful conditions are causally connected to longer-term establishment of genetic adaptations in the host. This includes evaluation of the scope and mechanisms by which the adaptive potential of the microbiome might contribute to rapid buildup of stable adaptive modifications in the host. Following these lines of investigation may uncover hitherto unrecognized mechanisms bridging ecological and evolutionary processes.

REFERENCES

Arora, T., and Sharma, R. (2011). Fermentation potential of the gut microbiome: implications for energy homeostasis and weight management. *Nutr. Rev.* 69, 99–106. doi: 10.1111/j.1753-4887.2010.00365.x

Atkins, P. W., and de Paula, J. (2002). *Physical Chemistry*, 7th Edn. Oxford: Oxford University Press.

Backhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., et al. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15718–15723. doi: 10.1073/pnas.0407076101

Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920. doi: 10.1126/science.1104816

Badola, P., Kumar, V. R., and Kulkarni, B. D. (1991). Effects of coupling nonlinear-systems with complex dynamics. *Phys. Lett. A* 155, 365–372. doi: 10.1016/0375-9601(91)91041-B

Behar, A., Yuval, B., and Jurkevitch, E. (2008). Gut bacterial communities in the Mediterranean fruit fly (*Ceratitis capitata*) and their impact on host longevity. *J. Insect Physiol.* 54, 1377–1383. doi: 10.1016/j.jinsphys.2008.07.011

Ben-Yosef, M., Aharon, Y., Jurkevitch, E., and Yuval, B. (2010). Give us the tools and we will do the job: symbiotic bacteria affect olive fly fitness in a diet-dependent fashion. *Proc. Biol. Sci.* 277, 1545–1552. doi: 10.1098/rspb.2009.2102

Bright, M., and Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* 8, 218–230. doi: 10.1038/nrmicro2262

Brummel, T., Ching, A., Seroude, L., Simon, A. F., and Benzer, S. (2004). *Drosophila* lifespan enhancement by exogenous bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12974–12979. doi: 10.1073/pnas.0405207101

Brune, A., and Friedrich, M. (2000). Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* 3, 263–269. doi: 10.1016/S1369-5274(00)00087-4

Cavanaugh, C. M., McKiness, Z. P., Newton, L. G., and Stewart, F. J. (2006). "Marine chemosynthetic symbioses," in *The Prokaryotes,* Vol. 1, *Symbiotic Associations, Biotechnology, Applied Microbiology,* eds M. Dworkin, E. Reosenberg, K.-H. Schleifer, and E. Stackebrandt (New york: Springer-Verlag).

Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* 6, 725–740. doi: 10.1038/nrmicro1992

Flatt, T. (2005). The evolutionary genetics of canalization. *Q. Rev. Biol.* 80, 287–316. doi: 10.1086/432265

Flint, H. J., Scott, K. P., Louis, P., and Duncan, S. H. (2012). The role of the gut microbiota in nutrition and health. *Nat. Rev. Gastroenterol. Hepatol.* 9, 577–589. doi: 10.1038/nrgastro.2012.156

Fridmann-Sirkis, Y., Stern, S., Elgart, M., Galili, M., Zeisel, A., Shental, A., et al. (2014). Delayed development induced by toxicity to the host can be inherited by a bacterial-dependent, transgenerational effect. *Front. Genet.* 5:27. doi: 10.3389/fgene.2014.00027

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., et al. (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469, 543–547. doi: 10.1038/nature09646

Gangaraju, V. K., Yin, H., Weiner, M. M., Wang, J., Huang, X. A., and Lin, H. (2011). *Drosophila* Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat. Genet.* 43, 153–158. doi: 10.1038/ng.743

Glass, L., and MacKey, M. C. (1988). *From Clocks to Chaos: The Rhythms of Life.* Princeton: Princeton University Press.

González-Miranda, J. M. (2004). *Synchronization and Control of Chaos – An Introduction for Scientists and Engineers.* London: Imperial College Press.

Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nat. Genet.* 38(Suppl.), S20–S24. doi: 10.1038/ng1803

Jablonka, E. (2013). Epigenetic inheritance and plasticity: the responsive germline. *Prog. Biophys. Mol. Biol.* 111, 99–107. doi: 10.1016/j.pbiomolbio.2012.08.014

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84, 131–176. doi: 10.1086/598822

Jaenike, J. (2012). Population genetics of beneficial heritable symbionts. *Trends Ecol. Evol.* 27, 226–232. doi: 10.1016/j.tree.2011.10.005

Kamada, N., Kim, Y. G., Sham, H. P., Vallance, B. A., Puente, J. L., Martens, E. C., et al. (2012). Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* 336, 1325–1329. doi: 10.1126/science.1222195

Kranich, J., Maslowski, K. M., and Mackay, C. R. (2011). Commensal flora and the regulation of inflammatory and autoimmune responses. *Semin. Immunol.* 23, 139–145. doi: 10.1016/j.smim.2011.01.011

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence.* Berlin: NY: Springer-Verlag. doi: 10.1007/978-3-642-69689-3

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D., and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr. Opin. Biotechnol.* 24, 160–168. doi: 10.1016/j.copbio.2012.08.005

Lee, N., Maurange, C., Ringrose, L., and Paro, R. (2005). Suppression of Polycomb group proteins by JNK signalling induces transdetermination in *Drosophila* imaginal discs. *Nature* 438, 234–237. doi: 10.1038/nature04120

Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S., and Carthew, R. W. (2009). A microRNA imparts robustness against environmental fluctuation during development. *Cell* 137, 273–282. doi: 10.1016/j.cell.2009.01.058

Lim, J. P., and Brunet, A. (2013). Bridging the transgenerational gap with epigenetic memory. *Trends Genet.* 29, 176–186. doi: 10.1016/j.tig.2012.12.008

Littman, D. R., and Pamer, E. G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host Microbe* 10, 311–323. doi: 10.1016/j.chom.2011.10.004

Markle, J. G., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., et al. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* 339, 1084–1088. doi: 10.1126/science.1233521

Martin, R., Miquel, S., Ulmer, J., Kechaou, N., Langella, P., and Bermudez-Humaran, L. G. (2013). Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. *Microb. Cell Fact.* 12:71. doi: 10.1186/1475-2859-12-71

Milton, C. C., Ulane, C. M., and Rutherford, S. (2006). Control of canalization and evolvability by Hsp90. *PLoS ONE* 1:e75. doi: 10.1371/journal.pone.0000075

Negri, I., Franchini, A., Gonella, E., Daffonchio, D., Mazzoglio, P. J., Mandrioli, M., et al. (2009). Unravelling the *Wolbachia* evolutionary role: the reprogramming of the host genomic imprinting. *Proc. Biol. Sci.* 276, 2485–2491. doi: 10.1098/rspb.2009.0324

Negri, I., Pellecchia, M., Mazzoglio, P. J., Patetta, A., and Alma, A. (2006). Feminizing *Wolbachia* in *Zyginidia pullula* (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. *Proc. Biol. Sci.* 273, 2409–2416. doi: 10.1098/rspb.2006.3592

Otsuka, K. (1991). Complex dynamics in coupled nonlinear element systems. *Int. J. Mod. Phys. B* 5, 1179–1214. doi: 10.1142/S0217979291000572

Ott, J. A. (1982). New mouthless interstitial worms from the sulfide system: symbiosis with Prokaryotes. *Mar. Ecol.* 3, 313–333. doi: 10.1111/j.1439-0485.1982.tb00282.x

Queitsch, C., Sangster, T. A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* 417, 618–624. doi: 10.1038/nature749

Ridley, E. V., Wong, A. C., and Douglas, A. E. (2013). Microbe-dependent and nonspecific effects of procedures to eliminate the resident microbiota from *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 79, 3209–3214. doi: 10.1128/AEM.00206-13

Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E. C. (2005). The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev.* 19, 2947–2952. doi: 10.1101/gad.1372505

Rosenberg, E., Sharon, G., and Zilber-Rosenberg, I. (2009). The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. *Environ. Microbiol.* 11, 2959–2962. doi: 10.1111/j.1462-2920.2009.01995.x

Rosenberg, E., and Zilber-Rosenberg, I. (2011). Symbiosis and development: the hologenome concept. *Birth Defects Res. C Embryo Today* 93, 56–66. doi: 10.1002/bdrc.20196

Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323. doi: 10.1038/nri2515

Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342. doi: 10.1038/24550

Salathia, N., and Queitsch, C. (2007). Molecular mechanisms of canalization: Hsp90 and beyond. *J. Biosci.* 32, 457–463. doi: 10.1007/s12038-007-0045-9

Samakovli, D., Thanou, A., Valmas, C., and Hatzopoulos, P. (2007). Hsp90 canalizes developmental perturbation. *J. Exp. Bot.* 58, 3513–3524. doi: 10.1093/jxb/erm191

Sarkar, B. C., Dandapathak, M., Sarkar, S., and Banerjee, T. (2013). Studies on the dynamics of two bilaterally coupled periodic gunn oscillators using melnikov techniques. *Prog. Electromagn. Res. M* 28, 213–228. doi: 10.2528/PIERM12120316

Sawarkar, R., and Paro, R. (2010). Interpretation of developmental signaling at chromatin: the Polycomb perspective. *Dev. Cell* 19, 651–661. doi: 10.1016/j.devcel.2010.10.012

Schmalhausen, I. I. (1949). *Factors of Evolution: The Theory of Stabilizing Selection.* Philadelphia: Blakiston Co (reprinted in 1987 by Chicago University Press).

Sgro, C. M., Wegener, B., and Hoffmann, A. A. (2010). A naturally occurring variant of Hsp90 that is associated with decanalization. *Proc. Biol. Sci.* 277, 2049–2057. doi: 10.1098/rspb.2010.0008

Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010a). Commensal bacteria play a role in mating preference

of *Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A.* 107, 20051–20056. doi: 10.1073/pnas.1009906107

Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010b). Commensal bacteria play a role in mating preference of *Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A.* 107, 20051–20056. doi: 10.1073/pnas.1009906107

Shin, S. C., Kim, S.-H., You, H., Kim, B., Kim, A. C., Lee, K.-A., et al. (2011). *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334, 670–674. doi: 10.1126/science.1212782

Siegal, M. L., and Bergman, A. (2002). Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10528–10532. doi: 10.1073/pnas.102303999

Specchia, V., Piacentini, L., Tritto, P., Fanti, L., D'Alessandro, R., Palumbo, G., et al. (2010). Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* 463, 662–665. doi: 10.1038/nature08739

Starr, D. J., and Cline, T. W. (2002). A host parasite interaction rescues *Drosophila* oogenesis defects. *Nature* 418, 76–79. doi: 10.1038/nature00843

Stern, S., Fridmann-Sirkis, Y., Braun, E., and Soen, Y. (2012). Epigenetically heritable alteration of fly development in response to toxic challenge. *Cell Rep.* 1, 528–542. doi: 10.1016/j.celrep.2012.03.012

Stern, S., Snir, O., Mizrachi, E., Galili, M., Zaltsman, I., and Soen, Y. (2014). Reduction in maternal Polycomb levels contributes to transgenerational inheritance of a response to toxic stress in flies. *J. Physiol.* doi: 10.1113/jphysiol.2014.271445 [Epub ahead of print].

Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J., and Leulier, F. (2011). *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab.* 14, 403–414. doi: 10.1016/j.cmet.2011.07.012

Stouthamer, R., Breeuwer, J. A., and Hurst, G. D. (1999). *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53, 71–102. doi: 10.1146/annurev.micro.53.1.71

Takahashi, K. (2014). Influence of bacteria on epigenetic gene control. *Cell. Mol. Life Sci.* 71, 1045–1054. doi: 10.1007/s00018-013-1487-x

Takahashi, K., Sugi, Y., Nakano, K., Tsuda, M., Kurihara, K., Hosono, A., et al. (2011). Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. *J. Biol. Chem.* 286, 35755–35762. doi: 10.1074/jbc.M111.271007

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565. doi: 10.1038/150563a0

Waddington, C. H. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology.* London: Allen & Unwin.

Wang, J., Li, F., Sun, R., Gao, X., Wei, H., Li, L. J., et al. (2013). Bacterial colonization dampens influenza-mediated acute lung injury via induction of M2 alveolar macrophages. *Nat. Commun.* 4:2106. doi: 10.1038/ncomms3106

Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472, 57–63. doi: 10.1038/nature09922

Werren, J. H. (1997). Biology of *Wolbachia. Annu. Rev. Entomol.* 42, 587–609. doi: 10.1146/annurev.ento.42.1.587

Wong, C. N. A., Ng, P., and Douglas, A. E. (2011). Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster. Environ. Microbiol.* 13, 1889–1900. doi: 10.1111/j.1462-2920.2011.02511.x

Wu, C.-I., Shen, Y., and Tang, T. (2009). Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res.* 19, 734–743. doi: 10.1101/gr.084640.108

Yamanaka, M., Nomura, T., and Kametaka, M. (1977). Influence of intestinal microbes on heat production in germ-free, gnotobiotic and conventional mice. *J. Nutr. Sci. Vitaminol.* 23, 221–226. doi: 10.3177/jnsv.23.221

Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* 32, 723–735. doi: 10.1111/j.1574-6976.2008.00123.x

Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure

Sara H. Sinclair [1,2,3,4], **Kristen E. Rennoll-Bankert** [2,3] **and J. S. Dumler** [1,2,3,4] *

[1] Graduate Program in Cellular and Molecular Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2] Department of Microbiology and Immunology, School of Medicine, University of Maryland Baltimore, Baltimore, MD, USA
[3] Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[4] Department of Pathology, School of Medicine, University of Maryland Baltimore, Baltimore, MD, USA

Obligate intracellular pathogenic bacteria evolved to manipulate their host cells with a limited range of proteins constrained by their compact genomes. The harsh environment of a phagocytic defense cell is one that challenges the majority of commensal and pathogenic bacteria; yet, these are the obligatory vertebrate homes for important pathogenic species in the *Anaplasmataceae* family. Survival requires that the parasite fundamentally alter the native functions of the cell to allow its entry, intracellular replication, and transmission to a hematophagous arthropod. The small genomic repertoires encode several eukaryotic-like proteins, including ankyrin A (AnkA) of *Anaplasma phagocytophilum* and Ank200 and tandem-repeat containing proteins of *Ehrlichia chaffeensis* that localize to the host cell nucleus and directly bind DNA. As a model, *A. phagocytophilum* AnkA appears to directly alter host cell gene expression by recruiting chromatin modifying enzymes such as histone deacetylases and methyltransferases or by acting directly on transcription in *cis*. While *cis* binding could feasibly alter limited ranges of genes and cellular functions, the complex and dramatic alterations in transcription observed with infection are difficult to explain on the basis of individually targeted genes. We hypothesize that nucleomodulins can act broadly, even genome-wide, to affect entire chromosomal neighborhoods and topologically associating chromatin domains by recruiting chromatin remodeling complexes or by altering the folding patterns of chromatin that bring distant regulatory regions together to coordinate control of transcriptional reprogramming. This review focuses on the *A. phagocytophilum* nucleomodulin AnkA, how it impacts host cell transcriptional responses, and current investigations that seek to determine how these multifunctional eukaryotic-like proteins facilitate epigenetic alterations and cellular reprogramming at the chromosomal level.

Keywords: epigenetics, nucleomodulin, DNA methylation, histone deacetylase, chromatin, *Anaplasma phagocytophilum*

INTRODUCTION

In order to infect mammalian hosts, bacterial pathogens evolved an array of mechanisms that serve to create an environment conducive for survival, replication, and spread. While many bacterial species survive in an extracellular environment, intracellular pathogens must be capable of both entering their host cells undetected and altering the cellular milieu in order to replicate. Traditionally, the ability of bacterial-derived proteins to induce disruptions of cell signaling or major cellular processes such as NF-κB, MAPK, and JAK/STAT pathways, are the predominant focus of host–pathogen interaction studies (Brodsky and Medzhitov, 2009). Recently, there has been an increasing interest in the ability of these intracellular pathogens to direct alterations in host cell gene expression that promote survival and replication (Paschos and Allday, 2010; Bierne et al., 2012; Silmon de Monerri and Kim, 2014). It is now well recognized that bacterial pathogens can reprogram host gene expression either directly or indirectly, by altering the accessibility of gene promoters via epigenetic modifications.

Eukaryotic DNA is highly organized and gene expression tightly regulated by an orchestrated network of proteins, RNAs, and other modulators. Histone octamers, or nucleosomes, organize DNA by acting as a bobbin on which the DNA winds. The charge of the histone proteins can be covalently modified in order to more tightly or loosely associate with DNA. Histone acetylation, which imparts a negative charge, is predominantly associated with an open configuration where promoters are easily accessed by RNA polymerases. Histone methylation or phosphorylation which impart a positive charge, cause DNA to more tightly associate with histone proteins, reducing promoter accessibility to transcription activating machinery. The process of modulating open and closed chromatin is further complicated by (i) the residue(s) of which histone protein(s) is/are modified, (ii) methylation of cytosine residues in DNA, and (iii) non-coding RNAs, and by the manner in which these mechanisms are intertwined. It therefore is no surprise that bacterial-derived proteins have evolved to interfere with host gene expression that improves bacterial fitness.

Over the last decade, examples of secreted bacterial effector proteins, ranging from those produced by *Listeria monocytogenes*, *Chlamydia trachomatis*, *Shigella flexneri*, and others, were found to target the host cell nucleus (Paschos and Allday, 2010; Bierne et al., 2012; Silmon de Monerri and Kim, 2014). These discoveries demonstrate the relationship of altered transcription and function of infected host cells and microbial survival, and in some cases, pathogenicity. *S. flexneri*, a bacterial pathogen that can cause dysentery, prevents NF-κB from binding its target gene promoters by altering the phosphorylation state of histone H3 at serine 10. The bacterium does so by secreting outer membrane protein F (OspF) that de-phosphorylates MAPKs in the nucleus resulting in a lack of histone H3 phosphorylation at serine 10 at a number of NF-κB-dependent genes (Arbibe et al., 2007). By altering NF-κB target gene transcription, *S. flexneri* suppresses the host cell inflammatory response promoting microbial survival and transmission (Arbibe et al., 2007). *L. monocytogenes* nuclear targeted protein A (LntA), blocks binding of heterochromatin inducing protein BAHD1 at interferon-stimulated genes resulting in upregulated expression (Rohde, 2011). Nuclear effector E (NUE) of *C. trachomatis* and RomA of *L. pneumophila* have methyltransferase activity and induce methylation of eukaryotic histones and altered host cell gene expression (Pennini et al., 2010; Rolando et al., 2013). Plant pathogens in the genus *Xanthomonas*, as well as *Ralstonia solanacearum*, *Burkholderia rhizoxinica*, and an endosymbiont of the plant pathogenic fungus *Rhizopus microsporus* all possess genes encoding transcription activator-like (TAL) protein effectors that when expressed, bind plant DNA and alter transcription and disease susceptibility (Sugio et al., 2007; Bogdanove, 2014). These examples are part of an expanding array of bacterial-derived proteins termed nucleomodulins that target host cell chromatin or chromatin-linked pathways to alter transcription, typically at one or a few host genes. To date, the only prokaryotic nucleomodulins shown to directly bind mammalian DNA and influence surrounding chromatin are from the *Anaplasmataceae* family. Ankyrin A (AnkA) of *Anaplasma phagocytophilum*, as well as Ank200 and several tandem-repeat containing proteins (TRPs) from *Ehrlichia chaffeensis*, have been shown to enter the nucleus and bind DNA, and interact with host epigenetic machinery or alter nearby histone octamers (Garcia-Garcia et al., 2009a; Luo et al., 2011; Zhu et al., 2011; Luo and McBride, 2012; Dunphy et al., 2013).

Targeting of individual genes or binding of an effector at a single chromatin region or small numbers of promoter loci could, in theory, lead to *cis* regulation of those loci and in part explain transcriptional alterations induced by infection. However, the degree of transcriptional alterations and the coordination of these events that dramatically affect cellular function programs are unlikely to be explained by individual targets given the complexity and repertoire of the human genome, at approximately 3251 Mb vs. genome sizes of *S. flexneri* (≤4.83 Mb; Jin et al., 2002), *C. trachomatis* (≤1.04 Mb; Stephens et al., 1998), *L. monocytogenes* (≤3.00 Mb; Evans et al., 2004), *L. pneumophila* (≤3.40 Mb; Chien et al., 2004), and *A. phagocytophilum* (≤1.47 Mb; Lin et al., 2011). We hypothesize that nucleomodulins can act broadly, even genome-wide, to affect entire chromosomal neighborhoods and topologically associating chromatin domains by recruiting chromatin remodeling complexes or by altering the folding patterns of chromatin that bring distant regulatory regions together to coordinate control of transcriptional reprogramming. This review will discuss the current knowledge of *A. phagocytophilum* subversion of host cells by nucleomodulins and how AnkA could play an even larger role than its documented effects on transcription.

Anaplasma phagocytophilum

Anaplasma phagocytophilum, transmitted by *Ixodes* spp. ticks, was discovered as the causative agent of human granulocytic anaplasmosis (HGA) in 1990 (Chen et al., 1994). While the infection is usually subclinical, manifestations in humans range from mild fever to severe infection requiring intensive care or even death (Chen et al., 1994). *A. phagocytophilum* is a Gram negative, obligate intracellular bacterium that is a parasite of neutrophils in mammalian hosts. Owing to their roles in protective inflammatory and immune responses toward microbial infections and their innate ability to recognize and kill pathogens, neutrophils are unlikely host cells for any microorganism. Yet, *A. phagocytophilum* colonizes these cells and thwarts their normal functions to create a hospitable environment for intracellular replication and subsequent transmission via tick bite. With a limited genome of approximately 1.5 Mb (Lin et al., 2011) the potential genetic reservoir for controlling an infected host cell, whether in the tick or mammal, is only a fraction of the tick or human genome. This simple observation suggests that to circumvent such extremes, its control mechanisms must be highly efficient, multifunctional, or target master regulators or similar checkpoints in eukaryotic cells.

One remarkable feature of *A. phagocytophilum*-infected cells is the marked change in transcriptional profiles that belies aberrant regulation of many key host pathways (Borjesson et al., 2005; de la Fuente et al., 2005; Lee et al., 2008). *A. phagocytophilum*-infected neutrophils are characterized by an "activated-deactivated" phenotype with major functional aberrations including decreased respiratory burst, delayed apoptosis, reduced transmigration across endothelial cell barriers, and decreased antimicrobial activities such as phagocytosis and microbial killing, while simultaneous increases in degranulation of vesicle contents including proteases and increased production of chemokines, are observed (Carlyon et al., 2002; Carlyon and Fikrig, 2003; Choi et al., 2005; Dumler et al., 2005; Garyu et al., 2005; Ge et al., 2005; Carlyon and Fikrig, 2006; Garcia-Garcia et al., 2009b). The net result of these changes is an increase in (i) inflammatory responses that recruit new host cells, (ii) prolonged survival of infected cells, (iii) an inability to kill internalized microbes, and (iv) net sequestration of infected cells within the intravascular compartment that is more readily accessed by the next tick bite (Rennoll-Bankert and Dumler, 2012). In fact, interference with any of these processes using *in vitro* and *in vivo* models leads to reduced fitness of the microbe underscoring how important these functional changes are in mammalian hosts.

The biological basis for such dramatic changes in neutrophil function is increasingly studied by methods that range from examination of individual pathways to genome-wide systems biology approaches. Transcriptional profiling of infected neutrophils and HL-60 cells, the latter a commonly used cell model for *A. phagocytophilum* infection, reveals altered transcription genome-wide,

confirming that changes in transcription are not restricted to a few genes and limited cellular functions, but likely play a role in most of the known functional changes induced by infection (Borjesson et al., 2005; de la Fuente et al., 2005; Lee et al., 2008). Transcriptional downregulation of two components of the NADPH oxidase, *CYBB* encoding NOX2 or gp91[phox] and the GTPase *RAC2*, that assemble in the membranes of activated cells to generate superoxide production and promote microbial killing plays a role in prolonged inhibition of respiratory burst (Banerjee et al., 2000; Carlyon et al., 2002). In addition, delayed apoptosis is achieved by maintained transcription of *BCL2* family members (Choi et al., 2005; Ge et al., 2005) and increased proinflammatory responses are due to upregulated transcription of cytokine and chemokine genes such as *IL8* (Klein et al., 2000; Akkoyunlu et al., 2001; Scorpio et al., 2004).

While many studies document these changes at functional and transcriptional levels, the precise mechanisms that organize and coordinate alterations to support improved *A. phagocytophilum* fitness are not well addressed. The biggest advance in understanding these processes occurred with the discovery of secreted prokaryotic effector proteins and their secretion systems. The genome of *A. phagocytophilum* encodes a type IV secretion system (T4SS) that allows protein effectors to translocate into infected host cells where they likely act by mechanisms similar to those of other bacterial secreted effectors that target cytosolic pathways such as signal transduction, cytoskeletal rearrangements, intracellular trafficking, etc. (Ohashi et al., 2002; IJdo et al., 2007; Gillespie et al., 2009). These observations also hold true for *A. phagocytophilum* effectors. However, not all effector proteins remain localized within the host cytosol, and those that enter the nucleus have direct access to genes as well as to a distinct and diverse array of proteins that could impact cellular function on a global scale.

AnkA OF *Anaplasma phagocytophilum*

Anaplasma phagocytophilum expresses a number of major immunoreactive proteins, including major surface protein 2/p44 (Msp2/p44) and AnkA, the latter of which was found by Caturegli et al. (2000) using immunoelectron microscopy to be transported into the host cell nucleus and bound within heterochromatin (Park et al., 2004). AnkA contains many EPIYA motifs that become tyrosine phosphorylated and recruit SHP-1 and ABL1 when introduced into mammalian cells which in turn regulates endosomal entry and intracellular infection (IJdo et al., 2007; Lin et al., 2011). Aside from this, AnkA contains multiple eukaryotic motifs including 8–15 or more ankyrin repeats, a putative bipartite nuclear localization signal (NLS), and a putative high mobility group N-chromatin unfolding domain (HMGN-CHUD; Caturegli et al., 2000). The ankyrin repeat domains are organized tandemly, creating highly stable spring-like structures that allow protein–protein or protein–DNA interactions and are commonly found in transcription factors and their regulatory proteins (Bork, 1993; Jernigan and Bordenstein, 2014). HMGN-CHUD domains facilitate binding to nucleosomes to alter chromatin structure and transcription of surrounding genes (Bustin, 1999). The presence of these motifs lends credence to the idea that AnkA plays a role in altering neutrophil transcription.

Park et al. (2004) confirmed that AnkA directly binds both DNA and nuclear proteins, and provided limited evidence of its capacity to bind broadly throughout the human genome. AnkA binds to regions at the promoter of *CYBB*, encoding the gp91[phox] component of phagocyte oxidase (Garcia-Garcia et al., 2009b). *CYBB* is known to be transcriptionally repressed with *A. phagocytophilum* infection, further suggesting that AnkA acts in *cis* to alter gene transcription (Thomas et al., 2005; Garcia-Garcia et al., 2009b). Moreover, transcription of *CYBB* is decreased in a dose dependent manner as nuclear AnkA concentrations increase (Garcia-Garcia et al., 2009b). When AnkA-expressing plasmids are transfected into HL-60 cells, *CYBB* expression is dampened, strengthening the link to transcriptional regulation (Garcia-Garcia et al., 2009b). In electrophoretic mobility shift assays, AnkA binds to the *CYBB* promoter at the same locations as other known transcriptional regulators such as CAATT displacement protein (CDP) and special-AT rich binding protein-1 (SATB1; Thomas et al., 2005; Garcia-Garcia et al., 2009b; Rennoll-Bankert and Dumler, 2012). Surprisingly, DNA binding by AnkA does not target a conserved DNA sequence motif or signature; rather it binds to regions rich in AT nucleotides that have specific structural qualities, including the ability to uncoil under superhelical stress (Park et al., 2004; Garcia-Garcia et al., 2009b). The latter feature is often observed in eukaryotic proteins that tether DNA strands together into the nuclear matrix at matrix attachment regions (MARs) to regulate transcription from distantly located but functionally related chromosomal regions; proteins that bind these regions are called MAR-binding proteins (MARBPs).

Interestingly, Garcia-Garcia et al. (2009a) showed that multiple downregulated defense genes in infected granulocytes were clustered on chromosomes. The spatial clustering of similarly regulated genes suggests that, in addition to *cis*-regulation as with *CYBB*, long ranges of chromosomes are affected – a process often attributed to epigenetic mechanisms such as DNA methylation and histone chromatin modifications. Using chromatin immunoprecipitation (ChIP) to investigate histone marks at defense gene promoters, acetylation of histone H3 was dramatically reduced, a finding often associated with silenced gene transcription (Garcia-Garcia et al., 2009a). To explain this, binding of histone deacetylase-1 (HDAC1) was found increased across multiple defense gene promoters (Garcia-Garcia et al., 2009a). Moreover, *A. phagocytophilum*-infected granulocytes have increased HDAC activity most likely explained by the increased quantity of both HDAC1 and HDAC2. Inhibition of *HDAC1*, but not *HDAC2* expression by siRNA or pharmacologic inhibition of HDAC activity impairs *A. phagocytophilum* propagation, whereas overexpression leads to increased intracellular propagation (Garcia-Garcia et al., 2009a). We modeled this process by using the wild type *CYBB* promoter, to which AnkA binds, and mutated forms unable to bind AnkA in order to demonstrate that AnkA binding leads to HDAC1 recruitment and silencing of expression at *CYBB* (Garcia-Garcia et al., 2009b). These results clearly link AnkA with HDAC1 toward facilitating widespread downregulation of antimicrobial responses.

Currently, most data examining the ability of HDAC1 and AnkA to modulate transcription focuses on a small number of loci,

including *CYBB* and up to 17 other gene promoters. While AnkA and HDAC1 are important contributors to transcriptional down-regulation of some genes in *A. phagocytophilum*-infected cells, neither has been determined to provide a broad mechanism that coordinates the transcriptional or functional alterations observed. Importantly, transcriptional profiling of *A. phagocytophilum*-infected granulocytes, including both primary neutrophils and cell lines, demonstrates that the majority of differentially expressed genes are upregulated (Borjesson et al., 2005). Thus, while HDAC1 recruitment by AnkA is important for down-regulating some genes, the majority of DEGs are likely regulated by alternative or additional mechanisms.

Of considerable interest is the interplay between chromatin structure and DNA methylation. Histone modifications, DNA methylation at CpG islands, and binding of methyl CpG DNA-binding proteins (MeCPs) via their methyl-DNA binding domains (MBDs) occur in concert during biological responses including neoplasia and cellular differentiation (Baylin and Herman, 2000; Baylin et al., 2001; Stirzaker et al., 2004; Srivastava et al., 2010; Reddington et al., 2013). Cytosine residues of CpG dinucleotides can by methylated by DNA methyltransferases (DNMTs) leading to sustained DNA methylation over multiple generations of cell divisions. It is believed in part that the methylation protects genes and transcriptional programs from inappropriate ectopic expression as cells enter various stages of tissue differentiation (Srivastava et al., 2010). DNMT1 is highly conserved and is thought to be primarily responsible for maintaining existing methyl CpGs (Baylin et al., 2001). In contrast, DNMT3a and 3b are believed to undergo transient *de novo* expression that induces CpG methylation in response to cellular stimuli (Okano et al., 1999). MBDs can recognize these newly methylated regions of DNA and are often associated with HDACs in MeCP1 and MeCP2 complexes, illustrating the potential for a direct link between DNA methylation and alterations in chromatin structure (Jones et al., 1998; Nan et al., 1998; Ng et al., 1999). Generally, hypermethylation of gene promoters occurs synchronously with histone deacetylation and decreased gene expression of many tumor suppressor genes (Baylin and Herman, 2000; Baylin et al., 2001; Stirzaker et al., 2004). This complex interplay has been best studied in cancer genomes, where it is currently unclear whether DNA methylation induces alterations in chromatin structure or if chromatin structure induces changes in DNA methylation. However, given the interplay between HDAC activity and DNA methylation, it is not unreasonable to hypothesize that *A. phagocytophilum* infection induces hypermethylation of the host genome during the process of global transcriptome reprogramming (**Figure 1**).

In addition to chromatin alterations induced by HDACs, MARs are responsible for dictating the three-dimensional architecture of chromatin loops and serve as tethering points for DNA to the nuclear matrix (Spector, 2003; Kumar et al., 2007; Kohwi-Shigematsu et al., 2012). Arranging chromatin into loops allows transcription factors attached to the matrix access to promoters and brings distal genomic loci and regulatory regions into a position of proximity for coordinated regulation (Arope et al., 2013). Interestingly, the AT-rich DNA docking sites of AnkA are similar to those of MARs. In fact, several transcription factors
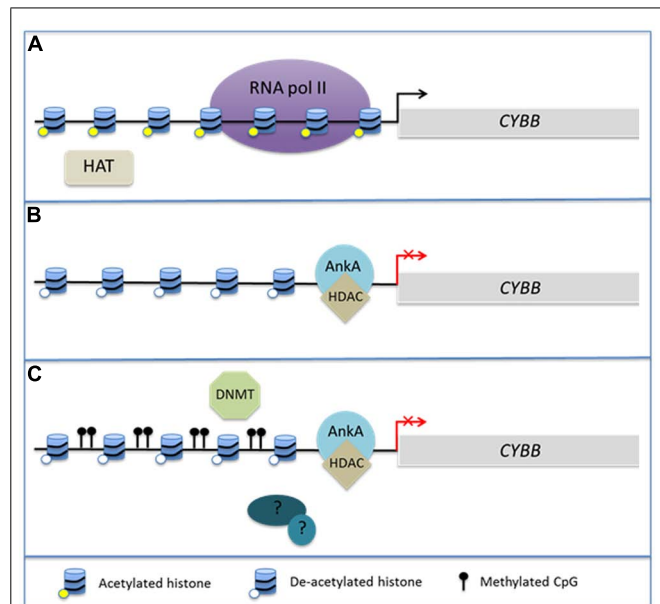
FIGURE 1 | AnkA alters chromatin structure at the *CYBB* promoter. During *Anaplasma phagocytophilum* infection, AnkA accumulates in the host cell nucleus where it can directly bind DNA and influence the transcription of *CYBB*, which encodes the gp91phox component of the NADPH oxidase. **(A)** In the absence of infection, *CYBB* is activated by inflammatory signals and is easily transcribed. **(B)** During infection, AnkA binds the proximal promoter and recruits HDAC which deacetylates nearby histones in order to inhibit transcription. **(C)** We propose that additional host or bacterial-derived chromatin modifying enzymes, such as DNMT methylation of host CpGs, may also be involved in altering the host epigenome to the bacterium's advantage.

including SATB1, bind MARs to coordinately modulate transcription from large genomic regions (Kumar et al., 2007). Like AnkA, SATB1 occupies the *CYBB* promoter during myeloid differentiation where it represses the transcription of *CYBB*, is involved in maintaining expression of *BCL2* (Gong et al., 2010) which is transcriptionally sustained during *A. phagocytophilum* infection (Ge et al., 2005) and is implicated in activation of cytokine expression in T cells (Hawkins et al., 2001; Beyer et al., 2011). SATB1 interactions with chromatin remodeling complexes include HDACs which are likely involved in some of these processes (Yasui et al., 2002). The complex looping of chromatin facilitated by MARs and MAR binding proteins allows for entire chromatin domains and territories to be remodeled despite relatively few binding sites (Cai et al., 2006). For example, SATB1 binds to only nine regions across the 200 kb T_H2 locus, yet it is a major regulator of T_H2 lymphocyte differentiation and function (Cai et al., 2006). Complex protein interactions involving chromatin remodeling and histone modification mediated by anchoring proteins like SATB1 make it plausible that DNA-binding bacterial nucleomodulins such as AnkA could target broad transcriptional programs that belie functions of host cells. While only limited data currently exist, AnkA binds to at least 23 distinct sites on 12 separate chromosomes with *A. phagocytophilum* infection of the human HL-60 promyelocytic cell line (Park et al., 2004) detailed mapping of AnkA genomic binding sites is now

in progress. We therefore think it is plausible that AnkA binds throughout the genome and exerts its effects, like SATB1, to both repress and activate transcription by tethering chromatin to the nuclear matrix and exposing promoters to chromatin remodeling complexes.

Anaplasma phagocytophilum INFECTION IS ASSOCIATED WITH TRANSCRIPTIONAL CHANGES OVER MEGABASES IN THE HUMAN GENOME

As a relatively new field, studies that examine the epigenome of cells infected by or as a consequence of bacterial infection, whether parasitic or symbiotic, are few and far between. Given the proven interaction of *A. phagocytophilum* AnkA with gene promoters, epigenetic alterations of nearby chromatin, and its MAR-binding attributes, we reanalyzed publicly available transcription microarray data generated from *A. phagocytophilum*-infected human peripheral blood neutrophils (Borjesson et al., 2005). We used updated bioinformatics tools, including an analysis of the SD estimates for differential transcription at each locus, and wherever possible, improved gene and gene feature localization using the NCBI GRCh38 human genome assembly (release date December 24, 2013). The differential expression of all 18,400 interrogated transcripts and variants, including 14,500 well-characterized human genes, was mapped by chromosomal position so that the relationship between the linear chromosome landscape and transcriptional activity could be visualized and investigated. To display physical locations on chromosomes with long regions of altered transcription, the fold change of expression for each locus was averaged with the nearest neighboring genes to create sliding windows for each chromosome. Overall, the median sliding window included 10.0 genes/gene features (IQR 2.0) spanning a median interval of 2.04 Mb (IQR 2.16) corresponding approximately to the lower end dimensions of chromosomal or gene expression dysregulation domains (Letourneau et al., 2014). The sliding window average along with individual gene transcriptional fold changes were plotted for each chromosome. To assess significance of fold changes in the sliding window, a similar sliding window was created to simultaneously show the average estimated SDs over each window.

All chromosomes showed linear regions of marked differential transcription covering megabases in scale on both p and q arms (**Figure 2**). In keeping with the original observations of Borjesson et al., most clusters were upregulated; however, the analysis also demonstrated a similar pattern of downregulated genes over long linear regions. Of interest was chromosome 17 with a 9.4 Mb cluster containing myeloid peroxidase (*MPO*), eosinophil peroxidase (*EPX*), and lactoperoxidase (*LPO*) which showed marked downregulation, and a 3.2 Mb cluster containing multiple upregulated chemokine genes (**Figure 3**). In addition to reduced transcript levels, *MPO* and *EPX* promoters were previously shown to be deacetylated upon *A. phagocytophilum* infection (Garcia-Garcia et al., 2009a). Furthermore, a 3.6 Mb cluster on chromosome 6 containing *FLOT1*, *TNF*, and the major histocompatibility complex (*MHC*) was upregulated (**Figure 4**). The
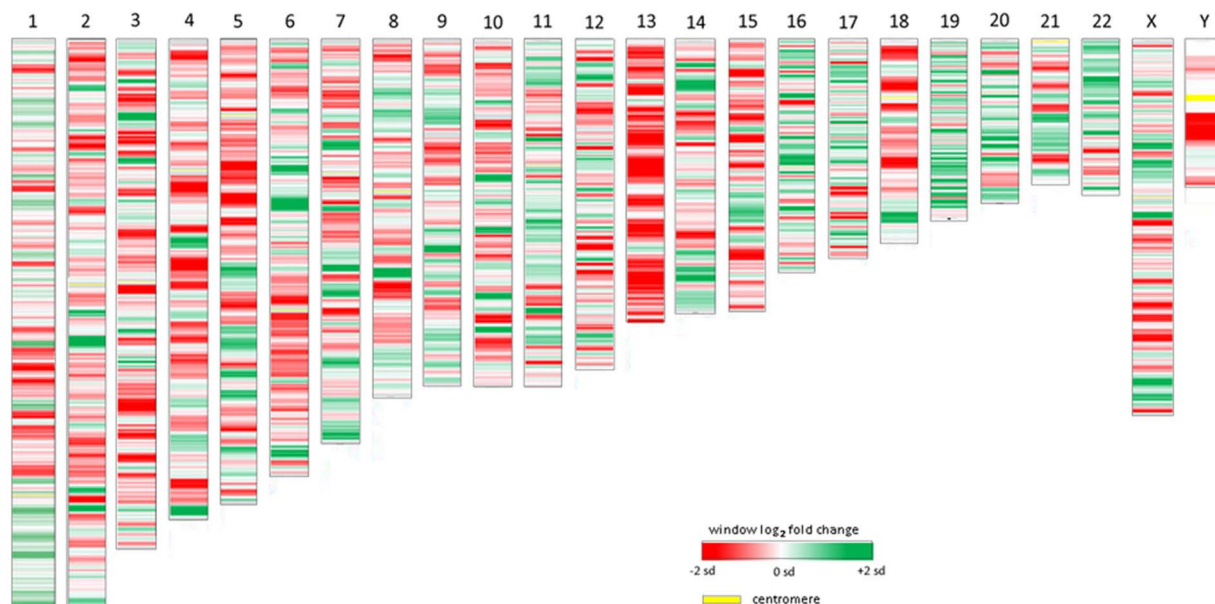
FIGURE 2 | Heat map demonstrations of long linear chromosomal differential transcription of *A. phagocytophilum*-infected human neutrophils vs. uninfected neutrophils. The human chromosome number is shown at the top. The colors represent estimates of the SD of differential transcription of the window at each position compared to the overall differential transcription for that chromosome; dark red = −2 SDs; dark green = +2 SDs.
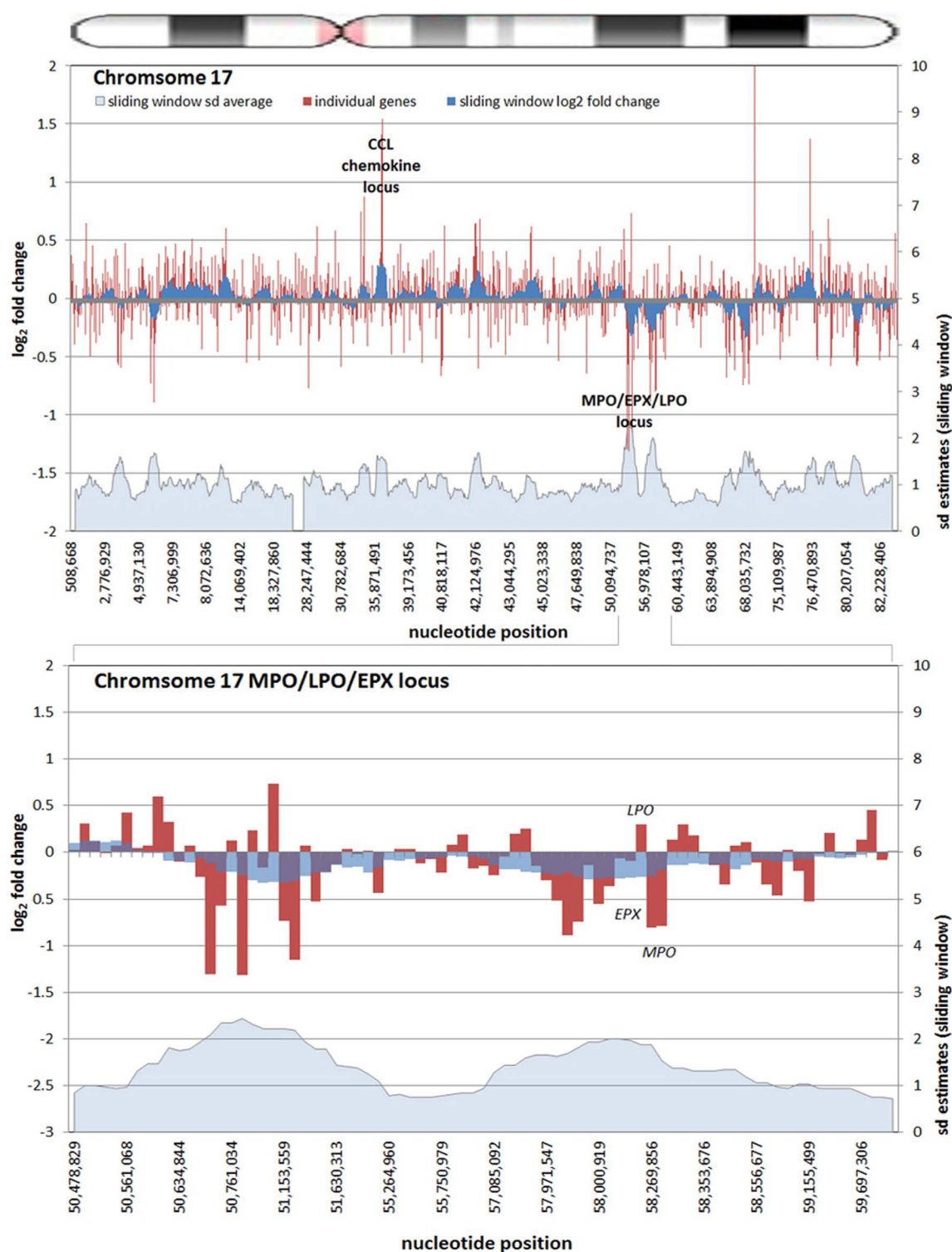
FIGURE 3 | Differential gene expression patterns cluster over large linear genomic regions in *A. phagocytophilum*-infected human peripheral blood neutrophils. The top panel shows human chromosome 17; the bottom panel shows the *MPO/EPX/LPO* locus on human chromosome 17 and the large genomic region that is upregulated with infection. The chromosome ideogram is shown at the top. Red bars (left axes) represent differential transcription of individual genes, including some replicates; the dark blue zones (left axes) show the sliding window average $\log_2$-fold differential transcription over the contiguous 9–11 genes (or 0.38–1.25 Mb). The light blue zone at the bottom (right axis) shows the sliding window average over the same region for estimated SDs of $\log_2$-fold differential transcription at each gene or gene feature. Data re-analyzed from Borjesson et al. (2005) http://www.ncbi.nlm.nih.gov/geo/, accession no. GSE2405.
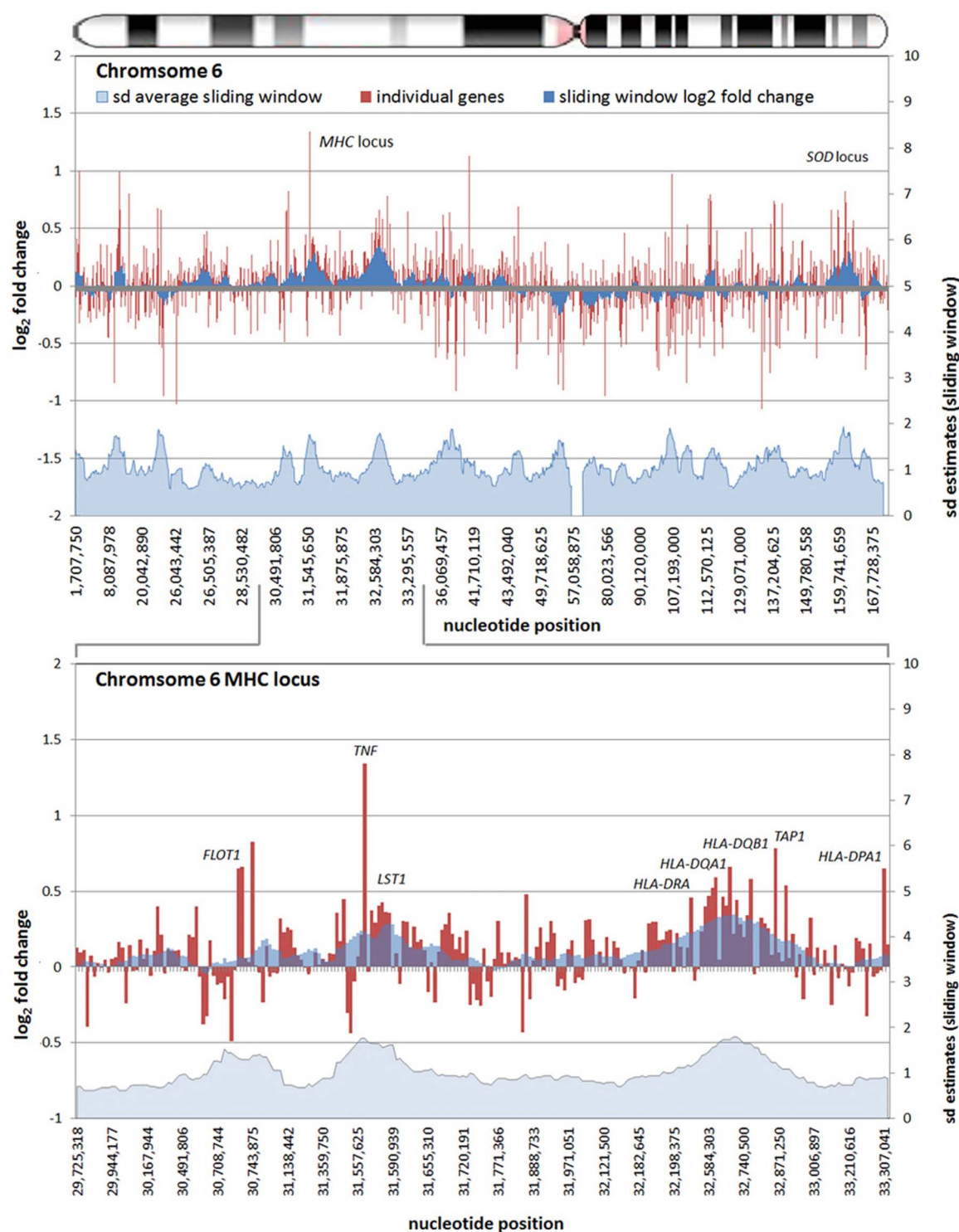
FIGURE 4 | Differential gene expression patterns comparing *A. phagocytophilum*-infected vs. non-infected human peripheral blood neutrophils are clustered over large linear genomic regions. The top panel shows human chromosome 6; the bottom panel shows the MHC locus on human chromosome 6 including a large genomic region spanning the *TNF* and *HLAD* loci that are upregulated with infection. The chromosome ideogram is shown at the top. Red bars (left axes) represent differential transcription of individual genes, including some replicates, the dark blue zones (left axes) show the sliding window average log$_2$-fold differential transcription over the contiguous 9–12 genes (or 0.42–3.24 Mb). The light blue zone at the bottom (right axis) shows the sliding window average over the same region for estimated SDs of log$_2$-fold differential transcription at each gene or gene feature. Data re-analyzed from Borjesson et al. (2005) http://www.ncbi.nlm.nih.gov/geo/, accession no. GSE2405.

chromosomal landscape of the *MHC* locus is well documented with respect to chromatin looping via MARS and histone marks (Ottaviani et al., 2008). It is difficult to determine with certainty whether these changes are the result of manipulation by microbial effectors like AnkA, or whether these represent host cellular responses to infection, or an amalgam of both. The limited genome-wide transcriptional alterations originally documented with exposure to heat-killed *A. phagocytophilum* argue for the former (Borjesson et al., 2005). Regardless, these observations clearly suggest that *A. phagocytophilum* infection targets large chromosomal territories to induce transcriptional alterations, in addition to specific genes. Given that HL-60 cells transfected to express AnkA have nearly identical differential gene expression patterns (Garcia-Garcia et al., 2009b) and reductions in induced respiratory burst compared to *A. phagocytophilum*-infected cells, a hypothesis that prokaryotic nucleomodulins evolved to modulate cell function by epigenetic alterations is compelling.

CONCLUSION AND FUTURE DIRECTIONS

Given the limited nature of bacterial genome repertoire, and the expansive genome repertoire, organization and complexity of transcriptional regulation in eukaryotes, the ability of prokaryotes to alter eukaryotic host cell functions at the epigenetic level is truly a remarkable phenomenon. The processes by which bacteria-derived proteins alter host cell signaling, endocytic or vesicular trafficking, and *cis* transcription of genes are important. However, these pathways become challenging as paradigms that can account for the marked and diverse changes in host cell transcription and functions during the course of infection. We believe that successful intracellular prokaryotes, whether parasitic and pathogenic, mutualistic or symbiotic, evolved multifunctional and complex proteins with broad effects that target key master regulators or checkpoints that regulate major cellular reprogramming events. The *A. phagocytophilum* genome is a fraction of the size of its human host, yet it manages to disrupt core functions of the cell by transforming the transcriptome and reprogramming the cell for improved microbial fitness, survival, and transmission. It is increasingly apparent that *A. phagocytophilum* and perhaps other intracellular prokaryotes manipulate their hosts with a high degree of efficacy, but by altering the "on" or "off" state of genes in a *cis* only fashion or targeting individual or small numbers of signaling pathways is unlikely to account for this alone, in essence, an "effector bottleneck". An analysis of available infection transcriptome data demonstrates that large chromosomal territories appear to be coordinately regulated – up or down. This observation, along with HDAC recruitment by AnkA, lends strong evidence to the idea that *A. phagocytophilum* induces global changes in host gene expression via a broad mechanism that involves epigenetic regulation of transcriptional programs which belie cellular functions.

It is currently unclear as to whether HDAC recruitment is the predominant mechanism by which AnkA exerts its chromatin modulating effects, including whether there are other host factors [e.g., polycomb repressive or hematopoietic associated factor-1 (HAF1) complexes] or additional bacterial-derived nucleomodulins that further contribute to the dramatic changes observed. To examine this, we designed a genome-level bioinformatics approach to predict nucleomodulins based on the combination of secretory and NLSs and studied the genomes of 12 phylogenetically diverse intracellular prokaryotic pathogens, including *Chlamydia pneumoniae*, *Mycobacterium tuberculosis*, and *Yersinia pestis* among others, and identified between 7 and 35 candidates for each (Borroto et al., 2009). Among these and by using iTRAQ proteomics approaches, we identified other proteins encoded in the *A. phagocytophilum* genome that localize to the host cell nucleus and are examining whether these too contribute to chromosomal structure alterations or work in synergy to enhance AnkA function (Gilmore et al., 2012). Elucidating the exact locations where AnkA or other nucleomodulins bind throughout the genome, and determining the architecture of the surrounding chromatin will be crucial to understanding this process. The interplay between HDACs, DNMTs, and methyl binding proteins suggests that *A. phagocytophilum* infection could also induce widespread DNA methylation perhaps as a mechanism for obtaining broad epigenetic changes and functional reprogramming.

Understanding the role of prokaryotic control over complex eukaryotic transcription machinery in lieu of the bottleneck that would occur with single effector targets at signaling pathways will allow for better understanding of the essential components of whole cell transcriptional re-programming. However, this has implications not only for host–pathogen or host–symbiont interactions, but across biology. The ability of proteins like AnkA of *A. phagocytophilum* and OspF of *S. flexneri* to alter inflammatory responses could provide opportunities for engineering of therapeutic agents which interfere with cellular responses in inflammatory diseases or other conditions where epigenetic factors control or contribute to disease.

REFERENCES

Akkoyunlu, M., Malawista, S. E., Anguita, J., and Fikrig, E. (2001). Exploitation of interleukin-8-induced neutrophil chemotaxis by the agent of human granulocytic ehrlichiosis. *Infect. Immun.* 69, 5577–5588. doi: 10.1128/IAI.69.9.5577-5588.2001

Arbibe, L., Kim, D. W., Batsche, E., Pedron, T., Mateescu, B., Muchardt, C., et al. (2007). An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses. *Nat. Immunol.* 8, 47–56. doi: 10.1038/ni1423

Arope, S., Harraghy, N., Pjanic, M., and Mermod, N. (2013). Molecular characterization of a human matrix attachment region epigenetic regulator. *PLoS ONE* 8:e79262. doi: 10.1371/journal.pone.0079262

Banerjee, R., Anguita, J., Roos, D., and Fikrig, E. (2000). Cutting edge: infection by the agent of human granulocytic ehrlichiosis prevents the respiratory burst by down-regulating gp91phox. *J. Immunol.* 164, 3946–3949. doi: 10.4049/jimmunol.164.8.3946

Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K., and Herman, J. G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* 10, 687–692. doi: 10.1093/hmg/10.7.687

Baylin, S. B., and Herman, J. G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* 16, 168–174. doi: 10.1016/S0168-9525(99)01971-X

Beyer, M., Thabet, Y., Muller, R. U., Sadlon, T., Classen, S., Lahl, K., et al. (2011). Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat. Immunol.* 12, 898–907. doi: 10.1038/ni.2084

Bierne, H., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2:a010272. doi: 10.1101/cshperspect.a010272

Bogdanove, A. J. (2014). Principles and applications of TAL effectors for plant physiology and metabolism. *Curr. Opin. Plant Biol.* 19C, 99–104. doi: 10.1016/j.pbi.2014.05.007

Borjesson, D. L., Kobayashi, S. D., Whitney, A. R., Voyich, J. M., Argue, C. M., and Deleo, F. R. (2005). Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *J. Immunol.* 174, 6364–6372. doi: 10.4049/jimmunol.174.10.6364

Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 17, 363–374. doi: 10.1002/prot.340170405

Borroto, C. J., Cole R., Dumler, J. S., and Garcia-Garcia, J. C. (2009). Identification of candidate nuclear effectors among intracellular bacteria. *Paper Presented at the Mid-Atlantic Regional Center for Excellence for Biodefense and Emerging Infectious Diseases [MARCE-2] Investigators' Meeting*, Airlie Center, Warrenton, VA.

Brodsky, I. E., and Medzhitov, R. (2009). Targeting of immune signalling networks by bacterial pathogens. *Nat. Cell Biol.* 11, 521–526. doi: 10.1038/ncb0509-521

Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell. Biol.* 19, 5237–5246.

Cai, S., Lee, C. C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* 38, 1278–1288. doi: 10.1038/ng1913

Carlyon, J. A., Chan, W. T., Galan, J., Roos, D., and Fikrig, E. (2002). Repression of Rac2 mRNA expression by *Anaplasma phagocytophila* is essential to the inhibition of superoxide production and bacterial proliferation. *J. Immunol.* 169, 7009–7018. doi: 10.4049/jimmunol.169.12.7009

Carlyon, J. A., and Fikrig, E. (2003). Invasion and survival strategies of *Anaplasma phagocytophilum*. *Cell Microbiol.* 5, 743–754. doi: 10.1046/j.1462-5822.2003.00323.x

Carlyon, J. A., and Fikrig, E. (2006). Mechanisms of evasion of neutrophil killing by *Anaplasma phagocytophilum*. *Curr. Opin. Hematol.* 13, 28–33. doi: 10.1097/01.moh.0000190109.00532.56

Caturegli, P., Asanovich, K. M., Walls, J. J., Bakken, J. S., Madigan, J. E., Popov, V. L., et al. (2000). ankA: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect. Immun.* 68, 5277–5283. doi: 10.1128/IAI.68.9.5277-5283.2000

Chen, S. M., Dumler, J. S., Bakken, J. S., and Walker, D. H. (1994). Identification of a granulocytotropic *Ehrlichia* species as the etiologic agent of human disease. *J. Clin. Microbiol.* 32, 589–595.

Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S. M., et al. (2004). The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305, 1966–1968. doi: 10.1126/science.1099776

Choi, K. S., Park, J. T., and Dumler, J. S. (2005). *Anaplasma phagocytophilum* delay of neutrophil apoptosis through the p38 mitogen-activated protein kinase signal pathway. *Infect. Immun.* 73, 8209–8218. doi: 10.1128/IAI.73.12.8209-8218.2005

de la Fuente, J., Ayoubi, P., Blouin, E. F., Almazan, C., Naranjo, V., and Kocan, K. M. (2005). Gene expression profiling of human promyelocytic cells in response to infection with *Anaplasma phagocytophilum*. *Cell Microbiol.* 7, 549–559. doi: 10.1111/j.1462-5822.2004.00485.x

Dumler, J. S., Choi, K. S., Garcia-Garcia, J. C., Barat, N. S., Scorpio, D. G., Garyu, J. W., et al. (2005). Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. *Emerg. Infect. Dis.* 11, 1828–1834. doi: 10.3201/eid1112.050898

Dunphy, P. S., Luo, T., and McBride, J. W. (2013). *Ehrlichia* moonlighting effectors and interkingdom interactions with the mononuclear phagocyte. *Microbes Infect.* 15, 1005–1016. doi: 10.1016/j.micinf.2013.09.011

Evans, M. R., Swaminathan, B., Graves, L. M., Altermann, E., Klaenhammer, T. R., Fink, R. C., et al. (2004). Genetic markers unique to *Listeria monocytogenes* serotype 4b differentiate epidemic clone II (hot dog outbreak strains) from other lineages. *Appl. Environ. Microbiol.* 70, 2383–2390. doi: 10.1128/AEM.70.4.2383-2390.2004

Garcia-Garcia, J. C., Barat, N. C., Trembley, S. J., and Dumler, J. S. (2009a). Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathog.* 5:e1000488. doi: 10.1371/journal.ppat.1000488

Garcia-Garcia, J. C., Rennoll-Bankert, K. E., Pelly, S., Milstone, A. M., and Dumler, J. S. (2009b). Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. *Infect. Immun.* 77, 2385–2391. doi: 10.1128/IAI.00023-09

Garyu, J. W., Choi, K. S., Grab, D. J., and Dumler, J. S. (2005). Defective phagocytosis in *Anaplasma phagocytophilum*-infected neutrophils. *Infect. Immun.* 73, 1187–1190. doi: 10.1128/IAI.73.2.1187-1190.2005

Ge, Y., Yoshiie, K., Kuribayashi, F., Lin, M., and Rikihisa, Y. (2005). *Anaplasma phagocytophilum* inhibits human neutrophil apoptosis via upregulation of Bfl-1, maintenance of mitochondrial membrane potential and prevention of caspase 3 activation. *Cell Microbiol.* 7, 29–38. doi: 10.1111/j.1462-5822.2004.00427.x

Gillespie, J. J., Ammerman, N. C., Dreher-Lesnick, S. M., Rahman, M. S., Worley, M. J., Setubal, J. C., et al. (2009). An anomalous type IV secretion system in rickettsia is evolutionarily conserved. *PLoS ONE* 4:e4833. doi: 10.1371/journal.pone.0004833

Gilmore, S. H., Garcia-Garcia, J. C., and Dumler, J. S. (2012). Identification of 6 novel nuclear translocated proteins in *A. phagocytophilum*. *Paper Presented at the 25th Meeting of the American Society of Rickettsiology*, Park City, UT, Abstract 130.

Gong, F., Sun, L., and Sun, Y. (2010). A novel SATB1 binding site in the BCL2 promoter region possesses transcriptional regulatory function. *J. Biomed. Res.* 24, 452–459. doi: 10.1016/S1674-8301(10)60060-7

Hawkins, S. M., Kohwi-Shigematsu, T., and Skalnik, D. G. (2001). The matrix attachment region-binding protein SATB1 interacts with multiple elements within the gp91phox promoter and is down-regulated during myeloid differentiation. *J. Biol. Chem.* 276, 44472–44480. doi: 10.1074/jbc.M104193200

IJdo, J. W., Carlson, A. C., and Kennedy, E. L. (2007). *Anaplasma phagocytophilum* AnkA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol.* 9, 1284–1296. doi: 10.1111/j.1462-5822.2006.00871.x

Jernigan, K. K., and Bordenstein, S. R. (2014). Ankyrin domains across the tree of life. *Peer J.* 2:e264. doi: 10.7717/peerj.264

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., et al. (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.* 30, 4432–4441. doi: 10.1093/nar/gkf566

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., et al. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 19, 187–191. doi: 10.1038/561

Klein, M. B., Hu, S., Chao, C. C., and Goodman, J. L. (2000). The agent of human granulocytic ehrlichiosis induces the production of myelosuppressing chemokines without induction of proinflammatory cytokines. *J. Infect. Dis.* 182, 200–205. doi: 10.1086/315641

Kohwi-Shigematsu, T., Kohwi, Y., Takahashi, K., Richards, H. W., Ayers, S. D., Han, H. J., et al. (2012). SATB1-mediated functional packaging of chromatin into loops. *Methods* 58, 243–254. doi: 10.1016/j.ymeth.2012.06.019

Kumar, P. P., Bischof, O., Purbey, P. K., Notani, D., Urlaub, H., Dejean, A., et al. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC Class I locus. *Nat. Cell Biol.* 9, 45–56. doi: 10.1038/ncb1516

Lee, H. C., Kioi, M., Han, J., Puri, R. K., and Goodman, J. L. (2008). *Anaplasma phagocytophilum*-induced gene expression in both human neutrophils and HL-60 cells. *Genomics* 92, 144–151. doi: 10.1016/j.ygeno.2008.05.005

Letourneau, A., Santoni, F. A., Bonilla, X., Sailani, M. R., Gonzalez, D., Kind, J., et al. (2014). Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* 508, 345–350. doi: 10.1038/nature13200

Lin, M., Kikuchi, T., Brewer, H. M., Norbeck, A. D., and Rikihisa, Y. (2011). Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Front. Microbiol.* 2:24. doi: 10.3389/fmicb.2011.00024

Luo, T., Kuriakose, J. A., Zhu, B., Wakeel, A., and McBride, J. W. (2011). *Ehrlichia chaffeensis* TRP120 interacts with a diverse array of eukaryotic proteins involved in transcription, signaling, and cytoskeleton organization. *Infect. Immun.* 79, 4382–4391. doi: 10.1128/IAI.05608-11

Luo, T., and McBride, J. W. (2012). *Ehrlichia chaffeensis* TRP32 interacts with host cell targets that influence intracellular survival. *Infect. Immun.* 80, 2297–2306. doi: 10.1128/IAI.00154-12

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., et al. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389. doi: 10.1038/30764

Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., et al. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* 23, 58–61. doi: 10.1038/12659

Ohashi, N., Zhi, N., Lin, Q., and Rikihisa, Y. (2002). Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect. Immun.* 70, 2128–2138. doi: 10.1128/IAI.70.4.2128-2138.2002

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/S0092-8674(00)81656-6

Ottaviani, D., Lever, E., Mitter, R., Jones, T., Forshew, T., Christova, R., et al. (2008). Reconfiguration of genomic anchors upon transcriptional activation of the human major histocompatibility complex. *Genome Res.* 18, 1778–1786. doi: 10.1101/gr.082313.108

Park, J., Kim, K. J., Choi, K. S., Grab, D. J., and Dumler, J. S. (2004). *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol.* 6, 743–751. doi: 10.1111/j.1462-5822.2004.00400.x

Paschos, K., and Allday, M. J. (2010). Epigenetic reprogramming of host genes in viral and microbial pathogenesis. *Trends Microbiol.* 18, 439–447. doi: 10.1016/j.tim.2010.07.003

Pennini, M. E., Perrinet, S., Dautry-Varsat, A., and Subtil, A. (2010). Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis*. *PLoS Pathog.* 6:e1000995. doi: 10.1371/journal.ppat.1000995

Reddington, J. P., Pennings, S., and Meehan, R. R. (2013). Non-canonical functions of the DNA methylome in gene regulation. *Biochem. J.* 451, 13–23. doi: 10.1042/BJ20121585

Rennoll-Bankert, K. E., and Dumler, J. S. (2012). Lessons from *Anaplasma phagocytophilum*: chromatin remodeling by bacterial effectors. *Infect. Disord. Drug Targets* 12, 380–387. doi: 10.2174/187152612804142242

Rohde, J. R. (2011). Microbiology. Listeria unwinds host DNA. *Science (New York, N.Y.)* 331, 1271–1272. doi: 10.1126/science.1203271

Rolando, M., Sanulli, S., Rusniok, C., Gomez-Valero, L., Bertholet, C., Sahr, T., et al. (2013). *Legionella pneumophila* effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. *Cell Host Microbe* 13, 395–405. doi: 10.1016/j.chom.2013.03.004

Scorpio, D. G., Akkoyunlu, M., Fikrig, E., and Dumler, J. S. (2004). CXCR2 blockade influences *Anaplasma phagocytophilum* propagation but not histopathology in the mouse model of human granulocytic anaplasmosis. *Clin. Diagn. Lab. Immunol.* 11, 963–968. doi: 10.1128/CDLI.11.5.963-968.2004

Silmon de Monerri, N. C., and Kim, K. (2014). Pathogens hijack the epigenome: a new twist on host–pathogen interactions. *Am. J. Pathol.* 184, 897–911. doi: 10.1016/j.ajpath.2013.12.022

Spector, D. L. (2003). The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* 72, 573–608. doi: 10.1146/annurev.biochem.72.121801.161724

Srivastava, S., Mishra, R. K., and Dhawan, J. (2010). Regulation of cellular chromatin state: insights from quiescence and differentiation. *Organogenesis* 6, 37–47. doi: 10.4161/org.6.1.11337

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L. W., et al. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science (New York, N.Y.)* 282, 754–759. doi: 10.1126/science.282.5389.754

Stirzaker, C., Song, J. Z., Davidson, B., and Clark, S. J. (2004). Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.* 64, 3871–3877. doi: 10.1158/0008-5472.CAN-03-3690

Sugio, A., Yang, B., Zhu, T., and White, F. F. (2007). Two type III effector genes of *Xanthomonas oryzae* Pv. *oryzae* control the induction of the host genes OsTFIIAγ1 and OsTFX1 during bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10720–10725. doi: 10.1073/pnas.0701742104

Thomas, V., Samanta, S., Wu, C., Berliner, N., and Fikrig, E. (2005). *Anaplasma phagocytophilum* modulates gp91phox gene expression through altered Interferon Regulatory Factor 1 and PU.1 levels and binding of CCAAT displacement protein. *Infect. Immun.* 73, 208–218. doi: 10.1128/IAI.73.1.208-218.2005

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419, 641–645. doi: 10.1038/nature01084

Zhu, B., Kuriakose, J. A., Luo, T., Ballesteros, E., Gupta, S., Fofanov, Y. et al. (2011). *Ehrlichia chaffeensis* TRP120 binds a G+C-rich motif in host cell DNA and exhibits eukaryotic transcriptional activator function. *Infect. Immun.* 79, 4370–4381. doi: 10.1128/IAI.05422-11

Bacterial SET domain proteins and their role in eukaryotic chromatin modification

Raúl Alvarez-Venegas *

Laboratory of Chromatin and Epigenetics, Department of Genetic Engineering, CINVESTAV Unidad-Irapuato, Irapuato, México

It has been shown by many researchers that SET-domain containing proteins modify chromatin structure and, as expected, genes coding for SET-domain containing proteins have been found in all eukaryotic genomes sequenced to date. However, during the last years, a great number of bacterial genomes have been sequenced and an important number of putative genes involved in histone post-translational modifications (histone PTMs) have been identified in many bacterial genomes. Here, I aim at presenting an overview of SET domain genes that have been identified in numbers of bacterial genomes based on similarity to SET domains of eukaryotic histone methyltransferases. I will argue in favor of the hypothesis that SET domain genes found in extant bacteria are of bacterial origin. Then, I will focus on the available information on pathogen and symbiont SET-domain containing proteins and their targets in eukaryotic organisms, and how such histone methyltransferases allow a pathogen to inhibit transcriptional activation of host defense genes.

Keywords: SET-domain, bacteria, pathogen, symbiont, histone methyltransferases, epigenetics

INTRODUCTION

In eukaryotic organisms, genomic DNA is packaged in the form of a complex structure known as chromatin. The basic unit of chromatin is the nucleosome, consisting of approximately two superhelical turns of DNA wrapped on a histone octamer composed of four histone species: a histone H3/H4 tetramer and two histone H2A/H2B dimmers (Kornberg, 1977). This nucleoprotein complex occurs basically every 200 ± 40 bp in all eukaryotic genomes. The repeating nucleosome units further assemble into higher-order structures stabilized by the linker histone H1.

Core histones are predominantly globular except for their N-terminal "tails," which are unstructured. A remarkable attribute of histones, and mainly of their tails, is the great number and type of modified residues they possess (Kouzarides, 2007). Thus, there are several histone post-translational modifications (histone PTMs) that correlate with either positive or negative transcriptional states. These modifications include acetylation, phosphorylation, methylation, ubiquitylation, and SUMOylation. Most modifications localize to the amino- and carboxy-terminal histone tails, and a few localize to the histone globular domains (Berger, 2007). It has been suggested that the combination of histone amino-terminal modifications on one or more histones represent a "histone code" that modulates gene expression, regulates chromatin structure, and determines cellular and epigenetic identities during development, therefore extending the information potential of the genetic code encoded in the DNA (Jenuwein and Allis, 2001; Casadio et al., 2013). Recently, with the discovery of several novel histone-binding modules the histone-code hypothesis predicts the existence of "reader proteins" that recognize chromatin covalent-modification marks to influence downstream events (Ruthenburg et al., 2007). On the other hand, all histone PTMs are removable. For example, histone deacetylases (HDACs) remove acetyl groups; Ser/Thr phosphatases remove phosphate groups; ubiquitin proteases remove mono-ubiquitin from H2B; arginine methylation is altered by deiminases; and two classes of lysine demethylases remove methyl groups from lysines: the LSD1/BHC110 class and the jumonji class (Berger, 2007).

One of the most extensively studied histone PTMs has been the methylation of lysine residues in histones. Accordingly, the first histone lysine methyltransferase (HKMT) to be identified was the human and mouse SUV39H1 that targets lysine 9 of histone H3 (H3K9) (Rea et al., 2000). Since then, numerous HKMTs have been identified, most of which methylate lysines within the histone N-terminal tails. Noticeably, all of the HKMTs that methylate N-terminal lysines contain the SET domain, a ∼130 amino acid catalytic domain initially found to be conserved in the PEV modifier SU(VAR)3-9, the Polycomb-group protein E(Z), and the trithorax-group protein TRX (Jenuwein et al., 1998). Crystal structures of SET-domain proteins have revealed that the SET domain is folded into several small β sheets (packed together with pre-SET and post-SET domains or regions) surrounding a knot-like structure. This "pseudo-knot" fold brings together the two most-conserved sequence motifs of the SET domain (RFINHXCXPN and ELXFDY) to form an active site in a location immediately next to the pocket where the methyl donor binds and to the peptide-binding cleft (Dillon et al., 2005). Moreover, it has also been shown that both N- and C-terminal flanking regions to the SET-domain are as well required for HKMT activity (Rea et al., 2000).

Histone lysine residues methylated *in vivo*, in animals, include H3K4, H3K9, H3K27, H3K36, H3K79, H4K20, H2BK5, and H1K26 (Barski et al., 2007). The first H3K4 (histone H3 lysine 4) methylase, Set1/COMPASS, was isolated from *Saccharomyces*

cerevisiae and was demonstrated to be capable of mono-, di-, and tri-methylate H3K4 (Miller et al., 2001; Roguev et al., 2001). Thus, methylation can occur several times on one lysine side chain and each level of modification may have different biological outcomes. Furthermore, of the large family of SET domains, a subset is encoded by bacterial pathogens and symbionts, which lack chromatin, suggesting a role in altering the host chromatin upon infection. For instance, pathogenic bacteria make use of a wide range of strategies to avoid elimination by their host. Then, aiming at histone modifications could allow a pathogen to inhibit transcriptional activation of host defense genes. Hence, pathogenic bacteria can be considered as "epimutagens" able to remodel the epigenome. Their effects might generate specific, long-lasting imprints on host cells, leading to a memory of infection that influences immunity and that might be the foundation of unexplained diseases.

Here, I will examine the available information on SET domain genes identified in bacterial genomes. I will analyze the hypothesis that SET domain genes found in bacteria are of bacterial origin. Then, I will concentrate on the information related to pathogen and symbiont SET-domain containing proteins, their targets in eukaryotic organisms, and on how such histone methyltransferases could allow a pathogen to inhibit transcriptional activation of host defense genes.

BACTERIAL SET DOMAIN

Considering that SET-domain containing proteins modify chromatin structure, as expected, genes coding for SET-domain containing proteins have been found in all eukaryotic genomes sequenced to date. However, during the last years, thanks to the recent advances in biological research such as next-generation sequencing, a great number of bacterial genomes have been sequenced and, what's more, an important number of putative genes involved in histone PTMs have been identified in many bacterial genomes. For example, from 390 completely and partially sequenced bacterial genomes available in 2007 at the National Center for Biotechnology Information database (NCBI), 83 bacterial species encoding putative SET domain proteins were retrieved (Alvarez-Venegas et al., 2007). However, a basic BLASTP search performed today retrieves more than 500 bacterial genomes (and counting; Supplementary Table ST1), including very closely related genomes, all of them containing SET-domain proteins. The number of hits range from one to four SET coding genes per genome. Interestingly, species like *Gemmata obscuriglobus* (a nonpathogenic spherical budding bacteria; NCBI locus WP_010044726), *Bacillus coahuilensis* (a Gram-positive, spore-forming bacterium from a highly saline desert lagoon; NCBI locus WP_010172611), *Opitutus terrae* PB90-1 (an obligatory anaerobic bacteria isolated from rice paddy soil; NCBI locus YP_001821399), *Burkholderia rhizoxinica* HKI 454 (an intracellular symbiont of a phytopathogenic fungus; NCBI locus CBW73645), *Methanoregula boonei* 6A8 (a novel acidiphilic, hydrogenotrophic methanogen; NCBI locus WP_012107610), and many more interesting bacterial species have recently been shown to code for SET-domain containing proteins (Supplementary Table ST1).

At first, proteins with SET domain present in bacterial genomes were considered to be the result of horizontal transfer from a eukaryotic host (Stephens et al., 1998; Aravind and Iyer, 2003). However, nowadays, the expanded collection of sequenced bacterial genomes, to include not only pathogenic but also free-living and environmental species, as well as methanogenic archaea, indicate that SET domain genes have existed in the bacterial domain of life (Alvarez-Venegas et al., 2007). Furthermore, Aravind et al. (2011) have proposed that SET domain methylases, which display the β-clip fold, first emerged in prokaryotes from the SAF superfamily of carbohydrate-binding domains (based on its representative members, SAS, type III AFP and FlgA). Thus, taking into account recent evidence that supports a chromatin-related role for at least a portion of the bacterial SET domain versions, it is likely that the SET domain had matured into a primitive chromatin-remodeling enzyme in prokaryotes, prior to its transfer to eukaryotes (Aravind et al., 2011).

On the other hand, phylogenetic analysis have shown that bacterial SET domain genes have undergone an evolution of their own, unrelated to the evolution of the eukaryotic SET domain genes (Alvarez-Venegas et al., 2007; Murata et al., 2007). This can be substantiated, for example, by performing a BLASTP search at NCBI, with any SET-domain. In all "distance trees" produced using BLAST pair-wise alignments, the Newick dendrograms (as well as any phylogenetic tree) produced at NCBI show all eukaryotic entries clustered as a monophyletic group. Important to our discussion is that eukaryotic SET proteins do not mix together with SET proteins of bacterial origin and that a similar distribution pattern is continuously reproduced with different combinations of prokaryotic or eukaryotic entries (data not shown). In addition, the branching arrangement of the respective monophylies of Archaea, Eukarya, and Bacteria rejects any sort of horizontal gene transfer (HGT) involving SET genes from bacteria and vertebrates, although some ambiguous cases may arise in such analyses and could be, for example, a consequence of the poor representation of certain genomes in present-day databases (Stanhope et al., 2001). In contrast, phylogenetic and chromosome analyses of Chlorobium, Bacillus, and Methanosarcinal SET domain-containing species support an ancient HGT between bacteria and Archeae (Alvarez-Venegas et al., 2007).

PATHOGEN AND SYMBIONT SET-DOMAIN CONTAINING PROTEINS

CHLAMYDIAE SET DOMAIN PROTEINS

One of the earliest reports on the identification of SET domain proteins in bacteria was the result of the genome sequencing of an obligate intracellular pathogen of humans that targets epithelial cells, *Chlamydia trachomatis* (Stephens et al., 1998). At that time, it was suggested that the SET domain protein CT737 in Chlamydia was the result of HGT based upon the assumption that the SET domain was found only in eukaryotic chromatin-associated proteins (Stephens et al., 1998).

Unlike eukaryotes, prokaryotes do not have histones nor highly ordered chromatin. However, pathogenic bacteria must employ a wide range of tactics to avoid eradication by their host. Then, targeting histone modifications could be one of those strategies that allow a pathogen to inhibit transcriptional

activation of host defense genes. With this in mind, Pennini et al. (2010) set out to characterize the chlamydial SET domain protein CT737 (named NUE, as the first "nuclear effector identified in chlamydiae"). They found a type three secretion (TTS) system signal in the N-terminal part of NUE. Also, Pennini and colleagues found that NUE acts as a TTS system effector protein, that it is secreted from bacteria, translocated to the host cell nucleus during *C. trachomatis* infection and associated with chromatin, particularly at late time points of infection. NUE has a histone methyltransferase activity that modified mammalian histones H2B, H3, and H4 but with a stronger activity toward H4 (**Table 1**). NUE itself automethylates, suggesting that automethylation enhances NUE enzymatic activity toward its substrate (Pennini et al., 2010). Consequently, the chlamydial SET domain protein seems to have evolved into a secreted protein capable to modify eukaryotic histones (**Figure 1**). Thus, by using its SET domain protein as an epigenetic control of host cells represents an advantage for Chlamydia in the persistence of chlamydial infection to maintain chronic disease progression.

On the other hand, *Chlamydophila pneumonia*, an obligatory intracellular eubacterium that causes acute respiratory diseases (e.g., pneumonia) and chronic inflammatory processes (e.g., asthma, bronchitis, and chronic obstructive pulmonary disease; Roulis et al., 2013), has also a SET-domain coding gene, recently characterized (NCBI locus BAA99086). The *C. pneumoniae* SET domain protein (named cpnSET by Murata et al., 2007), with a murine histone H3 methyltransferase activity, has a similar expression pattern of two chlamydial histone H1-like proteins, Hc1, and Hc2. In addition, cpnSET has a physical interaction with the Hc1 and Hc2 proteins, as determined by the yeast two-hybrid system. Furthermore, Hc1 is also methylated by cpnSET, indicative that cpnSET may play an important role in chlamydial cell maturation due to modification of chlamydial histone H1-like proteins (Hc1 proteins) during the alternate morphologies between elementary bodies (EBs) and reticulate bodies (RBs) of *C. pneumonia* (Murata et al., 2007). However, localization of cpnSET was shown mainly in chlamydial cells (Murata et al., 2007), raising the possibility that the *C. pneumoniae* cpnSET protein is exported into host cells, and then cpnSET and host histones may functionally interact with each other. This is something extremely possible if we consider that the cpnSET protein has the two ("bipartite") potential nuclear localization sequences (RHR at position 123 and KHRKKR at position 208) as those analyzed in the *C. trachomatis* translocated SET domain protein NUE (RRR at position 121 and KHRKKR at position 206; Pennini et al., 2010). Moreover, the chimeric N-terminal sequence of *C. pneumoniae* was secreted when expressed in *Shigella flexneri* ipaB (constitutive type three secretion—TTS-system), something highly indicative that chlamydiae SET domains proteins are TTS effectors (Pennini et al., 2010). Conversely, NUE did not methylate Hc1, as cpnSET did (**Figure 1**) (Murata et al., 2007; Pennini et al., 2010). This is something that deserves further investigation.

PROTEOBACTERIA: *LEGIONELLA PNEUMOPHILA*

The causative agent of a severe form of pneumonia called Legionnaires' disease, *Legionella pneumophila*, is a Gram-negative intracellular pathogen, ubiquitous in aquatic habitats where it survives and replicates within a wide range of protists, such as amoeba and ciliated protozoa. Upon transmission to humans, *L. pneumophila* invades and replicates in alveolar macrophages, evades the default endosome–lysosome pathway, remodels the phagosome, and escapes into the host cell cytosol. This ends in the expression of various bacterial virulence traits and bacterial escape to the extracellular environment, leading to the disease (Al-Quadan et al., 2012). *L. pneumophila* is a master manipulator of a variety of eukaryotic hosts ranging from unicellular amoebae to mammals. *L. pneumophila* facilitates this by taking control of numerous eukaryotic cellular functions through translocation of around 300 effectors into the host cell by the Dot/Icm type IVB secretion system. More than 70 of the injected effector proteins contain eukaryotic-like domains, including the ankyrin repeat, F-box, U-box, leucine-rich repeats and SET domain (Price and Abu Kwaik, 2013).

Recently, Rolando et al. (2013) identified one protein encoded by the gene *lpp1683* of the *L. pneumophila* strain Paris that contains a SET domain. It was shown that the Lpp1683 protein (named RomA, for regulator of methylation A) is a histone methyl transferase, with an apparent preference for histone H3. Specifically, RomA tri-methylates lysine 14 of histone H3 (H3K14), on free histones as well as in reconstituted oligonucleosomes. Then, Rolando and colleagues determined that RomA also conserved its activity *in vivo* during infection of human macrophages and amoeba, and that RomA localizes to the host nucleus in which it catalyzes a new histone methylation mark, H3K14. Deletion of a nuclear localization signal (NLS) in its N-terminal part altered the cellular distribution of RomA, leading to a predominant cytosolic localization. Fluorescent-based translocation assays showed that this protein is translocated in a Dot/Icm-dependent manner into the host cell. Next, it was shown that the H3K14 methylation mark, which replaces H3K14 acetylation (an active mark of ongoing transcription at the transcription start sites, TSSs), functions as a transcriptional repressive mark and results in global gene transcriptional repression, particularly in genes that are involved in innate immunity. Furthermore, deletion of RomA leaves the Paris strain defective in intracellular growth within both macrophages and amoebae, indicative that repression of host global transcription is important for *L. pneumophila* pathogenesis.

In contrast, Li et al. (2013) used the *L. pneumophila* Philadelphia-derivative Lp02 strain to characterize the protein LegAS4, a *L. pneumophila* type IV secretion system (TFSS) effector that contains a SET domain and tandem nuclear localization signals (NLS). LegAS4 is efficiently translocated from *L. pneumophila* into host macrophages and when ectopically expressed, LegAS4 was localized exclusively in the nuclei, specifically was localized in the host nucleolus, and associated with rDNA chromatin in which it catalyzes H3K4 dimethylation at the rDNA promoter and promoted rDNA transcription. Furthermore, LegAS4 interacted with host heterochromatin binding proteins HP1α and HP1γ, but extremely weakly with HP1β. The high-affinity HP1 binding is responsible for specific recruitment of LegAS4 to the transcriptionally silent rDNA chromatin region. Thus, *L. pneumophila* might exploit host ribosome activity for its own survival

Table 1 | Bacterial SET-domain proteins involved in eukaryotic chromatin modification.

Bacterial SET-domain protein	Bacterial species	GenBank accession number	Targeted histone(s)	Histone methyl mark	Targeted host/substrate	References
CT737 or NUE	*Chlamydia trachomatis*	NP_220256	H2B, H3, H4	n.d.	Mammalian histones	Pennini et al., 2010
cpnSET	*Chlamydophila pneumoniae*	BAA99086	H3	n.d.	Murine histones	Murata et al., 2007
RomA	*Legionella pneumophila* strain Paris	YP_124001	H3	H3K14me3	Free histones; oligonucleo-somes; macrophages	Rolando et al., 2013
LegAS4	*Legionella pneumophila* strain Philadelphia-Lp02	AAU27798	H3	H3K4me2; H3K9me3	Human macrophages	Li et al., 2013
LegAS4-like or *Bt*SET	*Burkholderia thailandensis*	YP_443833	H3	H3K4me and H3K4me2	Free histones	Li et al., 2013
*Ba*SET	*Bacillus anthracis*	YP_002869308	H1	Lysine residues	Macrophages	Mujtaba et al., 2013
Gö1-SET	*Methanosarcina mazei* strain Gö1	AAM32541	H4	H4K5	Bovine histones; MC1-α	Manzur and Zhou, 2005

Abbreviations: n.d., not determined; H3K14me3, tri-methylated lysine 14 of histone H3; H3K4me2, di-methylated lysine 4 of histone H3.

FIGURE 1 | Control of host gene expression by bacterial SET domain proteins and their mode of action. Schematic representation of Burkholderia, Legionella, Chlamydia, and Bacillus secreted factors involved in the control of gene expression of host cells, as detailed in the text. Me: histone methylation; H3K14me3: tri-methylated lysine 14 of histone H3; H3K4me2: di-methylated lysine 4 of histone H3; H3K9me3: tri-methylated lysine 9 of histone H3; H1K: histone H1 lysine.

advantages by stimulation of rDNA transcription (Li et al., 2013).

Why those two highly homologous effector proteins have such inconsistent phenotypes in two different strains of *L. pneumophila*? This could be the result of the differences in the amino acid sequences of both proteins, as suggested by Price and Abu Kwaik (2013). For example, RomA is missing an amino-terminal fragment present in LegAS4. In addition, there is a weak sequence homology at an eight amino acid stretch almost in the middle of the proteins. These sequence differences could alter the

structure of RomA from LegAS4 resulting in a change in histone methyltransferase substrate specificity. Also, LegAS4 has a robust localization to the nucleolus, which is not seen for RomA. The ability of LegAS4 to localize to the nucleolus is dependent on its tandem NLS, whereas RomA has three individual amino acid differences from LegAS4 in its corresponding NLS (Price and Abu Kwaik, 2013). Therefore, the structural differences between LegAS4 and RomA most likely explain the differential nucleolar protein targeting and distinct functional phenotypes of homologous effectors. It will be interesting to determine whether LegAS4

can also catalyze H4K14me3 in the nucleolus, as well as to determine the potential role of HP1 in RomA-mediated genome-wide repression.

Other bacterial pathogens, including *Bordetella bronchiseptica* and *Burkholderia thailandensis*, also possess LegAS4-like HKMTase effectors targeted to the host nucleolus. Specifically, the *B. thailandensis* type III effector BtSET upholds H3K4 methylation of rDNA chromatin and contributed to infection-induced rDNA transcription (Li et al., 2013). Thus, activation of rDNA transcription could be a common virulence strategy employed by bacterial pathogens for intracellular survival.

BACILLUS SET DOMAIN PROTEINS

Bacillus anthracis, the etiologic agent of anthrax disease, is a Gram-positive spore-forming organism found in soil environments. *B. anthracis* spores are taken up by macrophages and/or dendritic cells, and subsequently migrate in the draining lymph nodes where they germinate, leading to bacterial multiplication, and dissemination through the whole organism (Raymond et al., 2009). Respiratory, gastrointestinal, or cutaneous entry of *B. anthracis* spores into mammals can result in a rapid systemic infection and death (mainly, in the pulmonary form and in the absence of treatment). Accordingly, the ability to drive bacterial molecules directly into host cells is a major strategy used by diverse bacterial pathogens to destabilize the host transcriptional machinery and to overcome host defenses. Strategically, most effectors aim to stabilize the NF-κB/IκB (nuclear factor-κB/inhibitor of κB) complex for precluding nuclear localization and transcriptional activation of the nuclear factor NF-κB (Mujtaba et al., 2013).

Recently, Mujtaba et al. (2013) characterized a SET-domain protein in *B. anthracis* (named *Ba*SET) in order to determine its role in *B. anthracis* survival in infected hosts. It was shown that *Ba*SET is a specific histone H1 trimethylase that functions as a transcriptional repressor by reducing the activation of NF-κB response elements (NF-κB _RE) in a dose-dependent expression, as well as by repressing diverse NF-κB target gene promoters. Particularly, *Ba*SET methylates eight lysine residues of the histone H1, which could be the strategy of the bacillus to hypermethylate host chromatin to silence the host inflammatory response. Furthermore, *Ba*SET is secreted by *B. anthracis* and localizes to the nucleus of infected macrophages where it methylates Histone H1, although the mechanism by which *Ba*SET translocates outside the bacillus is unclear, as the *Ba*SET sequence does not display any secretion signal (Mujtaba et al., 2013). Additionally, an engineered *Ba*SET deletion mutant (*Ba*ΔSET) indicated that *Ba*SET is not involved in the formation and germination of *B. anthracis* heat-resistant endospores, but that it plays a major role in the virulence of *B. anthracis*, as deletion of the gene eliminates the capacity for the organism to cause disease and death as well as survival in the infected host.

It will be interesting to determine the function of SET domain proteins present in other Bacillus, for example: *Bacillus cereus* (a ubiquitous soil organism and an opportunistic human pathogen most commonly associated with food poisoning), and *Bacillus thuringiensis* (an insect pathogen that is widely used as a bio-pesticide) (Han et al., 2006). But even more interesting will be the study of SET domain proteins in bacteria like *Bacillus megaterium* (a commercially available, nonpathogenic host for the biotechnological industry), and *Bacillus coahuilensis*, an ancient and a moderately halophilic, Gram-positive and rod-shaped bacterium, isolated from a Chihuahuan desert lagoon in Cuatro Ciénegas, Coahuila, México (Alcaraz et al., 2008).

ARCHAEAL SET DOMAIN

On the basis of their molecular properties, Archaea has been defined as a separate domain of life. Archaea lack nuclear membranes and are therefore prokaryotes, however, they are genetically and biochemically as divergent from bacteria as are eukarya. Archaea contain a set of sequence-independent DNA-binding proteins some of which undergo post-translational modifications, similar to the histone modifications in eukaryotic chromatin (Reeve, 2003). Among these archaeal DNA-binding proteins are the so-called histone-like proteins. On the other hand, SET domain encoding genes have also been identified in Archaea (Aravind and Iyer, 2003). Consequently, in order to determine the functional significance of SET domain proteins in Archaea, Manzur and Zhou (2005) characterized the first archaeal SET protein from the acetate-utilizing archaeal methanogen, *Methanosarcina mazei* strain Gö1 (referred as Gö1-SET).

The Gö1-SET protein was shown to selectively methylate *in vitro* bovine histone H4 at Lys5. However, histone H4 is not present in *M. mazei*. Alternatively, *M. mazei* has three DNA interacting proteins: a histone-like protein and two homologous MC1 proteins (methanogen chromosomal 1, MC1-α and MC1-β proteins). In view of that, *in vitro* MTase assays using the three DNA interacting proteins showed that Gö1-SET selectively methylates MC1-α but not the MC1-β and the histone-like protein, and that likely Lys37 of MC1-α is the specific target of Gö1-SET (Manzur and Zhou, 2005). Thus, archaeal SET domain proteins, like Gö1-SET, may regulate structures of archaeal chromatin composed of MC1–DNA complexes and indicate that chromatin modification by methylation took place before the separation of the archaeal and eukaryotic lineages (Manzur and Zhou, 2005).

Recently, a crenarchaeal protein lysine methyltransferase (named aKMT4, or also aKMT), which shows structural and enzymatic similarity to the eukaryotic KMT4/Dot1 family, has been characterized from *Sulfolobus islandicus* (Chu et al., 2012; Niu et al., 2013). However, such protein does not have a SET-domain and belongs to the Dot 1 family of histone lysine methyltransferases (KMTs). Nonetheless, the detection and further characterization of aKMT4/aKMT and protein homologs in other crenarchaeal species will allow a better understanding of the mechanisms involved in lysine methylation in crenarchaea and will clarify the evolutionary relationships among methyltransferases from the three domains of life.

SYMBIONTS

Most of the research related to complex interactions between eukaryotes and bacteria has been related to associations with microbes that are pathogenic; usually, microbial infection has been considered as deleterious, or at best irrelevant, to vigor and reproduction. However, symbionts and their metabolic potential play essential roles for many eukaryotic organisms that may

benefit from enhanced fitness, survival, and even acquired virulence. Symbiosis is ubiquitous in terrestrial, freshwater, and marine ecosystems and it has played a crucial role in the appearance of major life forms on Earth and in the generation of biological diversity (Moran, 2006). Frequently, the associations are persistent for the hosts and are being transmitted vertically across generations. At times, the organisms involved in a symbiosis may be fully fused that they cannot live separately or be recognized as distinct entities without close scrutiny.

Recently, Partida-Martinez et al. (2007a,b) reported a unique symbiosis between bacteria belonging to the genus Burkholderia and the saprotrophic fungus *Rhizopus microsporus*. They have found that *Burkholderia rhizoxinica* is an intracellular symbiont of the phytopathogenic fungus *Rhizopus microspores* and that *B. rhizoxinica* is the producer of rhizoxin, the causative agent of rice seedling blight. This symbiosis represents an extraordinary example in which a fungus hosts a bacterial population for the production of a virulence factor. On the other hand, studies on the evolution of host resistance indicate that the fungus lost its ability to sporulate independently and became totally dependent on endobacteria for reproduction through spores (fungus formation of sporangia and spores is restored only upon reintroduction of endobacteria), thus warranting the persistence of the symbiosis and its efficient distribution through vegetative spores (Partida-Martinez et al., 2007a). Therefore, reproduction of the host is totally dependent on endofungal bacteria, which in return provides a highly potent toxin for defending the habitat and accessing nutrients from decaying plants.

Interestingly, *B. rhizoxinica* has a gene coding for a SET-domain protein (locus YP_004027789). However, the function of this SET protein is still unknown. On the other hand, *B. rhizoxinica* has a type II secretion pathway, an encoded type III secretion system shown to play a crucial role for the establishment of the symbiosis, and a putative type IV secretion system (Lackner et al., 2011). It is tempting to speculate that the SET-domain protein functions as a type III or type IV secretion system effector (just like SET proteins present in other proteobacteria), and that it might be translocated from *B. rhizoxinica* into the fungus *R. microspores*. The question is, for what purpose? If we take into consideration that a microbe that forms chronic infections in a host or in a host lineage may evolve to conserve or even to benefit its host, as this will help to maintain its immediate ecological resource (Moran, 2006), then we can only hypothesize that the bacterial SET protein might complement the array of post-translational modifications of histones in the fungus in order to improve bacterial intracellular replication through regulation of its host gene expression, and most important for the fungus, to formation of sporangia and spores. This is something that has to be scrutinized in the near future.

CONCLUDING REMARKS

Pathogenic or symbiotic bacteria make use of eukaryotic cell functions via specific interactions between microbial surface factors or secreted molecules and eukaryotic targets. These interactions, in turn, affect multiple signaling pathways and consequently promote a wide range of effects in host cells, including altered production of components involved in immune responses.

Thus, as expected, bacteria employ a variety of strategies to affect the host cell cycle and gene expression program for their own benefit.

Recent studies have found that microbes affect a diverse set of epigenetic factors such as DNA methylation, histone modifications, chromatin-associated complexes, and non-coding RNAs (ncRNAs) to alter chromatin structure and gene expression. Intriguingly, DNA methylation plays a critical role in epigenetic gene regulation in eukaryotes as well as in prokaryotes (Kumar and Rao, 2013); Argonaute proteins, key players in RNA interference (RNAi) and related gene silencing phenomena in diverse eukaryotic species, are also present in many bacterial and archaeal species (Makarova et al., 2009); and post-translational modifications (PTMs) of proteins (e.g., histone and histone-like protein modifications) are used by both prokaryotic and eukaryotic cells to regulate the activity of key proteins. Thus, epigenetic mechanisms are clearly implicated in modulating biological interactions between hosts and bacterial pathogens and symbionts.

Chromatin modifications during animal development and in response to diverse environmental factors contribute to adult phenotypic variability and susceptibility to a number of diseases, including cancer, neurodegenerative, and neurological diseases and autoimmune disorders (Portela and Esteller, 2010). Because epigenetic modifications of chromatin may be transmitted to daughter cells during cell division, leading to heritable changes in gene expression, it is likely that a bacterial infection could generate heritable marks (Bierne et al., 2012). Therefore, understanding whether histone modification and/or DNA methylation marks imposed by bacterial proteins are maintained over time is a whole new area of research.

In plants, heritable histone modifications (e.g., histone lysine methylation) reported to regulate plant immunity against bacterial pathogens, by plant endogenous SET-domain proteins, have been associated for instance to the phenomenon of "priming" (a potentiated induction of defense genes and antimicrobial compounds for protection against biotic and/or abiotic stress). For example, Jaskiewicz et al. (2011) have demonstrated that during the interaction *Arabidopsis thaliana-Pseudomonas syringae* pv. *maculicola*, histone modifications on WRKY gene promoters have been detected in leaves distal to localized foliar infection for an augmented response to secondary stress. Thus, pathogen exposure induces one or more systemic signals that are stored on gene promoters in remote leaves in the form of histone modifications, mainly in the form of trimethylation of Lys 4 on histone H3 (H3K4me3) (Jaskiewicz et al., 2011). In another Arabidopsis study, *Pseudomonas syringae* pv *tomato* DC3000 (*Pst*DC3000) inoculation increased resistance in subsequent generations. Progeny from *Pst*DC3000-inoculated Arabidopsis were primed to activate salicylic acid-inducible defense genes and were more resistant to PstDC3000 (Luna et al., 2012). This transgenerational systemic acquired resistance indicates an epigenetic basis of the phenomenon.

However, neither *P. s.* pv. *maculicola* nor *Pst*DC3000 have SET-domain coding genes. Thus, epigenetic changes can also contribute to and/or result from bacterial infectious diseases. Consequently, many open questions remain. For example, we should ask if events like priming take place only in plants that are

being attacked by pathogens that do not have SET domain proteins, or if all bacterial pathogens induce priming. How immune priming induced protection does takes place in animals infected by pathogenic bacteria containing SET-domain proteins? What's more, it is essential to know how histone modifications might contribute to host response to infection and/or if bacteria take control of histone modifications to drive a transcriptional program beneficial for infection. Also, it is imperative to determine if SET domain proteins present in all pathogenic bacteria with a secretion system are secreted from bacteria, translocated to the host cell nucleus during infection and if they associate with chromatin. With respect to symbionts, do bacterial SET proteins complement the collection of post-translational modifications of their hosts to facilitate bacterial intracellular replication through regulation of its host gene expression? How do hosts benefit from bacteria containing SET-domain proteins in a symbiotic relationship? What group of bacterial SET-domain proteins has evolved into a chromatin-related role similar to their eukaryotic counterparts? How many new histone modifications are performed by bacterial SET-domain proteins, like RomA?

All these questions open new opportunities for future research in the subject of bacterial pathogenesis and chromatin-based regulation of host genes and may help to better understand the pathophysiology of bacterial infections and to develop efficient therapeutic approaches to treat important diseases, as well as to increase crop productivity.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00065/abstract

REFERENCES
Alcaraz, L. D., Olmedo, G., Bonilla, G., Cerritos, R., Hernández, G., Cruz, A., et al. (2008). The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5803–5808. doi: 10.1073/pnas.0800981105

Al-Quadan, T., Price, C. T., and Abu Kwaik, Y. (2012). Exploitation of evolutionarily conserved amoeba and mammalian processes by Legionella. *Trends Microbiol.* 20, 299–306. doi: 10.1016/j.tim.2012.03.005

Alvarez-Venegas, R., Sadder, M., Tikhonov, A., and Avramova, Z. (2007). Origin of the bacterial SET domain genes: vertical or horizontal? *Mol. Biol. Evol.* 24, 482–497. doi: 10.1093/molbev/msl184

Aravind, L., Abhiman, S., and Iyer, L. M. (2011). Natural history of the eukaryotic chromatin protein methylation system. *Prog. Mol. Biol. Transl. Sci.* 10, 105–176. doi: 10.1016/B978-0-12-387685-0.00004-4

Aravind, L., and Iyer, L. M. (2003). Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. *Cell Cycle* 2, 369–376. doi: 10.4161/cc.2.4.419

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature* 447, 407–412. doi: 10.1038/nature05915

Bierne, B., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2:a010272. doi: 10.1101/cshperspect.a010272

Casadio, F., Lu, X., Pollock, S. B., Leroy, G., Garcia, B. A., Muir, T. W., et al. (2013). H3R42me2a is a histone modification with positive transcriptional effects. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14894–14899. doi: 10.1073/pnas.1312925110

Chu, Y., Zhang, Z., Wang, Q., Luo, Y., and Huang, L. (2012). Identification and characterization of a highly conserved crenarchaeal protein lysine methyltransferase with broad substrate specificity. *J. Bacteriol.* 194, 6917–6926. doi: 10.1128/JB.01535-12

Dillon, S. C., Zhang, X., Trievel, R. C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* 6:227. doi: 10.1186/gb-2005-6-8-227

Han, C. S., Xie, G., Challacombe, J. F., Altherr, M. R., Bhotika, S. S., Bruce, D., et al. (2006). Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. *J. Bacteriol.* 188, 3382–3390. doi: 10.1128/JB.188.9.3382-3390.2006

Jaskiewicz, M., Conrath, U., and Peterhänsel, C. (2011). Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Rep.* 12, 50–55. doi: 10.1038/embor.2010.186

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* 54, 80–93. doi: 10.1007/s000180050127

Kornberg, R. D. (1977). Structure of chromatin. *Annu. Rev. Biochem.* 46, 931–954. doi: 10.1146/annurev.bi.46.070177.004435

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Kumar, R., and Rao, D. N. (2013). Role of DNA methyltransferases in epigenetic regulation in bacteria. *Subcell. Biochem.* 61, 81–102. doi: 10.1007/978-94-007-4525-4_4

Lackner, G., Moebius, N., Partida-Martinez, L., and Hertweck, C. (2011). Complete genome sequence of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus microsporus*. *J. Bacteriol.* 193, 783–784. doi: 10.1128/JB.01318-10

Li, T., Lu, Q., Wang, G., Xu, H., Huang, H., Cai, T., et al. (2013). SET-domain bacterial effectors target heterochromatin protein 1 to activate host rDNA transcription. *EMBO Rep.* 14, 733–740. doi: 10.1038/embor.2013.86

Luna, E., Bruce, T. J., Roberts, M. R., Flors, V., and Ton, J. (2012). Next-generation systemic acquired resistance. *Plant Physiol.* 158, 844–853. doi: 10.1104/pp.111.187468

Makarova, K. S., Wolf, Y. I., van der Oost, J., and Koonin, E. V. (2009). Prokaryotic homologs of argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* 4:29. doi: 10.1186/1745-6150-4-29

Manzur, K. L., and Zhou, M.-M. (2005). An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1-α. *FEBS Lett.* 579, 3859–3865. doi: 10.1016/j.febslet.2005.05.026

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., et al. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12902–12907. doi: 10.1073/pnas.231473398

Moran, N. A. (2006). Symbiosis. *Curr. Biol.* 16, R866–R871. doi: 10.1016/j.cub.2006.09.019

Mujtaba, S., Winer, B. Y., Jaganathan, A., Patel, J., Sgobba, M., Schuch, R., et al. (2013). Anthrax SET protein: a potential virulence determinant that epigenetically represses NF-κB activation in infected macrophages. *J. Biol. Chem.* 288, 23458–23472. doi: 10.1074/jbc.M113.467696

Murata, M., Azuma, Y., Miura, K., Rahman, M. A., Matsutani, M., Aoyama, M., et al. (2007). Chlamydial SET domain protein functions as a histone methyltransferase. *Microbiology* 153, 585–592. doi: 10.1099/mic.0.29213-0

Niu, Y., Xia, Y., Wang, S., Li, J., Niu, C., Li, X., et al. (2013). A prototypic lysine methyltransferase 4 from archaea with degenerate sequence specificity methylates chromatin proteins Sul7d and Cren7 in different patterns. *J. Biol. Chem.* 288, 13728–13740. doi: 10.1074/jbc.M113.452979

Partida-Martinez, L. P., Groth, I., Schmitt, I., Richter, W., Roth, M., and Hertweck, C. (2007). *Burkholderia rhizoxinica* sp. nov. and *Burkholderia endofungorum* sp. nov., bacterial endosymbionts of the plant-pathogenic fungus *Rhizopus microsporus*. *Int. J. Syst. Evol. Microbiol.* 57(pt 11), 2583–2590. doi: 10.1099/ijs.0.64660-0

Partida-Martinez, L. P., Monajembashi, S., Greulich, K. O., and Hertweck, C. (2007a). Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism. *Curr. Biol.* 17, 773–777. doi: 10.1016/j.cub.2007.03.039

Pennini, M. E., Perrinet, S., Dautry-Varsat, A., and Subtil, A. (2010). Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis*. *PLoS Pathog.* 6:e1000995. doi: 10.1371/journal.ppat.1000995

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* 28, 1057–1068. doi: 10.1038/nbt.1685

Price, C. T., and Abu Kwaik, Y. (2013). One bacterial effector with two distinct catalytic activities by different strains. *EMBO Rep.* 14, 753–754. doi: 10.1038/embor.2013.126

Raymond, B., Batsche, E., Boutillon, F., Wu, Y.-Z., Leduc, D., Balloy, V., et al. (2009). Anthrax lethal toxin impairs IL-8 expression in epithelial cells through inhibition of histone H3 modification. *PLoS Pathog.* 5:e1000359. doi: 10.1371/journal.ppat.1000359

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406, 593–599. doi: 10.1038/35020506

Reeve, J. N. (2003). Archaeal chromatin and transcription. *Mol. Microbiol.* 48, 587–598. doi: 10.1046/j.1365-2958.2003.03439.x

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., et al. (2001). The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* 20, 7137–7148. doi: 10.1093/emboj/20.24.7137

Rolando, M., Sanulli, S., Rusniok, C., Gomez-Valero, L., Bertholet, C., Sahr, T., et al. (2013). Legionella pneumophila effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. *Cell Host. Microbe.* 13, 395–405. doi: 10.1016/j.chom.2013.03.004

Roulis, E., Polkinghorne, A., and Timms, P. (2013). *Chlamydia pneumoniae*: modern insights into an ancient pathogen. *Trends Microbiol.* 21, 120–128. doi: 10.1016/j.tim.2012.10.009

Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8, 983–994. doi: 10.1038/nrm2298

Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C., and Brown, J. R. (2001). Phylogenetic analysis do not support horizontal gene transfer from bacteria to vertebrates. *Nature* 411, 940–943. doi: 10.1038/35082058

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., et al. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282, 754–759. doi: 10.1126/science.282.5389.754

Epstein–Barr virus–host cell interactions: an epigenetic dialog?

Hans H. Niller[1]*, Kalman Szenthe[2] and Janos Minarovits[3]

[1] Institute of Medical Microbiology and Hygiene, University of Regensburg, Regensburg, Germany
[2] RT-Europe Nonprofit Research Ltd, Mosonmagyaróvár, Hungary
[3] Department of Oral Biology and Experimental Dental Research, Faculty of Dentistry, University of Szeged, Szeged, Hungary

Here, we wish to highlight the genetic exchange and epigenetic interactions between Epstein–Barr virus (EBV) and its host. EBV is associated with diverse lymphoid and epithelial malignancies. Their molecular pathogenesis is accompanied by epigenetic alterations which are distinct for each of them. While lymphoblastoid cell lines derived from B cells transformed by EBV *in vitro* are characterized by a massive demethylation and euchromatinization of the viral and cellular genomes, the primarily malignant lymphoid tumor Burkitt's lymphoma and the epithelial tumors nasopharyngeal carcinoma and EBV-associated gastric carcinoma are characterized by hypermethylation of a multitude of cellular tumor suppressor gene loci and of the viral genomes. In some cases, the viral latency and oncoproteins including the latent membrane proteins LMP1 and LMP2A and several nuclear antigens affect the level of cellular DNA methyltransferases or interact with the histone modifying machinery. Specific molecular mechanisms of the epigenetic dialog between virus and host cell remain to be elucidated.

Keywords: chromatin, CpG island, epigenome, viral latency, lytic cycle, pioneer transcription factor, tumor suppressor gene, tumor virus

EPSTEIN–BARR VIRUS—THE FIRST HUMAN TUMOR VIRUS

Epstein–Barr virus (EBV), the proto-typical gamma-herpesvirus infecting humans, has been discovered 50 years ago in cultured Burkitt's lymphoma (BL) cells (Epstein et al., 1964). EBV physiologically homes to memory B cells, a property which is reflected in the genus name *Lymphocryptovirus* for gamma-herpesviruses which "hide in lymphoid cells." EBV is the causative agent of infectious mononucleosis, and was the first known human tumor virus. In addition to BL, an endemic childhood tumor mainly of equatorial Africa, EBV plays a role in the origin or progression of other primarily malignant B cell tumors, such as Hodgkin lymphoma, and diverse AIDS-associated lymphomas including lymphomas of the central nervous system. A profound failure of T cell surveillance may allow early-onset post-transplant lymphoproliferative disorder (PTLD) which may be overcome upon a timely onset of T cell control, but may turn malignant if growing uncontrolled for too long a time span.

In the first decade of EBV research, EBV was considered a purely lymphotropic virus. Finding viral DNA in cellular DNA from biopsies of anaplastic carcinomas of the nasopharynx (NPC, nasopharyngeal carcinoma) by DNA hybridization did not change that general view, because NPC as a lymphoepithelial tumor contains a great many of infiltrating lymphocytes (zur Hausen et al., 1970). Localizing the virus specifically to the malignant epithelial cells first established EBV infection of non-lymphatic cells and paved the way for the novel concept of EBV as an epithelial tumor virus (Wolf et al., 1973). NPC is an endemic tumor with a strong preference for South East Asia, especially

Guangdong and Hong Kong, with an incidence rate of 20–30 cases per 100,000 persons per year, and virtually 100% of non-keratinizing and undifferentiated NPCs are EBV-associated. Almost two decades later on, the association of EBV with gastric carcinomas (EBVaGC) was established, too (Burke et al., 1990; Shibata et al., 1991). Contrary to the endemic tumor NPC, about 10% of the worldwide sporadic tumor gastric carcinoma (GC) are associated with EBV infection. Among gastric remnant carcinomas approximately 30% and among lymphoepithelioma-like GCs, approximately 80% are EBV-associated. Altogether, EBVaGC with an estimated more than 80,000 cases per year is probably the most frequent EBV-associated malignancy worldwide (reviewed in Niller et al., 2014a,b). Beyond the B cell lineage, EBV infection is currently associated with T cell lymphomas, epithelial tumors, and rarely with leiomyosarcoma, a neoplasm of mesodermal origin (McClain et al., 1995). Furthermore, the risk of autoimmune disease, including multiple sclerosis, is significantly increased after primary EBV-infection, and even more so after symptomatic primary infection, i.e., mononucleosis (Niller et al., 2008).

PATHOGENESIS OF EBV-ASSOCIATED TUMORS

The spectacular ability of EBV to transform and immortalize B cells dominated the first four decades of EBV research and tumor virology. General view was that the EBV-transformed cell was the origin of the endemic BL cell, too, although a fundamental difference of the epidemiology and pathogenesis between lymphoblastoid cell line (LCL)-like tumors on one side (early onset PTLD), and of primarily malignant EBV-associated lymphomas

Table 1 | Host cell-dependent expression of latent Epstein–Barr virus proteins.

Protein	Cell type and latency type			
	BL	NPC	CLL	LCL
	Type I	Type IIa	Type IIb	Type III
EBNA1	+	+	+	+
EBNA2	−	−	+	+
LMP1	−	+/−	−	+
LMP2A	−	+	+	+
BARF1	−	+	−	−

Key viral latency proteins or so-called oncoproteins are differentially expressed in EBV-associated malignancies and transformed cell lines. Depending on the specific viral gene expression patterns, specific viral latency classes are defined. BL, Burkitt's lymphoma; NPC, nasopharyngeal carcinoma; CLL, chronic lympho-cytic leukemia; LCL, lymphoblastoid cell line; EBNA, Epstein–Barr virus nuclear antigen; LMP, latent membrane protein; BARF1, BamHI A rightward frame 1.

on the other side (endemic BL) became evident. Early onset PTLD originate under conditions of severe immune suppression and depend on viral transforming functions, including EBV nuclear antigens (EBNAs) and latent membrane proteins (LMPs) that are expressed both in PTLDs *in vivo* and in LCLs immortalized *in vitro*. On the other hand, BL and Hodgkin lymphoma originate under conditions of hyperstimulation of the lymphoid germinal center reaction, and they do not depend on EBNA2 which the vast majority of them do not express (Klein, 1987; Lenoir and Bornkamm, 1987; **Table 1**). In this context, it is important to distinguish between morphological and oncogenic transformation (Niller et al., 2011). Our discovery of a binding site for the oncoprotein c-Myc in the central locus control region of the EBV genome suggested that the molecular pathogenesis of endemic BL does not depend on a previous EBNA2-transformed state of the B cell, but mostly on a dysbalance of pro- and anti-apoptotic functions in consequence of Myc-translocation, a molecular accident in a virus-infected B cell undergoing the germinal center reaction (Niller et al., 2003). The need to counter-balance the pro-apoptotic force of translocated *c-Myc* through anti-apoptotic functions, either encoded by the viral genome or induced by virus infection, in order for a BL to emerge has recently been re-emphasized (Mbulaiteye, 2013; Westhoff Smith and Sugden, 2013; Rickinson, 2014). Our differential pathogenesis model for EBV-associated lymphomas was controversial at first (Rossi and Bonetti, 2004; Thorley-Lawson, 2004; reviewed in Niller et al., 2012), but has now gained strong support by recent large-scale epigenomic analyses of LCLs and tumor cells (see below). Thus, finding the binding site for the oncoprotein c-Myc in the locus control region of EBV caused a conceptual shift away from the morphologically transformed cell and has turned out as a heuristic discovery (Niller et al., 2003).

GENETIC EXCHANGE BETWEEN HERPESVIRUSES AND HOST CELLS

Homology between herpesviral and human genes is now a common theme which was first highlighted by the finding of a gene for a functional thymidylate synthase (TS) in the genome of her-

pesvirus saimiri with an extremely high homology of 70% identical amino acids with the human *TS* gene. Various parameters suggested that the *TS* gene had been acquired in virus evolution by an ancestral herpesvirus from the cellular genome (Bodemer et al., 1986; Honess et al., 1986). The exchange of human and viral genes, and in the case of human herpesvirus 6 (HHV-6) the invasion of an entire herpesviral genome into the human germ line (Daibata et al., 1999) in about 0.8% of humans, must have happened on numerous occasions in evolutionary time. In the case of EBV, the intimate evolutionary relationship of virus and host cell is emphasized by the presence of several viral genes with sequence homology to cellular genes, i.e., *BHRF1* and *BALF1*, two anti-apoptotic *BCL2* homologs, *BILF1*, coding for a constitutively active G protein-coupled receptor (GPCR) homolog, and *BCRF1*, an *IL-10* (*interleukin 10*) gene homolog. The BCRF1 protein appears to be a functional homolog of IL-10, an immune suppressive cytokine secreted by Th2 cells with a sequence identity of about 70% (Hsu et al., 1990). The sequence homologies of the other three viral peptides are of lower degree than that. BHRF1 carries a 25% identity of a 150 amino acid C-terminal portion with the anti-apoptotic cellular protein BCL2 (Cleary et al., 1986). BALF1 shows homology at a similar degree in functionally important domains to its cellular counterparts BCL2 and BCLX (Marshall et al., 1999). BILF1 carries a homology of around 20% with the human chemokine receptor CXCR3A (Davis-Poynter and Farrell, 1996; Beisser et al., 2005; Paulsen et al., 2005). EBV-associated carcinoma cells express BARF1, identified initially as a lytic cycle protein that binds to hCSF-1 (human colony stimulating factor 1) as a viral decoy receptor, although its crystal structure is most closely related to CD80, a co-stimulatory molecule of antigen presenting cells (Seto et al., 2005; Tarbouriech et al., 2006; Elegheert et al., 2012).

Herpesviral DNA-binding proteins ICP8 of HSV1 and BALF2 of EBV, and additional homologous proteins of human herpesviruses which are required for viral replication belong to a class of "DDE/RNase H-like fold-family" nucleases, together with the recombination activating gene (RAG) 1 protein, essential for V(D)J recombination, and the Argonaute protein of the RNA-induced silencing complex (RISC). For ICP8 of HSV, divalent cation binding of the DDE-site was actually shown to be functional and required for viral replication (Bryant et al., 2012). Furthermore, inhibitors of HIV integrase, another RNase H-fold protein, inhibited the replication of viruses from all herpesvirus genera, too (Yan et al., 2014). Based on a co-regulatory transcriptional network for both *RAG-1* and *RAG-2*, and the genes for herpesviral DNA-binding proteins, and based on signature sequence homologies between *V(D)J* recombination sites and the viral terminal repeats, an evolutionary relationship between the RAG recombinase and herpesviral DNA-binding proteins was proposed. The *RAG* locus may have originally been introduced to host cells by a primordial herpesvirus (Dreyfus, 2009). We found a striking co-linearity of structural and functional elements between the cellular immunoglobulin gene loci and the left part of the EBV genome. Therefore, although speculative, we agree with the view that, in the case of the *RAG* genes, the appearance of the adaptive immune system may have been dependent on a primordial herpesvirus genome and, in the case of the B cell, may

have developed further as consequence of an evolutionary ping-pong game between EBV and the host cell (Niller et al., 2004).

EPIGENETIC INTERACTIONS BETWEEN EBV AND ITS HOST CELLS

Epstein–Barr virus infects both B lymphocytes and epithelial cells *in vivo*. It enters B cells after binding to CD21, a cell surface molecule absent from epithelial cells. However, EBV-infected B cells are capable to transfer the virus to epithelial cells lacking the EBV receptor. Both B cells and epithelial cells can support productive (lytic) EBV replication when all of the proteins and non-translated RNAs encoded by the viral genome are expressed in a sequential order. EBV also causes latent infections, typically in resting memory B cells and in various neoplastic cells that usually carry circular, double stranded viral genomes. During latency, only a restricted set of EBV promoters is active. The activity of latent EBV promoters depends on the phenotype of host cells, and it is controlled by the epigenetic regulatory machinery. Based on the epigenetic marks deposited on the viral chromatin by the cellular epigenetic machinery one can distinguish between viral epigenotypes that are associated with unique patterns of viral gene expression (reviewed in Minarovits, 2006). In parallel, certain latent EBV proteins characteristic for the major cell types carrying latent EBV genomes act as epigenetic regulators themselves: they alter the cellular epigenotype and gene expression pattern and may contribute to the development of malignant tumors. Thus, the situation is similar to an "epigenetic dialog," indeed. Typical examples of EBV latency types are summarized in **Table 1**, based on the nomenclature suggested by Klein et al. (2013; see also Laytragoon-Lewin et al., 1995; Niller et al., 2012).

EPIGENETIC CONTROL OF EBV LATENCY PROMOTERS

The epigenetic regulatory mechanisms of host cells not only control the preservation of cell type-specific gene expression patterns from cell generation to cell generation, but ensure the maintenance of host cell-dependent usage of latent EBV promoters as well. Epigenetic regulation is based on writing, reading, and erasing epigenetic marks on chromatin as well as on protein–DNA interactions that are stable even in mitotic, highly condensed chromatin (reviewed in Gopalakrishnan et al., 2008; Zaret et al., 2008; Sharma et al., 2010; Blomen and Boonstra, 2011). Euchromatic marks favor transcription whereas heterochromatic marks are associated with a more condensed chromatin structure that usually represses promoter activity.

DNA methylation is involved in silencing of most latent EBV promoters. It is well documented that the alternative promoters Wp and Cp, where transcripts coding for six EBNAs are initiated can be switched off by CpG methylation. In addition, LMP1p and LMP2Ap, the promoters for LMP1 and LMP2A transcripts are silenced by the activity of cellular DNA methyltransferases, too. DNA methylation does not play a role, however, in switching off Qp, an alternative promoter for EBNA1 transcripts (reviewed in Li and Minarovits, 2003; Niller et al., 2012). One may speculate, that EBNA2, the major transactivator protein encoded by the EBV genome that interacts with both histone acetyltransferases and histone deacetylases (reviewed in

Niller et al., 2009) may activate key cellular genes mediating genome-wide demethylation in LCLs. Acetylation of histone H3 and histone H4 molecules and di- or trimethylation of lysine 4 of histone H3 (H3K4me2 or H3K4me3) are euchromatic marks frequently associated with active Cp, Qp, LMP1p, and LMP2Ap (reviewed in Niller et al., 2012; Arvey et al., 2013). In principle, complexes formed by Polycomb and Trithorax group proteins that modify histone tails could also play a role in the regulation of latent EBV promoters but there are no data supporting such a mechanism. In contrast, Polycomb group protein EZH2 was observed to leave a heterochromatic histone mark (H3K27me3) at early lytic promoters of EBV that are silent during latency, and chromatin immunoprecipitation proved the association of the EZH2 methyltransferase with this class of viral promoters in the BL cell line Raji (Woellmer et al., 2012). The immediate-early promoter Zp, where transcripts for the lytic cycle initiating BZLF1 protein are initiated was also found to be repressed by H3K27me3 but also by H4K20me3 in Raji cells (Murata et al., 2012).

Pioneer transcription factors and variant histone molecules that bind to repressive chromatin areas and mark the genes to be activated have not been implicated in the control of EBV latency. Chromatin loops formed by binding of insulator proteins to EBV episomes may play a role, however, in the epigenetic regulation of the latency promoters Cp and Qp (Tempera et al., 2011).

EBV-ENCODED ONCOPROTEINS AS EPIGENETIC REGULATORS

The methylation patterns of NPC cells and EBVaGC cells differ from their normal counterparts: these neoplastic cells regularly carry hypermethylated genomic regions with silenced cellular promoters (Iizasa et al., 2012; Lo et al., 2012). Focal hypermethylations frequently inactivate tumor suppressor genes and may contribute to carcinogenesis and tumor progression. Because the EBV-encoded transmembrane proteins LMP1 and LMP2A are capable of upregulating the cellular DNA methyltransferases DNMT1, DNMT3A, and DNMT3B, it was suggested that hypermethylation of CpG rich sequences, the so called CpG islands, is mediated by LMP1 or LMP2A in EBV-associated carcinomas (reviewed in Niller et al., 2012). LMP1 and LMP2A are expressed in EBV positive Hodgkin lymphomas as well. Thus, they may contribute to gene silencing in these neoplasms, too.

It is worthy to note that, similarly to the EBV-associated carcinomas, LCLs established by *in vitro* EBV-infection of B cells also express LMP1 and LMP2A that could potentially upregulate DNA methyltransferases. It was observed, however, that the typical epigenetic change in LCLs is a widespread demethylation of the B cell methylome affecting one third of all cellular genes and 2.18 GB of the genome (Hansen et al., 2014). The mechanism of demethylation remains to be established. Furthermore, the EBV episomes carried by LCLs are also hypomethylated, in contrast to the overall hypermethylation of EBV genomes in latency type I BL lines, BL biopsies, and EBV-associated carcinomas (Minarovits et al., 1991; Fernandez et al., 2009). The viral oncoproteins EBNA3A and EBNA3C expressed during *in vitro* immortalization of B cells silence distinct cellular tumor suppressor genes by depositing a heterochromatic histone mark

via the Polycomb repressor complex PCR2 (reviewed in Allday, 2013). The dominant change in EBV immortalized B cells seems to be, however, a genome-wide decrease and redistribution of heterochromatic marks (Hernando et al., 2014). The EBV latency products eliciting the reprograming of the host cell epigenome in LCLs remain to be elucidated. Contrary to LCLs, primarily malignant lymphomas, i.e., BL and other lymphomas, are characterized not by a massive and wide-spread hypomethylation, but by a local hypermethylation of selected genomic loci (Martin-Subero et al., 2009a,b; Kreck et al., 2013; reviewed in Niller et al., 2014b).

In contrast to BL tumors, in which hypermethylated loci are strongly enriched for polycomb repressive complex (PRC) target genes of embryonic stem cells, hypermethylated genes in EBVaGC are not enriched for PRC targets (Martin-Subero et al., 2009a; Matsusaka et al., 2011). Contrary to *Helicobacter pylori*-associated GC, EBVaGC does not emerge from an "epigenetic field" in the gastric mucosa (Matsusaka et al., 2011; Niller et al., 2014a). Thus, EBV-associated epigenetic changes may quite quickly set the stage for malignancy (Au et al., 2005; Niller et al., 2014a). Notably, even transient EBV infection of epithelial cells leaves permanent epigenetic scars indicating past infection (Queen et al., 2013; Birdwell et al., 2014).

Although there are no data as to the interaction of host cell-encoded pioneer transcription factors and the EBV genome, certain viral proteins that bind to both viral and cellular DNA and remain associated with mitotic chromosomes may act as pioneer transcription factors. EBNA1, a nuclear protein expressed in all EBV latency types is a putative pioneer factor functionally resembling transcription factors of the FoxA family that act as epigenetic regulators controlling important developmental processes (Niller et al., 2012). EBNA1 was shown to bind to different sets of cellular promoters in cell lines of epithelial and lymphoid origin, and upregulation as well as downregulation of distinct gene batteries was observed (Canaan et al., 2009). In the BL cell line Raji, EBNA1 was shown to interact with a large number of cellular genes as well as LINE elements, and high affinity EBNA-binding sites were observed in a repetitive sequence in chromosome 11 (Lu et al., 2010).

BZLF1, an immediate-early EBV protein initiating productive viral replication that preferentially binds methylated DNA sequences was also suggested to act as a pioneer transcription factor (Woellmer et al., 2012). BZLF1 is expressed not only during the lytic cycle but transiently also during the establishment of latent infection of B cells *in vitro* (Kalla and Hammerschmidt, 2012). Because BZLF1 can bind to cellular promoters (Lan et al., 2013) and elicit epigenetic alterations (Woellmer et al., 2012), an "epigenetic dialog" between the latent EBV genomes and the host cell genome may occur, indeed: transient BZLF1 expression may change the cellular epigenotype followed by silencing of the BZLF1 promoter through the cellular epigenetic machinery. In parallel, EBV-encoded epigenetic regulators may leave their marks on the cellular epigenotype in a next phase of EBV-mediated B cell immortalization.

In conclusion, on an evolutionary time scale a genetic exchange between herpesviruses and their hosts is evident. Beginning with the early steps of viral infection, epigenetic interactions between virus and host cell are taking place. The multi-tiered epigenetic dialog between EBV and its host needs to be elucidated in greater molecular detail in order to understand the diverse outcomes of infection.

REFERENCES

Allday, M. J. (2013). EBV finds a polycomb-mediated, epigenetic solution to the problem of oncogenic stress responses triggered by infection. *Front. Genet.* 4:212. doi: 10.3389/fgene.2013.00212

Arvey, A., Tempera, I., and Lieberman, P. M. (2013). Interpreting the Epstein–Barr virus (EBV) epigenome using high-throughput data. *Viruses* 5, 1042–1054. doi: 10.3390/v5041042

Au, W. Y., Pang, A., Chan, E. C., Chu, K. M., Shek, T. W., and Kwong, Y. L. (2005). Epstein–Barr virus-related gastric adenocarcinoma: an early secondary cancer post hemopoietic stem cell transplantation. *Gastroenterology* 129, 2058–2063. doi: 10.1053/j.gastro.2005.10.011

Beisser, P. S., Verzijl, D., Gruijthuijsen, Y. K., Beuken, E., Smit, M. J., Leurs, R., et al. (2005). The Epstein–Barr virus BILF1 gene encodes a G protein-coupled receptor that inhibits phosphorylation of RNA-dependent protein kinase. *J. Virol.* 79, 441–449. doi: 10.1128/JVI.79.1.441-449.2005

Birdwell, C. E., Queen, K. J., Kilgore, P., Rollyson, P., Trutschl, M., Cvek, U., et al. (2014). Genome-wide DNA methylation as an epigenetic consequence of Epstein–Barr virus infection of immortalized keratinocytes. *J. Virol.* 88, 11442–11458. doi: 10.1128/JVI.00972-14

Blomen, V. A., and Boonstra, J. (2011). Stable transmission of reversible modifications: maintenance of epigenetic information through the cell cycle. *Cell. Mol. Life Sci.* 68, 27–44. doi: 10.1007/s00018-010-0505-5

Bodemer, W., Niller, H. H., Nitsche, N., Scholz, B., and Fleckenstein, B. (1986). Organization of the thymidylate synthase gene of herpesvirus saimiri. *J. Virol.* 60, 114–123.

Bryant, K. F., Yan, Z., Dreyfus, D. H., and Knipe, D. M. (2012). Identification of a divalent metal cation binding site in herpes simplex virus 1 (HSV-1) ICP8 required for HSV replication. *J. Virol.* 86, 6825–6834. doi: 10.1128/JVI.00374-12

Burke, A. P., Yen, T. S., Shekitka, K. M., and Sobin, L. H. (1990). Lymphoepithelial carcinoma of the stomach with Epstein–Barr virus demonstrated by polymerase chain reaction. *Mod. Pathol.* 3, 377–380.

Canaan, A., Haviv, I., Urban, A. E., Schulz, V. P., Hartman, S., Zhang, Z., et al. (2009). EBNA1 regulates cellular gene expression by binding cellular promoters. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22421–22426. doi: 10.1073/pnas.0911676106

Cleary, M. L., Smith, S. D., and Sklar, J. (1986). Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. *Cell* 47, 19–28. doi: 10.1016/0092-8674(86)90362-4

Daibata, M., Taguchi, T., Nemoto, Y., Taguchi, H., and Miyoshi, I. (1999). Inheritance of chromosomally integrated human herpesvirus 6 DNA. *Blood* 94, 1545–1549.

Davis-Poynter, N. J., and Farrell, H. E. (1996). Masters of deception: a review of herpesvirus immune evasion strategies. *Immunol. Cell Biol.* 74, 513–522. doi: 10.1038/icb.1996.84

Dreyfus, D. H. (2009). Paleo-immunology: evidence consistent with insertion of a primordial herpes virus-like element in the origins of acquired immunity. *PLoS ONE* 4:e5778. doi: 10.1371/journal.pone.0005778

Elegheert, J., Bracke, N., Pouliot, P., Gutsche, I., Shkumatov, A. V., Tarbouriech, N., et al. (2012). Allosteric competitive inactivation of hematopoietic CSF-1 signaling by the viral decoy receptor BARF1. *Nat. Struct. Mol. Biol.* 19, 938–947. doi: 10.1038/nsmb.2367

Epstein, M. A., Achong, B. G., and Barr, Y. M. (1964). Virus particles in cultured lymphoblasts from Burkitt's Lymphoma. *Lancet* 1, 702–703. doi: 10.1016/S0140-6736(64)91524-7

Fernandez, A. F., Rosales, C., Lopez-Nieva, P., Grana, O., Ballestar, E., Ropero, S., et al. (2009). The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res.* 19, 438–451. doi: 10.1101/gr.083550.108

Gopalakrishnan, S., Van Emburgh, B. O., and Robertson, K. D. (2008). DNA methylation in development and human disease. *Mutat. Res.* 647, 30–38. doi: 10.1016/j.mrfmmm.2008.08.006

Hansen, K. D., Sabunciyan, S., Langmead, B., Nagy, N., Curley, R., Klein, G., et al. (2014). Large-scale hypomethylated blocks associated with

Epstein–Barr virus-induced B-cell immortalization. *Genome Res.* 24, 177–184. doi: 10.1101/gr.157743.113

Hernando, H., Islam, A. B., Rodriguez-Ubreva, J., Forne, I., Ciudad, L., Imhof, A., et al. (2014). Epstein–Barr virus-mediated transformation of B cells induces global chromatin changes independent to the acquisition of proliferation. *Nucleic Acids Res.* 42, 249–263. doi: 10.1093/nar/gkt886

Honess, R. W., Bodemer, W., Cameron, K. R., Niller, H. H., Fleckenstein, B., and Randall, R. E. (1986). The A + T-rich genome of Herpesvirus saimiri contains a highly conserved gene for thymidylate synthase. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3604–3608.

Hsu, D. H., De Waal Malefyt, R., Fiorentino, D. F., Dang, M. N., Vieira, P., De Vries, J., et al. (1990). Expression of interleukin-10 activity by Epstein–Barr virus protein BCRF1. *Science* 250, 830–832. doi: 10.1126/science.2173142

Iizasa, H., Nanbo, A., Nishikawa, J., Jinushi, M., and Yoshiyama, H. (2012). Epstein–Barr virus (EBV)-associated gastric carcinoma. *Viruses* 4, 3420–3439. doi: 10.3390/v4123420

Kalla, M., and Hammerschmidt, W. (2012). Human B cells on their route to latent infection—early but transient expression of lytic genes of Epstein–Barr virus. *Eur. J. Cell Biol.* 91, 65–69. doi: 10.1016/j.ejcb.2011.01.014

Klein, E., Nagy, N., and Rasul, A. E. (2013). EBV genome carrying B lymphocytes that express the nuclear protein EBNA-2 but not LMP-1: type IIb latency. *Oncoimmunology* 2:e23035. doi: 10.4161/onci.23035

Klein, G. (1987). "In defense of the "old" Burkitt lymphoma scenario," in *Advances in Viral Oncology*, ed. G. Klein (New York: Raven Press), 207–211.

Kreck, B., Richter, J., Ammerpohl, O., Barann, M., Esser, D., Petersen, B. S., et al. (2013). Base-pair resolution DNA methylome of the EBV-positive endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing. *Leukemia* 27, 1751–1753. doi: 10.1038/leu.2013.4

Lan, Y. Y., Yeh, T. H., Lin, W. H., Wu, S. Y., Lai, H. C., Chang, F. H., et al. (2013). Epstein–Barr virus Zta upregulates matrix metalloproteinases 3 and 9 that synergistically promote cell invasion in vitro. *PLoS ONE* 8:e56121. doi: 10.1371/journal.pone.0056121

Laytragoon-Lewin, N., Chen, F., Avila-Carino, J., Zou, J. Z., Mellstedt, H., Ernberg, I., et al. (1995). Epstein Barr virus (EBV)-carrying cells of a chronic lymphocytic leukemia (CLL) subpopulation express EBNA1 and LMPs but not EBNA2 in vivo. *Int. J. Cancer* 63, 486–490.

Lenoir, G. M., and Bornkamm, G. W. (1987). "Burkitt's lymphoma, a human cancer model for the study of the multistep development of cancer: proposal for a new scenario," in *Advances in Viral Oncology*, ed. G. Klein (New York: Raven Press), 173–206.

Li, H., and Minarovits, J. (2003). Host cell-dependent expression of latent Epstein–Barr virus genomes: regulation by DNA methylation. *Adv. Cancer Res.* 89, 133–156. doi: 10.1016/S0065-230X(03)01004-2

Lo, K. W., Chung, G. T., and To, K. F. (2012). Deciphering the molecular genetic basis of NPC through molecular, cytogenetic, and epigenetic approaches. *Semin. Cancer Biol.* 22, 79–86. doi: 10.1016/j.semcancer.2011.12.011

Lu, F., Wikramasinghe, P., Norseen, J., Tsai, K., Wang, P., Showe, L., et al. (2010). Genome-wide analysis of host-chromosome binding sites for Epstein–Barr virus nuclear antigen 1 (EBNA1). *Virol. J.* 7:262. doi: 10.1186/1743-422X-7-262

Marshall, W. L., Yim, C., Gustafson, E., Graf, T., Sage, D. R., Hanify, K., et al. (1999). Epstein–Barr virus encodes a novel homolog of the bcl-2 oncogene that inhibits apoptosis and associates with Bax and Bak. *J. Virol.* 73, 5181–5185.

Martin-Subero, J. I., Ammerpohl, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., et al. (2009a). A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. *PLoS ONE* 4:e6986. doi: 10.1371/journal.pone.0006986

Martin-Subero, J. I., Kreuz, M., Bibikova, M., Bentink, S., Ammerpohl, O., Wickham-Garcia, E., et al. (2009b). New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood* 113, 2488–2497. doi: 10.1182/blood-2008-04-152900

Matsusaka, K., Kaneda, A., Nagae, G., Ushiku, T., Kikuchi, Y., Hino, R., et al. (2011). Classification of Epstein–Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. *Cancer Res.* 71, 7187–7197. doi: 10.1158/0008-5472.CAN-11-1349

Mbulaiteye, S. M. (2013). Burkitt Lymphoma: beyond discoveries. *Infect. Agent. Cancer* 8, 35. doi: 10.1186/1750-9378-8-35

McClain, K. L., Leach, C. T., Jenson, H. B., Joshi, V. V., Pollock, B. H., Parmley, R. T., et al. (1995). Association of Epstein–Barr virus with leiomyosarcomas in children with AIDS. *N. Engl. J. Med.* 332, 12–18.

Minarovits, J. (2006). Epigenotypes of latent herpesvirus genomes. *Curr. Top. Microbiol. Immunol.* 310, 61–80. doi: 10.1007/3-540-31181-5_5

Minarovits, J., Minarovits-Kormuta, S., Ehlin-Henriksson, B., Falk, K., Klein, G., and Ernberg, I. (1991). Host cell phenotype-dependent methylation patterns of Epstein–Barr virus DNA. *J. Gen. Virol.* 72, 1591–1599. doi: 10.1099/0022-1317-72-7-1591

Murata, T., Kondo, Y., Sugimoto, A., Kawashima, D., Saito, S., Isomura, H., et al. (2012). Epigenetic histone modification of Epstein–Barr virus BZLF1 promoter during latency and reactivation in Raji cells. *J. Virol.* 86, 4752–4761. doi: 10.1128/JVI.06768-11

Niller, H. H., Banati, F., Ay, E., and Minarovits, J. (2012). "Epigenetic changes in virus-associated neoplasms," in *Patho-Epigenetics of Disease*, eds J. Minarovits and H. H. Niller (New York: Springer), 179–225.

Niller, H. H., Banati, F., and Minarovits, J. (2014a). Epigenetic alterations in nasopharyngeal carcinoma and Epstein–Barr virus (EBV) associated gastric carcinoma—a lesson in contrasts. *J. Nasopharyng. Carcinoma* 1:e9. doi: http://dx.doi.org/10.15383/JNPC.9

Niller, H. H., Tarnai, Z., Decsi, G., Zsedenyi, A., Banati, F., et al. (2014b). Role of epigenetics in EBV regulation and pathogenesis. *Future Microbiol.* 9, 747–756. doi: 10.2217/fmb.14.41

Niller, H. H., and Minarovits, J. (2012). Similarities between the Epstein–Barr virus (EBV) nuclear protein EBNA1 and the pioneer transcription factor FoxA: is EBNA1 a "bookmarking" oncoprotein that alters the host cell epigenotype? *Pathogens* 1, 37–51. doi: 10.3390/pathogens1010037

Niller, H. H., Salamon, D., Ilg, K., Koroknai, A., Banati, F., Bauml, G., et al. (2003). The in vivo binding site for oncoprotein c-Myc in the promoter for Epstein–Barr virus (EBV) encoding RNA (EBER) 1 suggests a specific role for EBV in lymphomagenesis. *Med. Sci. Monit.* 9, HY1–HY9.

Niller, H. H., Salamon, D., Rahmann, S., Ilg, K., Koroknai, A., Banati, F., et al. (2004). A 30 kb region of the Epstein–Barr virus genome is colinear with the rearranged human immunoglobulin gene loci: implications for a "ping-pong evolution" model for persisting viruses and their hosts. A review. *Acta Microbiol. Immunol. Hung.* 51, 469–484. doi: 10.1556/AMicr.51.2004.4.7

Niller, H. H., Wolf, H., and Minarovits, J. (2008). Regulation and dysregulation of Epstein–Barr virus latency: implications for the development of autoimmune diseases. *Autoimmunity* 41, 298–328. doi: 10.1080/08916930802024772

Niller, H. H., Wolf, H., and Minarovits, J. (2009). Epigenetic dysregulation of the host cell genome in Epstein–Barr virus-associated neoplasia. *Semin. Cancer Biol.* 19, 158–164. doi: 10.1016/j.semcancer.2009.02.012

Niller, H. H., Wolf, H., and Minarovits, J. (2011). Viral hit and run-oncogenesis: genetic and epigenetic scenarios. *Cancer Lett.* 305, 200–217. doi: 10.1016/j.canlet.2010.08.007

Paulsen, S. J., Rosenkilde, M. M., Eugen-Olsen, J., and Kledal, T. N. (2005). Epstein–Barr virus-encoded BILF1 is a constitutively active G protein-coupled receptor. *J. Virol.* 79, 536–546. doi: 10.1128/JVI.79.1.536-546.2005

Queen, K. J., Shi, M., Zhang, F., Cvek, U., and Scott, R. S. (2013). Epstein–Barr virus-induced epigenetic alterations following transient infection. *Int. J. Cancer* 132, 2076–2086. doi: 10.1002/ijc.27893

Rickinson, A., B. (2014). Co-infections, inflammation and oncogenesis: future directions for EBV research. *Semin. Cancer Biol.* 26, 99–115. doi: 10.1016/j.semcancer.2014.04.004

Rossi, G., and Bonetti, F. (2004). EBV and Burkitt's lymphoma. *N. Engl. J. Med.* 350, 2621. doi: 10.1056/NEJM200406173502518

Seto, E., Yang, L., Middeldorp, J., Sheen, T. S., Chen, J. Y., Fukayama, M., et al. (2005). Epstein–Barr virus (EBV)-encoded BARF1 gene is expressed in nasopharyngeal carcinoma and EBV-associated gastric carcinoma tissues in the absence of lytic gene expression. *J. Med. Virol.* 76, 82–88. doi: 10.1002/jmv.20327

Sharma, S., Kelly, T. K., and Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis* 31, 27–36. doi: 10.1093/carcin/bgp220

Shibata, D., Tokunaga, M., Uemura, Y., Sato, E., Tanaka, S., and Weiss, L. M. (1991). Association of Epstein–Barr virus with undifferentiated gastric carcinomas

with intense lymphoid infiltration. Lymphoepithelioma-like carcinoma. *Am. J. Pathol.* 139, 469–474.

Tarbouriech, N., Ruggiero, F., De Turenne-Tessier, M., Ooka, T., and Burmeister, W. P. (2006). Structure of the Epstein–Barr virus oncogene BARF1. *J. Mol. Biol.* 359, 667–678. doi: 10.1016/j.jmb.2006.03.056

Tempera, I., Klichinsky, M., and Lieberman, P. M. (2011). EBV latency types adopt alternative chromatin conformations. *PLoS Pathog.* 7:e1002180. doi: 10.1371/journal.ppat.1002180

Thorley-Lawson, D. A. (2004). EBV and Burkitt's lymphoma. *N. Engl. J. Med.* 350, 2621.

Westhoff Smith, D., and Sugden, B. (2013). Potential cellular functions of Epstein–Barr nuclear antigen 1 (EBNA1) of Epstein–Barr virus. *Viruses* 5, 226–240. doi: 10.3390/v5010226

Woellmer, A., Arteaga-Salas, J. M., and Hammerschmidt, W. (2012). BZLF1 governs CpG-methylated chromatin of Epstein–Barr virus reversing epigenetic repression. *PLoS Pathog.* 8:e1002902. doi: 10.1371/journal.ppat. 1002902

Wolf, H., zur Hausen, H., and Becker, V. (1973). EB viral genomes in epithelial nasopharyngeal carcinoma cells. *Nat. New Biol.* 244, 245–247. doi: 10.1038/newbio244245a0

Yan, Z., Bryant, K. F., Gregory, S. M., Angelova, M., Dreyfus, D. H., Zhao, X. Z., et al. (2014). HIV integrase inhibitors block replication of alpha-, beta-, and gammaherpesviruses. *MBio* 5:e01318-14. doi: 10.1128/mBio. 01318-14

Zaret, K. S., Watts, J., Xu, J., Wandzioch, E., Smale, S. T., and Sekiya, T. (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb. Symp. Quant. Biol.* 73, 119–126. doi: 10.1101/sqb.2008.73.040

zur Hausen, H., Schulte-Holthausen, H., Klein, G., Henle, W., Henle, G., Clifford, P., and Santesson, L. (1970). EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature* 228, 1056–1058.

Epigenetic control of mobile DNA as an interface between experience and genome change

James A. Shapiro *

Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA

Mobile DNA in the genome is subject to RNA-targeted epigenetic control. This control regulates the activity of transposons, retrotransposons and genomic proviruses. Many different life history experiences alter the activities of mobile DNA and the expression of genetic loci regulated by nearby insertions. The same experiences induce alterations in epigenetic formatting and lead to trans-generational modifications of genome expression and stability. These observations lead to the hypothesis that epigenetic formatting directed by non-coding RNA provides a molecular interface between life history events and genome alteration.

Keywords: mutation, evolution, natural genetic engineering, mobile DNA, viruses, mobile genetic elements, non-coding RNA

INTRODUCTION

Understanding the functional organization of the genome and its evolutionary history are key goals of modern molecular biology. The task has become more interesting and complex as we learn more the details of cell processes involving the genome. Recent applications of high resolution technologies to genome expression in animals reveal a dynamic four-dimensional interactive control architecture incompatible with prior notions that genomes contain discrete functional segments of DNA ("genes") (Mercer and Mattick, 2013). This review will focus on the role of epigenetic regulation of viruses and mobile genetic elements as a key interface between the activities of these agents of evolutionary change and inputs from cell and organism life histories. The hypothesis developed as a result of the review is that disruption of epigenetic silencing constitutes a major target for life history activation of cellular functions for genome change. This likely occurs after genome replication, possibly by changes in small non-coding (snc) RNAs, typically on the order of 20–50 nucleotides long.

MOBILE DNA IS A MAJOR AND FUNCTIONALLY SIGNIFICANT COMPONENT OF GENOMES

One of the major surprises to come from the initial sequencing of the human genome was the high abundance of dispersed mobile repeat elements (Consortium, 2001). Today, we estimate that at least two-thirds of our genomes is composed of mobile DNA (De Koning et al., 2011). The human genome is not exceptional in its high content of mobile DNA (http://shapiro.bsd.uchicago.edu/TableII.1.shtml).

We increasingly recognize that viruses contribute to cell genomes (Kokosar and Kordis, 2013). They provide sequences for non-coding ncRNAs (Frias-Lasserre, 2012), sites for transcriptional control (Peaston et al., 2004; Dunn et al., 2005; Maksakova et al., 2006; Conley et al., 2008), and elements important in epigenetic regulation (Brunmeir et al., 2010; Conley and Jordan, 2012).

Similar transcriptional and epigenetic regulatory contributions are made by mobile genetic elements (http://shapiro.bsd.uchicago.edu/Table5C-1.MobileElementsFoundtobeExaptedascis-RegulatoryControlSitesinAnimals.html) (Youngson et al., 2005; Kinoshita et al., 2007; Suzuki et al., 2007; Fujimoto et al., 2008; Gehring et al., 2009; Pask et al., 2009; Nakayashiki, 2011).

Mobile DNA is a major source of novel coding information. One mechanism is the process known as "exonization," when splice signals are utilized in newly inserted DNA segments (http://shapiro.bsd.uchicago.edu/Origin_of_New_Protein_Domains.html). New coding sequences also form by reverse transcription of processed RNAs and genome insertion of the cDNAs, sometimes producing chimeric fusions with existing exons (http://shapiro.bsd.uchicago.edu/Table 5B. Reports of retrogenes in plant and animal genomes.html) (Long, 2001; Betrán et al., 2002; Fu et al., 2010).

It is now clear that mobile genetic elements play a key role in establishing and rewiring genomic networks (http://shapiro.bsd.uchicago.edu/Table5C-1.MobileElementsFoundtobeExaptedascis-RegulatoryControlSitesinAnimals.html) (Feschotte, 2008; Lindblad-Toh et al., 2011; Lowe et al., 2011; Testori et al., 2012; Kokosar and Kordis, 2013). Moreover, mobile element proliferation is a key factor in the formation of very large genomes (http://shapiro.bsd.uchicago.edu/Genome_Size.html).

The potential functional importance of distributed mobile DNA in genomes grows rapidly as evidence accumulates for pervasive genome transcription (http://shapiro.bsd.uchicago.edu/PervasiveGenomeTranscription.html) and for the regulatory role of non-coding RNAs (ncRNAs) in genome expression, including the functional juxtaposition of distant genome regions to activate transcription (http://shapiro.bsd.uchicago.edu/NonCodingRNAinGenomeExpression.html). Mobile elements participate in this long-range genomic communication and provide the sequences of many ncRNAs (Kapusta et al., 2013).

CELLS USE RNA-TARGETED EPIGENETIC CONTROL TO INHIBIT THE ACTIVITY OF MOBILE DNA

Given the high content of mobile DNA in many genomes, an important question is: what prevents all the mobility systems from destroying genome integrity? In eukaryotic cells, a major control mechanism is sncRNA-directed epigenetic formatting into silent chromatin (Law and Jacobsen, 2010; Castel and Martienssen, 2013).

Both prokaryotes and eukaryotes have systems for capturing fragments from invading DNA molecules and placing the fragments into special loci encoding sncRNAs (Dumesic and Madhani, 2014). In prokaryotes, these loci are called CRISPRs (clustered regular interspersed palindromic repeats) (http://shapiro.bsd.uchicago.edu/CRISPRs.html) (Marraffini and Sontheimer, 2010; Garrett et al., 2011; Bikard and Marraffini, 2013; Watanabe et al., 2013). The RNA transcripts from CRISPRs are processed into sncRNAs that target cleavage of homologous invading DNA and also inactivation of complementary mRNA (Djordjevic et al., 2012). The details of the RNA processing and interference activities are well-characterized, but the acquisition of DNA fragments is poorly understood. The process must be very rapid, because viral infection yields cells that survive the initial infection with appropriate fragments added to their CRISPR repertoire (Barrangou et al., 2007).

Virtually all eukaryotes investigated, with the notable exception of budding yeast, have mechanisms for sncRNA-directed chromatin silencing. They are based on members of the Argonaute family of sncRNA-processing proteins (http://shapiro.bsd.uchicago.edu/microRNA-directedchromatinsilencing.html). Plants and animals have independently evolved distinct mechanisms of processing the sncRNAs for the Argonaute family systems, but both groups use targeted epigenetic regulatory processes to defend against virus infection (Ding and Voinnet, 2007; Csorba et al., 2009) and prevent genome instability (**Table 1**). Like prokaryotes, *Drosophila* has specific genomic loci where it acquires fragments of invading DNA to encode the targeting sncRNAs (Brennecke et al., 2007, 2008; Handler et al., 2013).

LIFE HISTORY EVENTS DESTABILIZE GENOMES AND ACTIVATE MOBILE DNA

Anyone who has studied real-time genome changes quantitatively knows that mutation frequencies depend upon the treatment of the experimental organism prior to measurement. A wide variety of life history events influence the natural genetic engineering (NGE) functions that generate mutations, especially mobile elements (**Table 2**; Shapiro, 2011). In some cases, the genome instabilities are large scale and last multiple cell or organismal generations.

Many observations demonstrate responses of the circuits controlling NGE functions to biological and abiotic inputs. It is particularly significant that many such responses occur following exceptional cell interactions with viruses or with other cells, either by infection or by hybridization (**Table 2**). As we might expect, the introduction of alien DNA or chromatin into a cell often has disruptive effects on genome homeostasis (Shapiro, 2014).

Table 1 | Genome immunity by sncRNA targeting of mobile DNA (see also http://shapiro.bsd.uchicago.edu/Table1l.9.shtml for earlier references).

Organisms	sncRNA targets	References
Plants	Transposable elements	Rigal and Mathieu, 2011; Ng et al., 2012; Nuthikattu et al., 2013
Arabidopsis	Retrotransposons	Mirouze et al., 2009; Slotkin, 2010
Rice	Retrotransposons	Tian et al., 2011
Brassica	Retrotransposons	Zhang et al., 2013
Arabidopsis	Transposable elements	Mccue et al., 2012
Maize	Transposable elements	Barber et al., 2012; He et al., 2013
Plants	Viruses and viroids	Navarro et al., 2009; Pantaleo, 2011; Zhu and Guo, 2012; Ramesh et al., 2014
Rice, tobacco and *Laodelphgax striatellus* insect vector	Rice stripe virus	Xu et al., 2012b
Arabidopsis	Geminiviruses	Vanitharani et al., 2005; Raja et al., 2014
Caenorhabditis elegans germ-line	Transposons	Sijen and Plasterk, 2003; Buckley et al., 2012; Lee et al., 2012
Drosophila	Viruses	Van Rij et al., 2006
Drosophila somatic cells	Retrotransposons	Kawamura et al., 2008
Drosophila male germ-line	Retrotransposons	Kalmykova et al., 2005
Drosophila female germ-line	Transposons, retrotransposons and retroviruses	Brennecke et al., 2007, 2008
Drosophila female germ-line	Telomeric retrotransposons	Shpiz et al., 2009
Drosophila gonads	Transposons	Sienski et al., 2012
Drosophila somatic and germ-line cells	Transposons, retrotransposons and retroviruses	Handler et al., 2013
Drosophila tissue culture cells	Transposons	Chung et al., 2008
Shrimp	White spot syndrome DNA virus	Huang and Zhang, 2013; Sabin and Cherry, 2013
Mammalian cells	EMCV and NoV RNA viruses	Maillard et al., 2013
Human tissue culture cells	LINE retrotransposons	Yang and Kazazian, 2006

EPIGENETIC CHANGES IN RESPONSE TO LIFE HISTORY EVENTS

One of the most active research areas in the second decade of the 21st century is analyzing the impact of life history events on the epigenetic layers of cell regulatory architecture (**Table 3**) (Chinnusamy and Zhu, 2009a; Vandegehuchte and Janssen,

Table 2 | Life history events that lead to genome destabilization (see also http://shapiro.bsd.uchicago.edu/Table11.8.shtml for earlier references).

Organism	Life history event	Genome instability	References
Plant	Polyploidization	Transposon and retrotransposon activation	Bento et al., 2013
Rice	Introgression from wild rice (*Zizania*)	Genome-wide variation of all kinds, including transposon reactivation and transgenerational mobile element activation	Wang et al., 2009, 2010, 2013b
Apple	Polyploidization	Aneuploidy	Considine et al., 2012
Brassica	Intertribal hybridization; genome triplication; allopolyploidization	Retrotransposition; loss of tandem arrays; Homoeologous shuffling and chromosome compensation	Xiong et al., 2011; Fang et al., 2012; Zhang et al., 2013
Wheat, rye	Allopolyploidization	Loss of repetitive and non-coding DNA, including chromosome-specific sequences; rearrangement of syntenic blocks; transposon and retrotransposon activity	Bento et al., 2008, 2010, 2013; Kraitshtein et al., 2010; Yaakov and Kashkush, 2011b, 2012; Feldman and Levy, 2012; Tomas et al., 2012; Luo et al., 2012; Martis et al., 2013
Sunflower	Polyploidization	Chromosome rearrangements	Lim et al., 2008; Chester et al., 2012
Plants	Polyploidization	Rapid genome reshuffling	Tayale and Parisod, 2013
Plants	Polyploidization	Meiotic and fertilization abnormalities	Grandont et al., 2013
Animals	Polyploidization	Meiotic and fertilization abnormalities	Bogart and Bi, 2013; Choleva and Janko, 2013; Stenberg and Saura, 2013
Squalius alburnoides (Cyprinid fish)	Polyploidization	Rapid genome reshuffling; mobile element activity	Collares-Pereira et al., 2013
Arabidopsis	Oilseed rape mosaic virus infection	Increased homologous recombination	Yao et al., 2013
Arabidopsis	Heat shock	Transgenerational ONSEN retrotransposon activation	Matsunaga et al., 2012
Arabidopsis	Volatiles from UV-irradiated *Arabidopsis* or tobacco plants	Increased homologous recombination	Yao et al., 2011
Arabidopsis	Abiotic stresses (ionizing radiation, heavy metals, chlorine, temperature and water)	Somatic and heritable changes in homologous recombination, strand breakage	Boyko et al., 2010; Rahavi et al., 2011; Yao and Kovalchuk, 2011
Tobacco	Tobacco mosaic virus infection	Increased homologous recombination	Kathiria et al., 2010
Rice	Tissue culture cultivation	Genomic DNA fragment length polymorphisms	Wang et al., 2013a
Rice	Etoposide exposure	Increased transposon activity	Yang et al., 2012
Human	Human papillomavirus (HPV) integration	Extensive rearrangements, often focused on insertion site	Korzeniewski et al., 2011; Akagi et al., 2014

2013). The observed epigenetic responses include alterations to cytosine methylation in DNA (Chinnusamy and Zhu, 2009b), histone modifications in nucleosomes, and sncRNAs (Ruiz-Ferrer and Voinnet, 2009; Ng et al., 2012) as well as transgenerational inheritance of complex novel phenotypes (Zucchi et al., 2012), frequently induced by stress (Boyko and Kovalchuk, 2010). The phenomenon of hybrid vigor, or heterosis, increasingly is viewed as an alteration in sncRNA-targeted epigenetic formatting stimulated by the encounter of two distinct genome control regimes (Groszmann et al., 2011; Miller et al., 2012; Shivaprasad et al., 2012).

Many of the studies demonstrating induced epigenetic modifications also document accompanying genome instabilities and emphasize their evolutionary potential (Madlung and Wendel, 2013). It is noteworthy that many of the same stimuli are involved in both genomic and epigenomic responses in plants (Hegarty et al., 2013) and animals (Arkhipova and Rodriguez, 2013). The common stimuli include infection and symbiosis (Hamon and Cossart, 2008; Bierne et al., 2012; Takahashi, 2014), hybridization and changes in ploidy.

DIRECT INTERACTIONS BETWEEN NGE ACTIVITIES AND EPIGENETIC REGULATORY FUNCTIONS

In addition to disruption of sncRNA-targeted inhibition, there is limited but growing evidence that NGE functions acting on DNA molecules interact directly with epigenetic control factors. There is convincing evidence of the connection between NGE and the epigenome in DNA damage repair and retroviral or retrotransposon insertions into chromosomes.

EPIGENETIC INVOLVEMENT IN DNA PROOFREADING AND REPAIR

There are recent reports that a specific histone modification (H3K36me3) primes DNA mismatch repair (Schmidt and Jackson, 2013), that H3K56 acetylation affects mismatch repair (Kadyrova et al., 2013), that hypoacetylation of H3K56 by HDACs 1 and 2 facilitates recruitment of non-homologous end-joining

Table 3 | Life history events that induce epigenetic changes (see also http://shapiro.bsd.uchicago.edu/TableII.10.shtml for earlier references).

Organism	Life history event	Epigenetic change	References
Plants	Hybridization, polyploidization	sncRNA changes	Ng et al., 2012
Maize	Hybridization	rasRNA variation	Barber et al., 2012
Cotton	Allotetraploidization	Changes in mi- and siRNA content and levels	Pang et al., 2009
Brassica napus	Intertribal hybridization and introgression	Changes in cytosine methylation	Zhang et al., 2013
Wheat	Allopolyploidization	Multigenerational transposon methylation changes	Kraitshtein et al., 2010; Yaakov and Kashkush, 2011a,b
Wheat	Hybridization and polyploidization	Deregulation of sncRNAs	Kenan-Eichler et al., 2011
Solanaceae	Interspecific grafting	DNA methylation changes	Wu et al., 2013
Tobacco	Geminivirus and geminivirus-beta satellite infection	Suppression of DNA methylation-base silencing	Vanitharani et al., 2005; Buchmann et al., 2009; Yang et al., 2011
Tobacco	Tobacco mosaic virus infection	Heritable resistance to viral, bacterial and fungal pathogens	Kathiria et al., 2010
Rice	Drought exposure	Multigenerational DNA methylation changes	Zheng et al., 2013
Rice	Nitrogen deprivation	Heritable stress tolerance	Kou et al., 2011
Rice	Tissue culture cultivation	DNA methylation changes	Fukai et al., 2010; Wang et al., 2013a
Rice	Etoposide exposure	Multigenerational DNA methylation changes	Yang et al., 2012
Rice	Salt exposure	DNA methylation changes	Karan et al., 2012
Rice	Heavy metal exposure	Multigenerational DNA methylation changes	Ou et al., 2012
Rice	Abiotic stresses	Novel sncRNAs in the infloresences	Barrera-Figueroa et al., 2012
Pear seeds	Desiccation	DNA methylation changes	Michalak et al., 2013
Arabidopsis	Interspecific hybridization	Polycomb response complex changes	Burkart-Waco et al., 2013
Arabidopsis	Geminivirus (Cabbage leaf curl virus, CaLCuV) infection	Epigenetic silencing	Aregger et al., 2012
Arabidopsis	Stress response	Alteration of *Athila* family retrotransposon sncRNA	Mccue et al., 2012
Arabidopsis	Biotic stresses (bacteria, hormones)	Increased DNA methylation	Dowen et al., 2012
Arabidopsis	β-amino-butyric acid	Imprinted resistance (multigenerational) to *Pseduomonas syringae* and *Hyaloperonospora arabidopsidis* fungus	Slaughter et al., 2012
Arabidopsis	Salt exposure	DNA methylation, nucleosome composition	Bilichak et al., 2012
Arabidopsis	Hyperosmotic priming	Shortening and fractionation of H3K27me3 islands	Sani et al., 2013
Wild rye	Abiotic stresses	DNA methylation	Yu et al., 2013b
Neptune grass	Cadmium	DNA methylation and chromatin patterning	Greco et al., 2012
Plant and mammalian cells	Cadmium	DNA methylation and histone modification	Wang et al., 2012
Nematode (*Caenorhabditis elegans*)	Flock house virus expression	Transgenerational resistance transmitted by sncRNAs	Rechavi et al., 2011
Mosquito (*Aedes aegypti*)	*Wolbachia* infection	Disruption of cytosine methylation	Ye et al., 2013
Carp	Allotetraploidization	Localized hypermethylation	Xiao et al., 2013
Squalius alburnoides (fish)	Polyploidization	Alterations in sncRNA patterns	Inacio et al., 2012
Rats	Exposure to dioxin and endocrine disruptors of F0 generation	Transgenerational inheritance of adult onset diseases and sperm epimutations	Manikkam et al., 2012, 2013
Rats	Vinclozolin fungicide exposure of F0 males	Transgeneration changes to physiology, behavior, metabolic activity, and transcriptome in discrete brain nuclei, altered restraint stress responses	Crews et al., 2012
Pigs	Diet supplementation of F0 with methylating micronutrients	Transgenerational inheritance of extra fat and DNA methylation changes	Braunschweig et al., 2012
Mouse neuronal cells	Short-term hypoxia	DNA methylation changes	Hartley et al., 2013
Humans	High fat diet	DNA methylation changes	Jacobsen et al., 2012
Humans	Early life trauma	DNA methylation changes	Labonte et al., 2012

(Continued)

Table 3 | Continued

Organism	Life History Event	Epigenetic change	References
Humans	Cadmium	DNA hypo-methylation	Hossain et al., 2012
Human lymphocytes	Epstein-Bar virus (EBV) infection	Hypermethylation of tumor suppressor loci, DNA methylation changes	Leonard et al., 2011; Kaneda et al., 2012; Queen et al., 2013
Human liver cells	Hepatitis B virus infection	DNA methylation, histone and sncRNA changes	Tian et al., 2013; Rongrui et al., 2014
Gastric epithelium	*Helicobacter pylori* infection	DNA methylation and histone changes	Ding et al., 2010; Alvarez et al., 2013; Chiariotti et al., 2013
Schwann cells	*Mycobacterium leprae* infection	Reprogramming to stem cell-like state	Masaki et al., 2013

(NHEJ) proteins (Miller et al., 2010; Munoz-Galvan et al., 2013), and that nucleosome remodeling is integral to DS break repair (Seeber et al., 2013). Longstanding observations document the involvement of a specific histone, gamma-H2AX, in DS break repair and NHEJ (Kinner et al., 2008; Altaf et al., 2009; Dickey et al., 2009b; Redon et al., 2009; Firsanov et al., 2011; Chen et al., 2013). A direct role in chromatin remodeling for DNA repair has been claimed for another H2 analog, H2A.Z (Xu et al., 2012a).

Published reports indicate that H2AX incorporation into chromatin suppresses conversion of single-strand nicks to DS breaks (Franco et al., 2006) and affects the processing of the ends of broken DNA molecules (Helmink et al., 2011). H2AX operates in phosphorylated form (Rogakou et al., 1998; Kinner et al., 2008).

Beyond the role of H2AX, chromatin dynamics play an essential role in DNA repair and genome homeostasis (Lahue and Frizzell, 2012; Shi and Oberdoerffer, 2012). Many reports claim repair roles for chromatin regulators, remodeling complexes and nucleosome exchange factors (Ryan and Owen-Hughes, 2011):

– in DNA damage tolerance (Conaway and Conaway, 2009; Falbo et al., 2009);
– after exposure to ionizing radiation (Hunt et al., 2013);
– in UV damage responses (Palomera-Sanchez and Zurita, 2011; Yu et al., 2011);
– in DS break repair by NHEJ and HR (Van Attikum and Gasser, 2005; Williams et al., 2007; Robert et al., 2011; Xu and Price, 2011; Price and D'andrea, 2013);
– in PolyADP-dependent DNA repair (Ahel et al., 2009);
– in NER as well as DS break repair (Osley et al., 2007; Czaja et al., 2012; Yu et al., 2013a).

Nucleosome disassembly is probably necessary for certain repair processes (Linger and Tyler, 2007; Amouroux et al., 2010; Gospodinov and Herceg, 2013), and histone modifications affect damage-induced checkpoint signaling (Chen and Tyler, 2008). Once repair is complete, nucleosome modifications are reversed, and H2AX~P is eliminated from chromatin (Svetlova et al., 2010). So-called "bystander" cells, which are not subjected to DNA damage but are in the same culture as irradiated cells, also display H2AX phosphorylation (Sokolov et al., 2007; Dickey et al., 2009a, 2011).

A key feature of genome repair is that H2AX-marked damaged DNA mobilizes to subnuclear "repair centers" where homologous recombination and NHEJ proteins also localize (Lisby and Rothstein, 2005; Plate et al., 2008; Bekker-Jensen and Mailand, 2010). A role for chromatin in mobilization of damaged DNA has been proposed (Seeber et al., 2013), but multiple sources of evidence are lacking.

RETROVIRAL AND RETROTRANSPOSON INTEGRASES

A more extensive case for NGE-chromatin interactions comes from analysis of retroviral and retrotransposon insertion specificities (Zhang and Mager, 2012). Each type of retrovirus displays a characteristic insertion specificity for its provirus (Lewinski et al., 2006). A number of targeting mechanisms involve epigenetic formatting molecules.

In budding yeast, Ty1 retrotransposon integrase contacts an H2A/H2B interface upstream of RNA polymerase III initiation sites (Baller et al., 2012; Bridier-Nahmias and Lesage, 2012; Mularoni et al., 2012). Histone deacetylase Hos2 and Trithorax group protein Set3 stimulate this nucleosome-targeted integration (Mou et al., 2006), and chromatin remodeling factor Isw2p is also implicated (Bachman et al., 2005). In contrast, the Ty5 retrotransposon inserts in silent chromatin, targeted by binding of its integrase to the Sir4 heterochromatin nucleating factor (Xie et al., 2001; Dai et al., 2007; Brady et al., 2008; Baller et al., 2011).

HIV and other lentiviral targeted integration into actively transcribed regions of the genome is associated with transcription-associated histone modifications, including H3 acetylation, H4 acetylation, and H3 K4 methylation, but is disfavored in regions rich in transcription-inhibiting modifications, which include H3K27me3 and DNA CpG methylation (Wang et al., 2007). The specificity results from integrase tethering by the LEDGF/p75 chromatin-binding growth factor (Vanegas et al., 2005; Llano et al., 2006; Ciuffi, 2008; Meehan and Poeschla, 2010; Zheng et al., 2010; Christ and Debyser, 2013). Replacing the LEDGF/p75 domain that interacts with expressed chromatin by the CBX1 domain, which binds histones H3K9me2 or H3K9me3 found in pericentric heterochromatin, targets HIV insertions to silent chromatin regions (Gijsbers et al., 2010).

Murine leukemia virus (MuLV) insertion targeting to initiation sites upstream of actively transcribed regions involves integrase interactions with bromodomain proteins BRD2, BRD3, and BRD4 (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013a). Interestingly, chromatin recognition bromodomain

protein BRD4 antagonizes HIV provirus reactivation (Zhu et al., 2012).

Certain retrotransposons are specifically targeted to centromeres (Wolfgruber et al., 2009; Birchler and Presting, 2012; Tsukahara et al., 2012; Sharma et al., 2013b), which have a special chromatin configuration characterized by centromeric versions of H3 (Henikoff and Dalal, 2005; Vos et al., 2006; Partridge, 2008; Zhang et al., 2008a). Centromeric retrotransposons in rice are highly associated with H3K9me2, a hallmark for heterochromatin (Neumann et al., 2007). Some centromeric retrotransposons encode integrase proteins with histone-binding chromodomains at their carboxy-termini (Neumann et al., 2011). Chromodomains recognize lysine methylation (Blus et al., 2011; Yap and Zhou, 2011; Eissenberg, 2012).

It is probably not coincidental that the most widely distributed group of retrotransposons among all eukaryotic clades are the "chromoviruses," which are so named because they have chromodomains in their integrase proteins (Gorinsek et al., 2004; Kordis, 2005; Novikov et al., 2012; Weber et al., 2013). A chromodomain has been reported to target fungal chromovirus MAGGY insertions to heterochromatin marked by H3K9me2/me3 (Gao et al., 2008). An integrase chromodomain also participates in activator protein-targeted insertion of fission yeast retrotransposon Tf1 upstream of RNA polymerase II transcription start sites (Hizi and Levin, 2005; Chatterjee et al., 2009).

DNA TRANSPOSONS

In contrast with many retrotransposons that interact with nucleosomes, the DNA transposon Hermes inserts preferentially in budding yeast into nucleosome-free regions of the genome (Gangadharan et al., 2010). The widely used P element DNA transposons in *Drosophila* show targeting (called "P element homing") by incorporating binding sites for various regulatory factors, including chromatin insulators (Bender and Hudson, 2000; Fujioka et al., 2009) and Polycomb group response elements (Kassis, 2002; Cheng et al., 2012).

EPIGENETIC REFORMATTING AFTER DNA REPLICATION AND ncRNAs AS POTENTIAL AGENTS FOR TRANSMITTING EXPERIENCE TO THE GENOME

While the evidence is increasingly abundant for effects of different life history events on epigenetic regulation in general, and on genome homeostasis in particular, it is far from clear how those effects occur (Lim and Brunet, 2013). We know very little about the connections between cell sensors and epigenetic (re)formatting complexes (Erdel et al., 2011; Narlikar et al., 2013). Deciphering those connections is currently an important research goal.

DNA replication provides a key decision point for maintaining or changing chromatin configurations (Poot et al., 2005; Liu and Gong, 2011; Mermoud et al., 2011). The replication apparatus must disassemble chromatin for polymerization and then reassemble chromatin once replication is complete. Replication takes place only in dividing cells, and transgenerational inheritance of epigenetic states must involve the proliferating cells that give rise to gametes. Transfer of outside information from somatic tissues to the germline has been reported in mammals (Sharma, 2013; Skinner et al., 2013). And epigenetic windows of susceptibility to environmental insults have been suggested during sperm development (Soubry et al., 2014). Since there is no segregated germ line in plants and eukaryotic microbes, the same cells that experience environmental inputs can also be the progenitors of gametes.

A number of different factors have been found or hypothesized to participate in post-replication chromatin restoration: histone chaperones (Budhavarapu et al., 2013), RNA editing and sncRNAs (Savva et al., 2013), chromatin remodeler SMARCAD1 (Mermoud et al., 2011), chromatin assembly factor 1 (Huang and Jiao, 2012), histone chaperon FACT (Winkler and Luger, 2011) and Swi/Snf complexes (Neves-Costa and Varga-Weisz, 2006; Ryan and Owen-Hughes, 2011; Zhu et al., 2013), and ISW1 complexes (Erdel and Rippe, 2011).

One frequently overlooked feature of post-replication reestablishment of epigenetic formatting is where in the nucleus it might occur. Replication takes place in specialized "replication factories" (Vago et al., 2009; Guillou et al., 2010). Does chromatin reestablishment occur in the same location or does it involve migration of newly replicated DNA segments to distinct subnuclear "chromatin factories," like the ones that exist in the nucleolus for heterochromatin formation on rRNA-encoding DNA (Guetg and Santoro, 2012)? If so, such post-replication relocalization would be guided by the nucleoskeleton and lncRNAs (Mercer and Mattick, 2013; Mercer et al., 2013) and might present an attractive target for stress response and sensory input signaling (Weiner et al., 2012).

It is notable that changes to ncRNAs are frequently cited with regard to the impact of life history events on the genome (Sunkar et al., 2007; Khraiwesh et al., 2012; Lelandais-Briere et al., 2012; Nakaminami et al., 2012; Amaral et al., 2013). In the plant literature, there is documentation of numerous ncRNA changes in response to particular biotic and abiotic stress regimes (**Table 4**).

A number of observations about resistance to biotic and abiotic stresses are consistent with a key role for ncRNA changes in life history responses. Several viruses encode siRNA suppressors to overcome host defenses (Jiang et al., 2012; Omarov and Scholthof, 2012; Guo and Lu, 2013). Transgenic constructs encoding constitutive miRNA expression can lead to salt and drought tolerance in creeping bentgrass (Zhou et al., 2013), to immunity against blast fungus in rice (Li et al., 2014), and in *Arabidopsis* to greater salt and alkalinity sensitivity (Gao et al., 2011). Acquired aphid resistance in *Arabidopsis* involves sncRNA changes (Kettles et al., 2013), and most acquired stress resistances in plants display transgenerational epigenetic inheritance (Holeski et al., 2012; Luna and Ton, 2012; Slaughter et al., 2012).

SPECULATIVE CONCLUSIONS ABOUT AN EPIGENETIC INTERFACE BETWEEN EXPERIENCE AND GENOME CHANGE

Mobile DNA and other NGE functions are the key agents for adaptively significant changes in genome organization and DNA sequences. The data reviewed and tabulated above establish the importance of RNA-directed chromatin formatting in the regulation and operation of mobile elements, viruses and DNA repair

Table 4 | Changes in non-coding RNAs in response to life history events.

Stress or input	Organism(s)	References
Salt	Multiple plants	Ding et al., 2009; Qin et al., 2011; Macovei and Tuteja, 2012; Carnavale Bottino et al., 2013; Li et al., 2013; Ren et al., 2013; Zhuang et al., 2014
Drought	Multiple plants	Barrera-Figueroa et al., 2011; Li et al., 2011a; Qin et al., 2011; Wang et al., 2011; Eldem et al., 2012; Ferreira et al., 2012; Ding et al., 2013; Gentile et al., 2013; Shuai et al., 2013
Waterlogging	Maize, poplar	Zhang et al., 2008b; Liu et al., 2012; Ren et al., 2012; Zhai et al., 2013
Cold stress	Wheat	Tang et al., 2012
Aluminum	Soybeans	Chen et al., 2012a; Zeng et al., 2012
Cadmium	Radish	Xu et al., 2013
Boron	Barley	Ozhuner et al., 2013
ethylene	*Medicago truncatula* (barrel clover)	Chen et al., 2012b
Ozone	*Arabidopsis*	Iyer et al., 2012
Hypoxia	*Arabidopsis*	Moldovan et al., 2010
Low phosphorous	Maize	Zhang et al., 2012
Low nitrate	Maize	Xu et al., 2011
Sulfur deprivation	*Chlamydomonas reinhardtii*	Shu and Hu, 2012
Abiotic stresses	Multiple plants	Kulcheski et al., 2011; Li et al., 2011b; Barrera-Figueroa et al., 2012; Sun et al., 2012; Zhan et al., 2012; Ballen-Taborda et al., 2013
Physiological stressors and invasive plant infection	Rice blast fungus, *Magnaporthe oryzae*	Raman et al., 2013
Virus infection	Multiple plants, rice	Du et al., 2011; Sha et al., 2014
Viral and bacterial infections	Multiple plants, cassava (*Xanthomonas* infection)	Perez-Quintero et al., 2012; Zvereva and Pooggin, 2012; Pelaez and Sanchez, 2013; Quintero et al., 2013
Bacterial/phytoplasma infection	Multiple plants, lime trees	Zhang et al., 2011; Ehya et al., 2013
Powdery mildew infection	Wheat	(Xin et al., 2011) miRNAs (Xin et al., 2010)
Verticillium wilt infection	Cotton, eggplant	Yin et al., 2012; Yang et al., 2013

functions. In addition, there is a remarkable correlation between the life history events that activate NGE functions to destabilize genomes and those that lead to alteration of chromatin states and DNA methylation patterns.

The preceding observations lead to the plausible hypothesis that epigenetic regulation serves as a key interface between organismal life history and the agents that restructure genomic DNA. This hypothesis is supported by the limited number of cases where empirical observations have established direct molecular connections between NGE functions and components of the epigenetic control system: histones, nucleosomes, and chromatin reformatting complexes.

If, as I expect, further research bolsters the epigenome-NGE correlations and connections documented above, then we need to ask: what components(s) of the epigenetic control apparatus communicate information about experience to NGE operators? We do not know the answer to this fundamental question. However, the data reported in **Table 4** indicate that ncRNAs are good candidates for key intermediates in the experience-genome signal transduction process. If this is so, then ncRNAs are logical molecular targets for modulating genome change toward potentially adaptive outcomes. Let us hope that research aimed at examining this proposal deepens our understanding of how life history impacts both epigenetic and genome change operations (**Tables 2–4**), whether or not my speculation ultimately proves to be correct.

REFERENCES

Ahel, D., Horejsi, Z., Wiechens, N., Polo, S. E., Garcia-Wilson, E., Ahel, I., et al. (2009). Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* 325, 1240–1243. doi: 10.1126/science.1177321

Akagi, K., Li, J., Broutian, T. R., Padilla-Nash, H., Xiao, W., Jiang, B., et al. (2014). Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* 24, 185–199. doi: 10.1101/gr.164806.113

Altaf, M., Auger, A., Covic, M., and Cote, J. (2009). Connection between histone H2A variants and chromatin remodeling complexes. *Biochem. Cell Biol.* 87, 35–50. doi: 10.1139/O08-140

Alvarez, M. C., Santos, J. C., Maniezzo, N., Ladeira, M. S., Da Silva, A. L., Scaletsky, I. C., et al. (2013). MGMT and MLH1 methylation in Helicobacter pylori-infected children and adults. *World J. Gastroenterol.* 19, 3043–3051. doi: 10.3748/wjg.v19.i20.3043

Amaral, P. P., Dinger, M. E., and Mattick, J. S. (2013). Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Brief. Funct. Genomics* 12, 254–278. doi: 10.1093/bfgp/elt016

Amouroux, R., Campalans, A., Epe, B., and Radicella, J. P. (2010). Oxidative stress triggers the preferential assembly of base excision repair complexes on open chromatin regions. *Nucleic Acids Res.* 38, 2878–2890. doi: 10.1093/nar/gkp1247

Aregger, M., Borah, B. K., Seguin, J., Rajeswaran, R., Gubaeva, E. G., Zvereva, A. S., et al. (2012). Primary and secondary siRNAs in geminivirus-induced gene silencing. *PLoS Pathog.* 8:e1002941. doi: 10.1371/journal.ppat.1002941

Arkhipova, I. R., and Rodriguez, F. (2013). Genetic and epigenetic changes involving (retro)transposons in animal hybrids and polyploids. *Cytogenet. Genome Res.* 140, 295–311. doi: 10.1159/000352069

Bachman, N., Gelbart, M. E., Tsukiyama, T., and Boeke, J. D. (2005). TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to S. cerevisiae tDNAs. *Genes Dev.* 19, 955–964. doi: 10.1101/gad.1299105

Ballen-Taborda, C., Plata, G., Ayling, S., Rodriguez-Zapata, F., Becerra Lopez-Lavalle, L. A., Duitama, J., et al. (2013). Identification of cassava MicroRNAs under abiotic stress. *Int. J. Genomics* 2013, 857986. doi: 10.1155/2013/857986

Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J., and Voytas, D. F. (2012). A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. *Genome Res.* 22, 704–713. doi: 10.1101/gr.129585.111

Baller, J. A., Gao, J., and Voytas, D. F. (2011). Access to DNA establishes a secondary target site bias for the yeast retrotransposon Ty5. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20351–20356. doi: 10.1073/pnas.1103665108

Barber, W. T., Zhang, W., Win, H., Varala, K. K., Dorweiler, J. E., Hudson, M. E., et al. (2012). Repeat associated small RNAs vary among parents and following hybridization in maize. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10444–10449. doi: 10.1073/pnas.1202073109

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140

Barrera-Figueroa, B. E., Gao, L., Diop, N. N., Wu, Z., Ehlers, J. D., Roberts, P. A., et al. (2011). Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. *BMC Plant Biol.* 11:127. doi: 10.1186/1471-2229-11-127

Barrera-Figueroa, B. E., Gao, L., Wu, Z., Zhou, X., Zhu, J., Jin, H., et al. (2012). High throughput sequencing reveals novel and abiotic stress-regulated microRNAs in the inflorescences of rice. *BMC Plant Biol.* 12:132. doi: 10.1186/1471-2229-12-132

Bekker-Jensen, S., and Mailand, N. (2010). Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair (Amst.)* 9, 1219–1228. doi: 10.1016/j.dnarep.2010.09.010

Bender, W., and Hudson, A. (2000). P element homing to the Drosophila bithorax complex. *Development* 127, 3981–3992. Available online at: http://dev.biologists.org/content/127/18/3981.long

Bento, M., Gustafson, P., Viegas, W., and Silva, M. (2010). Genome merger: from sequence rearrangements in triticale to their elimination in wheat-rye addition lines. *Theor. Appl. Genet.* 121, 489–497. doi: 10.1007/s00122-010-1325-6

Bento, M., Pereira, H. S., Rocheta, M., Gustafson, P., Viegas, W., and Silva, M. (2008). Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in triticale. *PLoS ONE* 3:e1402. doi: 10.1371/journal.pone.0001402

Bento, M., Tomas, D., Viegas, W., and Silva, M. (2013). Retrotransposons represent the most labile fraction for genomic rearrangements in polyploid plant species. *Cytogenet. Genome Res.* 140, 286–294. doi: 10.1159/000353308

Betrán, E., Thornton, K., and Long, M. (2002). Retroposed new genes out of the X in Drosophila. *Genome Res.* 12, 1854–1859. doi: 10.1101/gr.6049

Bierne, H., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2:a010272. doi: 10.1101/cshperspect.a010272

Bikard, D., and Marraffini, L. A. (2013). Control of gene expression by CRISPR-Cas systems. *F1000Prime Rep.* 5:47. doi: 10.12703/p5-47

Bilichak, A., Ilnystkyy, Y., Hollunder, J., and Kovalchuk, I. (2012). The progeny of *Arabidopsis thaliana* plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. *PLoS ONE* 7:e30515. doi: 10.1371/journal.pone.0030515

Birchler, J. A., and Presting, G. G. (2012). Retrotransposon insertion targeting: a mechanism for homogenization of centromere sequences on nonhomologous chromosomes. *Genes Dev.* 26, 638–640. doi: 10.1101/gad.191049.112

Blus, B. J., Wiggins, K., and Khorasanizadeh, S. (2011). Epigenetic virtues of chromodomains. *Crit. Rev. Biochem. Mol. Biol.* 46, 507–526. doi: 10.3109/10409238.2011.619164

Bogart, J. P., and Bi, K. (2013). Genetic and genomic interactions of animals with different ploidy levels. *Cytogenet. Genome Res.* 140, 117–136. doi: 10.1159/000351593

Boyko, A., Golubov, A., Bilichak, A., and Kovalchuk, I. (2010). Chlorine ions but not sodium ions alter genome stability of Arabidopsis thaliana. *Plant Cell Physiol.* 51, 1066–1078. doi: 10.1093/pcp/pcq048

Boyko, A., and Kovalchuk, I. (2010). Transgenerational response to stress in *Arabidopsis thaliana. Plant Signal. Behav.* 5, 995–998. doi: 10.1371/journal.pone.0009514

Brady, T. L., Schmidt, C. L., and Voytas, D. F. (2008). Targeting integration of the Saccharomyces Ty5 retrotransposon. *Methods Mol. Biol.* 435, 153–163. doi: 10.1007/978-1-59745-232-8_11

Braunschweig, M., Jagannathan, V., Gutzwiller, A., and Bee, G. (2012). Investigations on transgenerational epigenetic response down the male line in F2 pigs. *PLoS ONE* 7:e30583. doi: 10.1371/journal.pone.0030583

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128, 1089–1103. doi: 10.1016/j.cell.2007.01.043

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., and Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392. doi: 10.1126/science.1165171

Bridier-Nahmias, A., and Lesage, P. (2012). Two large-scale analyses of Ty1 LTR-retrotransposon de novo insertion events indicate that Ty1 targets nucleosomal DNA near the H2A/H2B interface. *Mob. DNA* 3:22. doi: 10.1186/1759-8753-3-22

Brunmeir, R., Lagger, S., Simboeck, E., Sawicka, A., Egger, G., Hagelkruys, A., et al. (2010). Epigenetic regulation of a murine retrotransposon by a dual histone modification mark. *PLoS Genet.* 6:e1000927. doi: 10.1371/journal.pgen.1000927

Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G., and Bisaro, D. M. (2009). Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.* 83, 5005–5013. doi: 10.1128/jvi.01771-08

Buckley, B. A., Burkhart, K. B., Gu, S. G., Spracklin, G., Kershner, A., Fritz, H., et al. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* 489, 447–451. doi: 10.1038/nature11352

Budhavarapu, V. N., Chavez, M., and Tyler, J. K. (2013). How is epigenetic information maintained through DNA replication? *Epigenetics Chromatin* 6:32. doi: 10.1186/1756-8935-6-32

Burkart-Waco, D., Ngo, K., Dilkes, B., Josefsson, C., and Comai, L. (2013). Early disruption of maternal-zygotic interaction and activation of defense-like responses in Arabidopsis interspecific crosses. *Plant Cell* 25, 2037–2055. doi: 10.1105/tpc.112.108258

Carnavale Bottino, M., Rosario, S., Grativol, C., Thiebaut, F., Rojas, C. A., Farrinelli, L., et al. (2013). High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane. *PLoS ONE* 8:e59423. doi: 10.1371/journal.pone.0059423

Castel, S. E., and Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14, 100–112. doi: 10.1038/nrg3355

Chatterjee, A. G., Leem, Y. E., Kelly, F. D., and Levin, H. L. (2009). The chromodomain of Tf1 integrase promotes binding to cDNA and mediates target site selection. *J. Virol.* 83, 2675–2685. doi: 10.1128/JVI.01588-08

Chen, C. C., and Tyler, J. (2008). Chromatin reassembly signals the end of DNA repair. *Cell Cycle* 7, 3792–3797. doi: 10.4161/cc.7.24.7188

Chen, L., Wang, T., Zhao, M., Tian, Q., and Zhang, W. H. (2012a). Identification of aluminum-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *Planta* 235, 375–386. doi: 10.1007/s00425-011-1514-9

Chen, L., Wang, T., Zhao, M., and Zhang, W. (2012b). Ethylene-responsive miRNAs in roots of *Medicago truncatula* identified by high-throughput sequencing at whole genome level. *Plant Sci.* 184, 14–19. doi: 10.1016/j.plantsci.2011.11.007

Chen, W. T., Alpert, A., Leiter, C., Gong, F., Jackson, S. P., and Miller, K. M. (2013). Systematic identification of functional residues in mammalian histone H2AX. *Mol. Cell. Biol.* 33, 111–126. doi: 10.1128/mcb.01024-12

Cheng, Y., Kwon, D. Y., Arai, A. L., Mucci, D., and Kassis, J. A. (2012). P-element homing is facilitated by engrailed polycomb-group response elements in *Drosophila melanogaster. PLoS ONE* 7:e30437. doi: 10.1371/journal.pone.0030437

Chester, M., Gallagher, J. P., Symonds, V. V., Cruz Da Silva, A. V., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. U.S.A.* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Chiariotti, L., Angrisano, T., Keller, S., Florio, E., Affinito, O., Pallante, P., et al. (2013). Epigenetic modifications induced by *Helicobacter pylori* infection through a direct microbe-gastric epithelial cells cross-talk. *Med. Microbiol. Immunol.* 202, 327–337. doi: 10.1007/s00430-013-0301-6

Chinnusamy, V., and Zhu, J. K. (2009a). Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139. doi: 10.1016/j.pbi.2008.12.006

Chinnusamy, V., and Zhu, J. K. (2009b). RNA-directed DNA methylation and demethylation in plants. *Sci. China C Life Sci.* 52, 331–343. doi: 10.1007/s11427-009-0052-1

Choleva, L., and Janko, K. (2013). Rise and persistence of animal polyploidy: evolutionary constraints and potential. *Cytogenet. Genome Res.* 140, 151–170. doi: 10.1159/000353464

Christ, F., and Debyser, Z. (2013). The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy. *Virology* 435, 102–109. doi: 10.1016/j.virol.2012.09.033

Chung, W. J., Okamura, K., Martin, R., and Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr. Biol.* 18, 795–802. doi: 10.1016/j.cub.2008.05.006

Ciuffi, A. (2008). Mechanisms governing lentivirus integration site selection. *Curr. Gene Ther.* 8, 419–429. doi: 10.2174/156652308786848021

Collares-Pereira, M. J., Matos, I., Morgado-Santos, M., and Coelho, M. M. (2013). Natural pathways towards polyploidy in animals: the *Squalius alburnoides* fish complex as a model system to study genome size and genome reorganization in polyploids. *Cytogenet. Genome Res.* 140, 97–116. doi: 10.1159/000351729

Conaway, R. C., and Conaway, J. W. (2009). The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.* 34, 71–77. doi: 10.1016/j.tibs.2008.10.010

Conley, A. B., and Jordan, I. K. (2012). "Endogenous retroviruses and the epigenome," in *Viruses: Essential Agents of Life*, ed G. Witzany (Dordrecht: Spriinger), 309–323.

Conley, A. B., Piriyapongsa, J., and Jordan, I. K. (2008). Retroviral promoters in the human genome. *Bioinformatics* 24, 1563–1567. doi: 10.1093/bioinformatics/btn243

Considine, M. J., Wan, Y., D'antuono, M. F., Zhou, Q., Han, M., Gao, H., et al. (2012). Molecular genetic features of polyploidization and aneuploidization reveal unique patterns for genome duplication in diploid Malus. *PLoS ONE* 7:e29449. doi: 10.1371/journal.pone.0029449

Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

Crews, D., Gillette, R., Scarpino, S. V., Manikkam, M., Savenkova, M. I., and Skinner, M. K. (2012). Epigenetic transgenerational inheritance of altered stress responses. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9143–9148. doi: 10.1073/pnas.1118514109

Csorba, T., Pantaleo, V., and Burgyan, J. (2009). RNA silencing: an antiviral mechanism. *Adv. Virus Res.* 75, 35–71. doi: 10.1016/s0065-3527(09)07502-2

Czaja, W., Mao, P., and Smerdon, M. J. (2012). The emerging roles of ATP-dependent chromatin remodeling enzymes in nucleotide excision repair. *Int. J. Mol. Sci.* 13, 11954–11973. doi: 10.3390/ijms130911954

Dai, J., Xie, W., Brady, T. L., Gao, J., and Voytas, D. F. (2007). Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin. *Mol. Cell* 27, 289–299. doi: 10.1016/j.molcel.2007.06.010

De Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384

De Rijck, J., De Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N., et al. (2013). The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* 5, 886–894. doi: 10.1016/j.celrep.2013.09.040

Dickey, J. S., Baird, B. J., Redon, C. E., Sokolov, M. V., Sedelnikova, O. A., and Bonner, W. M. (2009a). Intercellular communication of cellular stress monitored by gamma-H2AX induction. *Carcinogenesis* 30, 1686–1695. doi: 10.1093/carcin/bgp192

Dickey, J. S., Redon, C. E., Nakamura, A. J., Baird, B. J., Sedelnikova, O. A., and Bonner, W. M. (2009b). H2AX: functional roles and potential applications. *Chromosoma* 118, 683–692. doi: 10.1007/s00412-009-0234-4

Dickey, J. S., Zemp, F. J., Altamirano, A., Sedelnikova, O. A., Bonner, W. M., and Kovalchuk, O. (2011). H2AX phosphorylation in response to DNA double-strand break formation during bystander signalling: effect of microRNA knock-down. *Radiat. Prot. Dosimetry* 143, 264–269. doi: 10.1093/rpd/ncq470

Ding, D., Zhang, L., Wang, H., Liu, Z., Zhang, Z., and Zheng, Y. (2009). Differential expression of miRNAs in response to salt stress in maize roots. *Ann. Bot.* 103, 29–38. doi: 10.1093/aob/mcn205

Ding, S. W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. *Cell* 130, 413–426. doi: 10.1016/j.cell.2007.07.039

Ding, S. Z., Goldberg, J. B., and Hatakeyama, M. (2010). Helicobacter pylori infection, oncogenic pathways and epigenetic mechanisms in gastric carcinogenesis. *Future Oncol.* 6, 851–862. doi: 10.2217/fon.10.37

Ding, Y., Tao, Y., and Zhu, C. (2013). Emerging roles of microRNAs in the mediation of drought stress response in plants. *J. Exp. Bot.* 64, 3077–3086. doi: 10.1093/jxb/ert164

Djordjevic, M., Djordjevic, M., and Severinov, K. (2012). CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs. *Biol. Direct* 7:24. doi: 10.1186/1745-6150-7-24

Dowen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Dowen, J. M., Nery, J. R., et al. (2012). Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2183–E2191. doi: 10.1073/pnas.1209329109

Du, P., Wu, J., Zhang, J., Zhao, S., Zheng, H., Gao, G., et al. (2011). Viral infection induces expression of novel phased microRNAs from conserved cellular microRNA precursors. *PLoS Pathog.* 7:e1002176. doi: 10.1371/journal.ppat.1002176

Dumesic, P. A., and Madhani, H. D. (2014). Recognizing the enemy within: licensing RNA-guided genome defense. *Trends Biochem. Sci.* 39, 25–34. doi: 10.1016/j.tibs.2013.10.003

Dunn, C. A., Van De Lagemaat, L. N., Baillie, G. J., and Mager, D. L. (2005). Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5. *Gene* 364, 2–12. doi: 10.1016/j.gene.2005.05.045

Ehya, F., Monavarfeshani, A., Mohseni Fard, E., Karimi Farsad, L., Khayam Nekouei, M., Mardi, M., et al. (2013). Phytoplasma-responsive microRNAs modulate hormonal, nutritional, and stress signalling pathways in Mexican lime trees. *PLoS ONE* 8:e66372. doi: 10.1371/journal.pone.0066372

Eissenberg, J. C. (2012). Structural biology of the chromodomain: form and function. *Gene* 496, 69–78. doi: 10.1016/j.gene.2012.01.003

Eldem, V., Celikkol Akcay, U., Ozhuner, E., Bakir, Y., Uranbey, S., and Unver, T. (2012). Genome-wide identification of miRNAs responsive to drought in peach (Prunus persica) by high-throughput deep sequencing. *PLoS ONE* 7:e50298. doi: 10.1371/journal.pone.0050298

Erdel, F., Krug, J., Langst, G., and Rippe, K. (2011). Targeting chromatin remodelers: signals and search mechanisms. *Biochim. Biophys. Acta* 1809, 497–508. doi: 10.1016/j.bbagrm.2011.06.005

Erdel, F., and Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes–where, when and why? *FEBS J.* 278, 3608–3618. doi: 10.1111/j.1742-4658.2011.08282.x

Falbo, K. B., Alabert, C., Katou, Y., Wu, S., Han, J., Wehr, T., et al. (2009). Involvement of a chromatin remodeling complex in damage tolerance during DNA replication. *Nat. Struct. Mol. Biol.* 16, 1167–1172. doi: 10.1038/nsmb.1686

Fang, L., Cheng, F., Wu, J., and Wang, X. (2012). The impact of genome triplication on tandem gene evolution in *Brassica rapa. Front. Plant Sci.* 3:261. doi: 10.3389/fpls.2012.00261

Feldman, M., and Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. *Genetics* 192, 763–774. doi: 10.1534/genetics.112.146316

Ferreira, T. H., Gentile, A., Vilela, R. D., Costa, G. G., Dias, L. I., Endres, L., et al. (2012). microRNAs associated with drought response in the bioenergy crop sugarcane (Saccharum spp.). *PLoS ONE* 7:e46703. doi: 10.1371/journal.pone.0046703

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/nrg2337

Firsanov, D. V., Solovjeva, L. V., and Svetlova, M. P. (2011). H2AX phosphorylation at the sites of DNA double-strand breaks in cultivated mammalian cells and tissues. *Clin. Epigenetics* 2, 283–297. doi: 10.1007/s13148-011-0044-4

Franco, S., Gostissa, M., Zha, S., Lombard, D. B., Murphy, M. M., Zarrin, A. A., et al. (2006). H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* 21, 201–214. doi: 10.1016/j.molcel.2006.01.005

Frias-Lasserre, D. (2012). Non coding RNAs and viruses in the framework of the phylogeny of the genes, epigenesis and heredity. *Int. J. Mol. Sci.* 13, 477–490. doi: 10.3390/ijms13010477

Fu, B., Chen, M., Zou, M., Long, M., and He, S. (2010). The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genomics* 11:657. doi: 10.1186/1471-2164-11-657

Fujimoto, R., Kinoshita, Y., Kawabe, A., Kinoshita, T., Takashima, K., Nordborg, M., et al. (2008). Evolution and control of imprinted FWA genes in the genus Arabidopsis. *PLoS Genet.* 4:e1000048. doi: 10.1371/journal.pgen.1000048

Fujioka, M., Wu, X., and Jaynes, J. B. (2009). A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development* 136, 3077–3087. doi: 10.1242/dev.036467

Fukai, E., Umehara, Y., Sato, S., Endo, M., Kouchi, H., Hayashi, M., et al. (2010). Derepression of the plant Chromovirus LORE1 induces germline transposition in regenerated plants. *PLoS Genet.* 6:e1000868. doi: 10.1371/journal.pgen.1000868

Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J., and Craig, N. L. (2010). DNA transposon Hermes inserts into DNA in nucleosome-free regions *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21966–21972. doi: 10.1073/pnas.1016382107

Gao, P., Bai, X., Yang, L., Lv, D., Pan, X., Li, Y., et al. (2011). osa-MIR393: a salinity- and alkaline stress-related microRNA gene. *Mol. Biol. Rep.* 38, 237–242. doi: 10.1007/s11033-010-0100-8

Gao, X., Hou, Y., Ebina, H., Levin, H. L., and Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* 18, 359–369. doi: 10.1101/gr.7146408

Garrett, R. A., Vestergaard, G., and Shah, S. A. (2011). Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.* 19, 549–556. doi: 10.1016/j.tim.2011.08.002

Gehring, M., Bubb, K. L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324, 1447–1451. doi: 10.1126/science.1171609

Gentile, A., Ferreira, T. H., Mattos, R. S., Dias, L. I., Hoshino, A. A., Carneiro, M. S., et al. (2013). Effects of drought on the microtranscriptome of field-grown sugarcane plants. *Planta* 237, 783–798. doi: 10.1007/s00425-012-1795-7

Gijsbers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., Mcneely, M., et al. (2010). LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. *Mol. Ther.* 18, 552–560. doi: 10.1038/mt.2010.36

Gorinsek, B., Gubensek, F., and Kordis, D. (2004). Evolutionary genomics of chromoviruses in eukaryotes. *Mol. Biol. Evol.* 21, 781–798. doi: 10.1093/molbev/msh057

Gospodinov, A., and Herceg, Z. (2013). Shaping chromatin for repair. *Mutat. Res.* 752, 45–60. doi: 10.1016/j.mrrev.2012.10.001

Grandont, L., Jenczewski, E., and Lloyd, A. (2013). Meiosis and its deviations in polyploid plants. *Cytogenet. Genome Res.* 140, 171–184. doi: 10.1159/000351730

Greco, M., Chiappetta, A., Bruno, L., and Bitonti, M. B. (2012). In *Posidonia oceanica* cadmium induces changes in DNA methylation and chromatin patterning. *J. Exp. Bot.* 63, 695–709. doi: 10.1093/jxb/err313

Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J., and Dennis, E. S. (2011). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2617–2622. doi: 10.1073/pnas.1019217108

Guetg, C., and Santoro, R. (2012). Formation of nuclear heterochromatin: the nucleolar point of view. *Epigenetics* 7, 811–814. doi: 10.4161/epi.21072

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., et al. (2010). Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* 24, 2812–2822. doi: 10.1101/gad.608210

Guo, X., and Lu, R. (2013). Characterization of virus-encoded RNA interference suppressors in *Caenorhabditis elegans*. *J. Virol.* 87, 5414–5423. doi: 10.1128/jvi.00148-13

Gupta, S. S., Maetzig, T., Maertens, G. N., Sharif, A., Rothe, M., Weidner-Glunde, M., et al. (2013). Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* 87, 12721–12736. doi: 10.1128/jvi.01942-13

Hamon, M. A., and Cossart, P. (2008). Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe* 4, 100–109. doi: 10.1016/j.chom.2008.07.009

Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., et al. (2013). The genetic makeup of the Drosophila piRNA pathway. *Mol. Cell* 50, 762–777. doi: 10.1016/j.molcel.2013.04.031

Hartley, I., Elkhoury, F. F., Heon Shin, J., Xie, B., Gu, X., Gao, Y., et al. (2013). Long-lasting changes in DNA methylation following short-term hypoxic exposure in primary hippocampal neuronal cultures. *PLoS ONE* 8:e77859. doi: 10.1371/journal.pone.0077859

He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., et al. (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol.* 14:R57. doi: 10.1186/gb-2013-14-6-r57

Hegarty, M., Coate, J., Sherman-Broyles, S., Abbott, R., Hiscock, S., and Doyle, J. (2013). Lessons from natural and artificial polyploids in higher plants. *Cytogenet. Genome Res.* 140, 204–225. doi: 10.1159/000353361

Helmink, B. A., Tubbs, A. T., Dorsett, Y., Bednarski, J. J., Walker, L. M., Feng, Z., et al. (2011). H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes. *Nature* 469, 245–249. doi: 10.1038/nature09585

Henikoff, S., and Dalal, Y. (2005). Centromeric chromatin: what makes it unique? *Curr. Opin. Genet. Dev.* 15, 177–184. doi: 10.1016/j.gde.2005.01.004

Hizi, A., and Levin, H. L. (2005). The integrase of the long terminal repeat-retrotransposon tf1 has a chromodomain that modulates integrase activities. *J. Biol. Chem.* 280, 39086–39094. doi: 10.1074/jbc.M506363200

Holeski, L. M., Jander, G., and Agrawal, A. A. (2012). Transgenerational defense induction and epigenetic inheritance in plants. *Trends Ecol. Evol.* 27, 618–626. doi: 10.1016/j.tree.2012.07.011

Hossain, M. B., Vahter, M., Concha, G., and Broberg, K. (2012). Low-level environmental cadmium exposure is associated with DNA hypomethylation in Argentinean women. *Environ. Health Perspect.* 120, 879–884. doi: 10.1289/ehp.1104600

Huang, H., and Jiao, R. (2012). Roles of chromatin assembly factor 1 in the epigenetic control of chromatin plasticity. *Sci. China Life Sci.* 55, 15–19. doi: 10.1007/s11427-012-4269-z

Huang, T., and Zhang, X. (2013). Host defense against DNA virus infection in shrimp is mediated by the siRNA pathway. *Eur. J. Immunol.* 43, 137–146. doi: 10.1002/eji.201242806

Hunt, C. R., Ramnarain, D., Horikoshi, N., Iyengar, P., Pandita, R. K., Shay, J. W., et al. (2013). Histone modifications and DNA double-strand break repair after exposure to ionizing radiations. *Radiat. Res.* 179, 383–392. doi: 10.1667/rr3308.2

Inacio, A., Pinho, J., Pereira, P. M., Comai, L., and Coelho, M. M. (2012). Global analysis of the small RNA transcriptome in different ploidies and genomic combinations of a vertebrate complex–the *Squalius alburnoides*. *PLoS ONE* 7:e41158. doi: 10.1371/journal.pone.0041158

Iyer, N. J., Jia, X., Sunkar, R., Tang, G., and Mahalingam, R. (2012). microRNAs responsive to ozone-induced oxidative stress in *Arabidopsis thaliana*. *Plant Signal. Behav.* 7, 484–491. doi: 10.4161/psb.19337

Jacobsen, S. C., Brons, C., Bork-Jensen, J., Ribel-Madsen, R., Yang, B., Lara, E., et al. (2012). Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men. *Diabetologia* 55, 3341–3349. doi: 10.1007/s00125-012-2717-8

Jiang, L., Wei, C., and Li, Y. (2012). Viral suppression of RNA silencing. *Sci. China Life Sci.* 55, 109–118. doi: 10.1007/s11427-012-4279-x

Kadyrova, L. Y., Mertz, T. M., Zhang, Y., Northam, M. R., Sheng, Z., Lobachev, K. S., et al. (2013). A reversible histone h3 acetylation cooperates with mismatch repair and replicative polymerases in maintaining genome stability. *PLoS Genet.* 9:e1003899. doi: 10.1371/journal.pgen.1003899

Kalmykova, A. I., Klenov, M. S., and Gvozdev, V. A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res.* 33, 2052–2059. doi: 10.1093/nar/gki323

Kaneda, A., Matsusaka, K., Aburatani, H., and Fukayama, M. (2012). Epstein-Barr virus infection as an epigenetic driver of tumorigenesis. *Cancer Res.* 72, 3445–3450. doi: 10.1158/0008-5472.can-11-3919

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., et al. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 9:e1003470. doi: 10.1371/journal.pgen.1003470

Karan, R., Deleon, T., Biradar, H., and Subudhi, P. K. (2012). Salt stress induced variation in DNA methylation pattern and its influence on gene expression in contrasting rice genotypes. *PLoS ONE* 7:e40203. doi: 10.1371/journal.pone.0040203

Kassis, J. A. (2002). Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. *Adv. Genet.* 46, 421–438. doi: 10.1016/S0065-2660(02)46015-4

Kathiria, P., Sidler, C., Golubov, A., Kalischuk, M., Kawchuk, L. M., and Kovalchuk, I. (2010). Tobacco mosaic virus infection results in an increase in recombination frequency and resistance to viral, bacterial, and fungal pathogens in the progeny of infected tobacco plants. *Plant Physiol.* 153, 1859–1870. doi: 10.1104/pp.110.157263

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., et al. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453, 793–797. doi: 10.1038/nature06938

Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., et al. (2011). Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics* 188, 263–272. doi: 10.1534/genetics.111.128348

Kettles, G. J., Drurey, C., Schoonbeek, H. J., Maule, A. J., and Hogenhout, S. A. (2013). Resistance of *Arabidopsis thaliana* to the green peach aphid, *Myzus persicae*, involves camalexin and is regulated by microRNAs. *New Phytol.* 198, 1178–1190. doi: 10.1111/nph.12218

Khraiwesh, B., Zhu, J. K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta* 1819, 137–148. doi: 10.1016/j.bbagrm.2011.05.001

Kinner, A., Wu, W., Staudt, C., and Iliakis, G. (2008). Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res.* 36, 5678–5694. doi: 10.1093/nar/gkn550

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W. J., Koornneef, M., et al. (2007). Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* 49, 38–45. doi: 10.1111/j.1365-313X.2006.02936.x

Kokosar, J., and Kordis, D. (2013). Genesis and regulatory wiring of retroelement-derived domesticated genes: a phylogenomic perspective. *Mol. Biol. Evol.* 30, 1015–1031. doi: 10.1093/molbev/mst014

Kordis, D. (2005). A genomic perspective on the chromodomain-containing retrotransposons: chromoviruses. *Gene* 347, 161–173. doi: 10.1016/j.gene.2004.12.017

Korzeniewski, N., Spardy, N., Duensing, A., and Duensing, S. (2011). Genomic instability and cancer: lessons learned from human papillomaviruses. *Cancer Lett.* 305, 113–122. doi: 10.1016/j.canlet.2010.10.013

Kou, H. P., Li, Y., Song, X. X., Ou, X. F., Xing, S. C., Ma, J., et al. (2011). Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* L.). *J. Plant Physiol.* 168, 1685–1693. doi: 10.1016/j.jplph.2011.03.017

Kraitshtein, Z., Yaakov, B., Khasdan, V., and Kashkush, K. (2010). Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* 186, 801–812. doi: 10.1534/genetics.110.120790

Kulcheski, F. R., De Oliveira, L. F., Molina, L. G., Almerao, M. P., Rodrigues, F. A., Marcolino, J., et al. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 12:307. doi: 10.1186/1471-2164-12-307

Labonte, B., Suderman, M., Maussion, G., Navaro, L., Yerko, V., Mahar, I., et al. (2012). Genome-wide epigenetic regulation by early-life trauma. *Arch. Gen. Psychiatry* 69, 722–731. doi: 10.1001/archgenpsychiatry.2011.2287

Lahue, R. S., and Frizzell, A. (2012). Histone deacetylase complexes as caretakers of genome stability. *Epigenetics* 7, 806–810. doi: 10.4161/epi.20922

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719

Lee, H. C., Gu, W., Shirayama, M., Youngman, E., Conte, D. Jr., and Mello, C. C. (2012). *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell* 150, 78–87. doi: 10.1016/j.cell.2012.06.016

Lelandais-Briere, C., Sorin, C., Crespi, M., and Hartmann, C. (2012). [Non-coding RNAs involved in plant responses to environmental constraints]. *Biol. Aujourdhui.* 206, 313–322. doi: 10.1051/jbio/2012032

Leonard, S., Wei, W., Anderton, J., Vockerodt, M., Rowe, M., Murray, P. G., et al. (2011). Epigenetic and transcriptional changes which follow Epstein-Barr virus infection of germinal center B cells and their relevance to the pathogenesis of Hodgkin's lymphoma. *J. Virol.* 85, 9568–9577. doi: 10.1128/jvi.00468-11

Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* 2:e60. doi: 10.1371/journal.ppat.0020060

Li, B., Duan, H., Li, J., Deng, X. W., Yin, W., and Xia, X. (2013). Global identification of miRNAs and targets in *Populus euphratica* under salt stress. *Plant Mol. Biol.* 81, 525–539. doi: 10.1007/s11103-013-0010-y

Li, B., Qin, Y., Duan, H., Yin, W., and Xia, X. (2011a). Genome-wide characterization of new and drought stress responsive microRNAs in *Populus euphratica*. *J. Exp. Bot.* 62, 3765–3779. doi: 10.1093/jxb/err051

Li, H., Dong, Y., Yin, H., Wang, N., Yang, J., Liu, X., et al. (2011b). Characterization of the stress associated microRNAs in Glycine max by deep sequencing. *BMC Plant Biol.* 11:170. doi: 10.1186/1471-2229-11-170

Li, Y., Lu, Y. G., Shi, Y., Wu, L., Xu, Y. J., Huang, F., et al. (2014). Multiple rice microRNAs are involved in immunity against the blast fungus *Magnaporthe oryzae*. *Plant Physiol.* 164, 1077–1092. doi: 10.1104/pp.113.230052

Lim, J. P., and Brunet, A. (2013). Bridging the transgenerational gap with epigenetic memory. *Trends Genet.* 29, 176–186. doi: 10.1016/j.tig.2012.12.008

Lim, K. Y., Soltis, D. E., Soltis, P. S., Tate, J., Matyasek, R., Srubarova, H., et al. (2008). Rapid chromosome evolution in recently formed polyploids in Tragopogon (Asteraceae). *PLoS ONE* 3:e3353. doi: 10.1371/journal.pone.0003353

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482. doi: 10.1038/nature10530

Linger, J. G., and Tyler, J. K. (2007). Chromatin disassembly and reassembly during DNA repair. *Mutat. Res.* 618, 52–64. doi: 10.1016/j.mrfmmm.2006.05.039

Lisby, M., and Rothstein, R. (2005). Localization of checkpoint and repair proteins in eukaryotes. *Biochimie* 87, 579–589. doi: 10.1016/j.biochi.2004.10.023

Liu, Q., and Gong, Z. (2011). The coupling of epigenome replication with DNA replication. *Curr. Opin. Plant Biol.* 14, 187–194. doi: 10.1016/j.pbi.2010.12.001

Liu, Z., Kumari, S., Zhang, L., Zheng, Y., and Ware, D. (2012). Characterization of miRNAs in response to short-term waterlogging in three inbred lines of Zea mays. *PLoS ONE* 7:e39786. doi: 10.1371/journal.pone.0039786

Llano, M., Saenz, D. T., Meehan, A., Wongthida, P., Peretz, M., Walker, W. H., et al. (2006). An essential role for LEDGF/p75 in HIV integration. *Science* 314, 461–464. doi: 10.1126/science.1132319

Long, M. (2001). Evolution of novel genes. *Curr. Opin. Genet. Dev.* 11, 673–680. doi: 10.1016/S0959-437X(00)00252-5

Lowe, C. B., Kellis, M., Siepel, A., Raney, B. J., Clamp, M., Salama, S. R., et al. (2011). Three periods of regulatory innovation during vertebrate evolution. *Science* 333, 1019–1024. doi: 10.1126/science.1202702

Luna, E., and Ton, J. (2012). The epigenetic machinery controlling transgenerational systemic acquired resistance. *Plant Signal. Behav.* 7, 615–618. doi: 10.4161/psb.20155

Luo, J., Hao, M., Zhang, L., Chen, J., Zhang, L., Yuan, Z., et al. (2012). Microsatellite mutation rate during allohexaploidization of newly resynthesized wheat. *Int. J. Mol. Sci.* 13, 12533–12543. doi: 10.3390/ijms131012533

Macovei, A., and Tuteja, N. (2012). microRNAs targeting DEAD-box helicases are involved in salinity stress response in rice (*Oryza sativa* L.). *BMC Plant Biol.* 12:183. doi: 10.1186/1471-2229-12-183

Madlung, A., and Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* 140, 270–285. doi: 10.1159/000351430

Maillard, P. V., Ciaudo, C., Marchais, A., Li, Y., Jay, F., Ding, S. W., et al. (2013). Antiviral RNA interference in mammalian cells. *Science* 342, 235–238. doi: 10.1126/science.1241930

Maksakova, I. A., Romanish, M. T., Gagnier, L., Dunn, C. A., Van De Lagemaat, L. N., and Mager, D. L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2:e2. doi: 10.1371/journal.pgen.0020002

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M. K. (2012). Dioxin (TCDD) induces epigenetic transgenerational inheritance of adult onset disease and sperm epimutations. *PLoS ONE* 7:e46249. doi: 10.1371/journal.pone.0046249

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M. K. (2013). Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. *PLoS ONE* 8:e55387. doi: 10.1371/journal.pone.0055387

Marraffini, L. A., and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181–190. doi: 10.1038/nrg2749

Martis, M. M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrana, J., Kubalakova, M., et al. (2013). Reticulate evolution of the rye genome. *Plant Cell* 25, 3685–3698. doi: 10.1105/tpc.113.114553

Masaki, T., Qu, J., Cholewa-Waclaw, J., Burr, K., Raaum, R., and Rambukkana, A. (2013). Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection. *Cell* 152, 51–67. doi: 10.1016/j.cell.2012.12.014

Matsunaga, W., Kobayashi, A., Kato, A., and Ito, H. (2012). The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in *Arabidopsis thaliana*. *Plant Cell Physiol.* 53, 824–833. doi: 10.1093/pcp/pcr179

Mccue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet.* 8:e1002474. doi: 10.1371/journal.pgen.1002474

Meehan, A. M., and Poeschla, E. M. (2010). Chromatin tethering and retroviral integration: recent discoveries and parallels with DNA viruses. *Biochim. Biophys. Acta* 1799, 182–191. doi: 10.1016/j.bbagrm.2009.10.001

Mercer, T. R., Edwards, S. L., Clark, M. B., Neph, S. J., Wang, H., Stergachis, A. B., et al. (2013). DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* 45, 852–859. doi: 10.1038/ng.2677

Mercer, T. R., and Mattick, J. S. (2013). Understanding the regulatory and transcriptional complexity of the genome through structure. *Genome Res.* 23, 1081–1088. doi: 10.1101/gr.156612.113

Mermoud, J. E., Rowbotham, S. P., and Varga-Weisz, P. D. (2011). Keeping chromatin quiet: how nucleosome remodeling restores heterochromatin after replication. *Cell Cycle* 10, 4017–4025. doi: 10.4161/cc.10.23.18558

Michalak, M., Barciszewska, M. Z., Barciszewski, J., Plitta, B. P., and Chmielarz, P. (2013). Global changes in DNA methylation in seeds and seedlings of *Pyrus communis* after seed desiccation and storage. *PLoS ONE* 8:e70693. doi: 10.1371/journal.pone.0070693

Miller, K. M., Tjeertes, J. V., Coates, J., Legube, G., Polo, S. E., Britton, S., et al. (2010). Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat. Struct. Mol. Biol.* 17, 1144–1151. doi: 10.1038/nsmb.1899

Miller, M., Zhang, C., and Chen, Z. J. (2012). Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. *G3 (Bethesda)* 2, 505–513. doi: 10.1534/g3.112.002162

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., et al. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* 461, 427–430. doi: 10.1038/nature08328

Moldovan, D., Spriggs, A., Yang, J., Pogson, B. J., Dennis, E. S., and Wilson, I. W. (2010). Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in Arabidopsis. *J. Exp. Bot.* 61, 165–177. doi: 10.1093/jxb/erp296

Mou, Z., Kenny, A. E., and Curcio, M. J. (2006). Hos2 and Set3 promote integration of Ty1 retrotransposons at tRNA genes in *Saccharomyces cerevisiae*. *Genetics* 172, 2157–2167 doi: 10.1534/genetics.105.054072

Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S. J., and Boeke, J. D. (2012). Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res.* 22, 693–703. doi: 10.1101/gr.129460.111

Munoz-Galvan, S., Jimeno, S., Rothstein, R., and Aguilera, A. (2013). Histone H3K56 acetylation, Rad52, and non-DNA repair factors control double-strand break repair choice with the sister chromatid. *PLoS Genet.* 9:e1003237. doi: 10.1371/journal.pgen.1003237

Nakaminami, K., Matsui, A., Shinozaki, K., and Seki, M. (2012). RNA regulation in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819, 149–153. doi: 10.1016/j.bbagrm.2011.07.015

Nakayashiki, H. (2011). The Trickster in the genome: contribution and control of transposable elements. *Genes Cells* 16, 827–841. doi: 10.1111/j.1365-2443.2011.01533.x

Narlikar, G. J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 154, 490–503. doi: 10.1016/j.cell.2013.07.011

Navarro, B., Pantaleo, V., Gisel, A., Moxon, S., Dalmay, T., Bisztray, G., et al. (2009). Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. *PLoS ONE* 4:e7686. doi: 10.1371/journal.pone.0007686

Neumann, P., Navratilova, A., Koblizkova, A., Kejnovsky, E., Hribova, E., Hobza, R., et al. (2011). Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob. DNA* 2:4. doi: 10.1186/1759-8753-2-4

Neumann, P., Yan, H., and Jiang, J. (2007). The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 176, 749–761. doi: 10.1534/genetics.107.071902

Neves-Costa, A., and Varga-Weisz, P. (2006). The roles of chromatin remodelling factors in replication. *Results Probl. Cell Differ.* 41, 91–107. doi: 10.1007/400_007

Ng, D. W., Lu, J., and Chen, Z. J. (2012). Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* 15, 154–161. doi: 10.1016/j.pbi.2012.01.007

Novikov, A., Smyshlyaev, G., and Novikova, O. (2012). Evolutionary history of LTR retrotransposon chromodomains in plants. *Int. J. Plant Genomics* 2012:874743. doi: 10.1155/2012/874743

Nuthikattu, S., Mccue, A. D., Panda, K., Fultz, D., Defraia, C., Thomas, E. N., et al. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* 162, 116–131. doi: 10.1104/pp.113.216481

Omarov, R. T., and Scholthof, H. B. (2012). Biological chemistry of virus-encoded suppressors of RNA silencing: an overview. *Methods Mol. Biol.* 894, 39–56. doi: 10.1007/978-1-61779-882-5_3

Osley, M. A., Tsukuda, T., and Nickoloff, J. A. (2007). ATP-dependent chromatin remodeling factors and DNA damage repair. *Mutat. Res.* 618, 65–80. doi: 10.1016/j.mrfmmm.2006.07.011

Ou, X., Zhang, Y., Xu, C., Lin, X., Zang, Q., Zhuang, T., et al. (2012). Transgenerational inheritance of modified DNA methylation patterns and enhanced tolerance induced by heavy metal stress in rice (*Oryza sativa* L.). *PLoS ONE* 7:e41143. doi: 10.1371/journal.pone.0041143

Ozhuner, E., Eldem, V., Ipek, A., Okay, S., Sakcali, S., Zhang, B., et al. (2013). Boron stress responsive microRNAs and their targets in barley. *PLoS ONE* 8:e59543. doi: 10.1371/journal.pone.0059543

Palomera-Sanchez, Z., and Zurita, M. (2011). Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage. *DNA Repair (Amst.)* 10, 119–125. doi: 10.1016/j.dnarep.2010.10.010

Pang, M., Woodward, A. W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., et al. (2009). Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biol.* 10:R122. doi: 10.1186/gb-2009-10-11-r122

Pantaleo, V. (2011). Plant RNA silencing in viral defence. *Adv. Exp. Med. Biol.* 722, 39–58. doi: 10.1007/978-1-4614-0332-6_3

Partridge, J. F. (2008). Centromeric chromatin in fission yeast. *Front. Biosci.* 13, 3896–3905. doi: 10.2741/2977

Pask, A. J., Papenfuss, A. T., Ager, E. I., Mccoll, K. A., Speed, T. P., and Renfree, M. B. (2009). Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome Biol.* 10:R1. doi: 10.1186/gb-2009-10-1-r1

Peaston, A. E., Evsikov, A. V., Graber, J. H., De Vries, W. N., Holbrook, A. E., Solter, D., et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606. doi: 10.1016/j.devcel.2004.09.004

Pelaez, P., and Sanchez, F. (2013). Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Front. Plant Sci.* 4:343. doi: 10.3389/fpls.2013.00343

Perez-Quintero, A. L., Quintero, A., Urrego, O., Vanegas, P., and Lopez, C. (2012). Bioinformatic identification of cassava miRNAs differentially expressed in response to infection by *Xanthomonas axonopodis* pv. manihotis. *BMC Plant Biol.* 12:29. doi: 10.1186/1471-2229-12-29

Plate, I., Hallwyl, S. C., Shi, I., Krejci, L., Muller, C., Albertsen, L., et al. (2008). Interaction with RPA is necessary for Rad52 repair center formation and for its mediator activity. *J. Biol. Chem.* 283, 29077–29085. doi: 10.1074/jbc.M804881200

Poot, R. A., Bozhenok, L., Van Den Berg, D. L., Hawkes, N., and Varga-Weisz, P. D. (2005). Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance? *Cell Cycle* 4, 543–546. doi: 10.4161/cc.4.4.1624

Price, B. D., and D'andrea, A. D. (2013). Chromatin remodeling at DNA double-strand breaks. *Cell* 152, 1344–1354. doi: 10.1016/j.cell.2013.02.011

Qin, Y., Duan, Z., Xia, X., and Yin, W. (2011). Expression profiles of precursor and mature microRNAs under dehydration and high salinity shock in

Populus euphratica. Plant Cell Rep. 30, 1893–1907. doi: 10.1007/s00299-011-1096-9

Queen, K. J., Shi, M., Zhang, F., Cvek, U., and Scott, R. S. (2013). Epstein-Barr virus-induced epigenetic alterations following transient infection. *Int. J. Cancer* 132, 2076–2086. doi: 10.1002/ijc.27893

Quintero, A., Perez-Quintero, A. L., and Lopez, C. (2013). Identification of ta-siRNAs and cis-nat-siRNAs in cassava and their roles in response to cassava bacterial blight. *Genom. Proteomics Bioinform.* 11, 172–181. doi: 10.1016/j.gpb.2013.03.001

Rahavi, M. R., Migicovsky, Z., Titov, V., and Kovalchuk, I. (2011). Transgenerational adaptation to heavy metal salts in Arabidopsis. *Front. Plant Sci.* 2:91. doi: 10.3389/fpls.2011.00091

Raja, P., Jackel, J. N., Li, S., Heard, I. M., and Bisaro, D. M. (2014). Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *J. Virol.* 88, 2611–2622. doi: 10.1128/JVI.02305-13

Raman, V., Simon, S. A., Romag, A., Demirci, F., Mathioni, S. M., Zhai, J., et al. (2013). Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, Magnaporthe oryzae. *BMC Genomics* 14:326. doi: 10.1186/1471-2164-14-326

Ramesh, S. V., Ratnaparkhe, M. B., Kumawat, G., Gupta, G. K., and Husain, S. M. (2014). Plant miRNAome and antiviral resistance: a retrospective view and prospective challenges. *Virus Genes.* 48, 1–14. doi: 10.1007/s11262-014-1038-z

Rechavi, O., Minevich, G., and Hobert, O. (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans. Cell* 147, 1248–1256. doi: 10.1016/j.cell.2011.10.042

Redon, C. E., Dickey, J. S., Bonner, W. M., and Sedelnikova, O. A. (2009). gamma-H2AX as a biomarker of DNA damage induced by ionizing radiation in human peripheral blood lymphocytes and artificial skin. *Adv. Space Res.* 43, 1171–1178. doi: 10.1016/j.asr.2008.10.011

Ren, Y., Chen, L., Zhang, Y., Kang, X., Zhang, Z., and Wang, Y. (2012). Identification of novel and conserved Populus tomentosa microRNA as components of a response to water stress. *Funct. Integr. Genomics* 12, 327–339. doi: 10.1007/s10142-012-0271-6

Ren, Y., Chen, L., Zhang, Y., Kang, X., Zhang, Z., and Wang, Y. (2013). Identification and characterization of salt-responsive microRNAs in Populus tomentosa by high-throughput sequencing. *Biochimie* 95, 743–750. doi: 10.1016/j.biochi.2012.10.025

Rigal, M., and Mathieu, O. (2011). A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim. Biophys. Acta* 1809, 452–458. doi: 10.1016/j.bbagrm.2011.04.001

Robert, T., Vanoli, F., Chiolo, I., Shubassi, G., Bernstein, K. A., Rothstein, R., et al. (2011). HDACs link the DNA damage response, processing of double-strand breaks and autophagy. *Nature* 471, 74–79. doi: 10.1038/nature09803

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* 273, 5858–5868.

Rongrui, L., Na, H., Zongfang, L., Fanpu, J., and Shiwen, J. (2014). Epigenetic mechanism involved in the HBV/HCV-related hepatocellular carcinoma tumorigenesis. *Curr. Pharm. Des.* 20, 1715–1725. doi: 10.2174/13816128113199990533

Ruiz-Ferrer, V., and Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.* 60, 485–510. doi: 10.1146/annurev.arplant.043008.092111

Ryan, D. P., and Owen-Hughes, T. (2011). Snf2-family proteins: chromatin remodellers for any occasion. *Curr. Opin. Chem. Biol.* 15, 649–656. doi: 10.1016/j.cbpa.2011.07.022

Sabin, L. R., and Cherry, S. (2013). Small creatures use small RNAs to direct antiviral defenses. *Eur. J. Immunol.* 43, 27–33. doi: 10.1002/eji.201243201

Sani, E., Herzyk, P., Perrella, G., Colot, V., and Amtmann, A. (2013). Hyperosmotic priming of Arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol.* 14:R59. doi: 10.1186/gb-2013-14-6-r59

Savva, Y. A., Jepson, J. E., Chang, Y. J., Whitaker, R., Jones, B. C., St. Laurent, G., et al. (2013). RNA editing regulates transposon-mediated heterochromatic gene silencing. *Nat. Commun.* 4:2745. doi: 10.1038/ncomms3745

Schmidt, C. K., and Jackson, S. P. (2013). On your mark, get SET(D2), go! H3K36me3 primes DNA mismatch repair. *Cell* 153, 513–515. doi: 10.1016/j.cell.2013.04.018

Seeber, A., Hauer, M., and Gasser, S. M. (2013). Nucleosome remodelers in double-strand break repair. *Curr. Opin. Genet. Dev.* 23, 174–184. doi: 10.1016/j.gde.2012.12.008

Sha, A., Zhao, J., Yin, K., Tang, Y., Wang, Y., Wei, X., et al. (2014). Virus-based microRNA silencing in plants. *Plant Physiol.* 164, 36–47. doi: 10.1104/pp.113.231100

Shapiro, J. A. (2011). *Evolution: a View from the 21st Century.* Upper Saddle River, NJ: FT Press Science.

Shapiro, J. A. (2014). The physiology of the Read-Write (RW) genome. *J. Physiol.* (in press).

Sharma, A. (2013). Transgenerational epigenetic inheritance: focus on soma to germline information transfer. *Prog. Biophys. Mol. Biol.* 113, 439–446. doi: 10.1016/j.pbiomolbio.2012.12.003

Sharma, A., Larue, R. C., Plumb, M. R., Malani, N., Male, F., Slaughter, A., et al. (2013a). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12036–12041. doi: 10.1073/pnas.1307157110

Sharma, A., Wolfgruber, T. K., and Presting, G. G. (2013b). Tandem repeats derived from centromeric retrotransposons. *BMC Genomics* 14:142. doi: 10.1186/1471-2164-14-142

Shi, L., and Oberdoerffer, P. (2012). Chromatin dynamics in DNA double-strand break repair. *Biochim. Biophys. Acta* 1819, 811–819. doi: 10.1016/j.bbagrm.2012.01.002

Shivaprasad, P. V., Dunn, R. M., Santos, B. A., Bassett, A., and Baulcombe, D. C. (2012). Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.* 31, 257–266. doi: 10.1038/emboj.2011.458

Shpiz, S., Kwon, D., Rozovsky, Y., and Kalmykova, A. (2009). rasiRNA pathway controls antisense expression of Drosophila telomeric retrotransposons in the nucleus. *Nucleic Acids Res.* 37, 268–278. doi: 10.1093/nar/gkn960

Shu, L., and Hu, Z. (2012). Characterization and differential expression of microRNAs elicited by sulfur deprivation in *Chlamydomonas reinhardtii. BMC Genomics* 13:108. doi: 10.1186/1471-2164-13-108

Shuai, P., Liang, D., Zhang, Z., Yin, W., and Xia, X. (2013). Identification of drought-responsive and novel Populus trichocarpa microRNAs by high-throughput sequencing and their targets using degradome analysis. *BMC Genomics* 14:233. doi: 10.1186/1471-2164-14-233

Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980. doi: 10.1016/j.cell.2012.10.040

Sijen, T., and Plasterk, R. H. (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* 426, 310–314. doi: 10.1038/nature02107

Skinner, M. K., Haque, C. G., Nilsson, E., Bhandari, R., and Mccarrey, J. R. (2013). Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line. *PLoS ONE* 8:e66318. doi: 10.1371/journal.pone.0066318

Slaughter, A., Daniel, X., Flors, V., Luna, E., Hohn, B., and Mauch-Mani, B. (2012). Descendants of primed Arabidopsis plants exhibit resistance to biotic stress. *Plant Physiol.* 158, 835–843. doi: 10.1104/pp.111.191593

Slotkin, R. K. (2010). The epigenetic control of the Athila family of retrotransposons in Arabidopsis. *Epigenetics* 5, 483–490. doi: 10.4161/epi.5.6.12119

Sokolov, M. V., Dickey, J. S., Bonner, W. M., and Sedelnikova, O. A. (2007). gamma-H2AX in bystander cells: not just a radiation-triggered event, a cellular response to stress mediated by intercellular communication. *Cell Cycle* 6, 2210–2212. doi: 10.4161/cc.6.18.4682

Soubry, A., Hoyo, C., Jirtle, R. L., and Murphy, S. K. (2014). A paternal environmental legacy: evidence for epigenetic inheritance through the male germ line. *Bioessays* 36, 359–371. doi: 10.1002/bies.201300113

Stenberg, P., and Saura, A. (2013). Meiosis and its deviations in polyploid animals. *Cytogenet. Genome Res.* 140, 185–203. doi: 10.1159/000351731

Sun, G., Stewart, C. N., Jr., Xiao, P., and Zhang, B. (2012). MicroRNA expression analysis in the cellulosic biofuel crop switchgrass (*Panicum virgatum*) under abiotic stress. *PLoS ONE* 7:e32017. doi: 10.1371/journal.pone.0032017

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J. K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci.* 12, 301–309. doi: 10.1016/j.tplants.2007.05.001

Suzuki, S., Ono, R., Narita, T., Pask, A. J., Shaw, G., Wang, C., et al. (2007). Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet.* 3:e55. doi: 10.1371/journal.pgen.0030055

Svetlova, M. P., Solovjeva, L. V., and Tomilin, N. V. (2010). Mechanism of elimination of phosphorylated histone H2AX from chromatin after repair of DNA double-strand breaks. *Mutat. Res.* 685, 54–60. doi: 10.1016/j.mrfmmm.2009.08.001

Takahashi, K. (2014). Influence of bacteria on epigenetic gene control. *Cell. Mol. Life Sci.* 71, 1045–1054. doi: 10.1007/s00018-013-1487-x

Tang, Z., Zhang, L., Xu, C., Yuan, S., Zhang, F., Zheng, Y., et al. (2012). Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. *Plant Physiol.* 159, 721–738. doi: 10.1104/pp.112.196048

Tayale, A., and Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140, 79–96. doi: 10.1159/000351318

Testori, A., Caizzi, L., Cutrupi, S., Friard, O., De Bortoli, M., Cora, D., et al. (2012). The role of transposable elements in shaping the combinatorial interaction of transcription factors. *BMC Genomics* 13:400. doi: 10.1186/1471-2164-13-400

Tian, Y., Yang, W., Song, J., Wu, Y., and Ni, B. (2013). Hepatitis B virus X protein-induced aberrant epigenetic modifications contributing to human hepatocellular carcinoma pathogenesis. *Mol. Cell. Biol.* 33, 2810–2816. doi: 10.1128/mcb.01205-13

Tian, Z., Yu, Y., Lin, F., Sanmiguel, P. J., Wing, R. A., Mccouch, S. R., et al. (2011). Exceptional lability of a genomic complex in rice and its close relatives revealed by interspecific and intraspecific comparison and population analysis. *BMC Genomics* 12:142. doi: 10.1186/1471-2164-12-142

Tomas, D., Bento, M., Viegas, W., and Silva, M. (2012). Involvement of disperse repetitive sequences in wheat/rye genome adjustment. *Int. J. Mol. Sci.* 13, 8549–8561. doi: 10.3390/ijms13078549

Tsukahara, S., Kawabe, A., Kobayashi, A., Ito, T., Aizu, T., Shin-I, T., et al. (2012). Centromere-targeted de novo integrations of an LTR retrotransposon of Arabidopsis lyrata. *Genes Dev.* 26, 705–713. doi: 10.1101/gad.183871.111

Vago, R., Leva, V., Biamonti, G., and Montecucco, A. (2009). DNA ligase I and Nbs1 proteins associate in a complex and colocalize at replication factories. *Cell Cycle* 8, 2600–2607. doi: 10.4161/cc.8.16.9352

Van Attikum, H., and Gasser, S. M. (2005). ATP-dependent chromatin remodeling and DNA double-strand break repair. *Cell Cycle* 4, 1011–1014. doi: 10.4161/cc.4.8.1887

Van Rij, R. P., Saleh, M. C., Berry, B., Foo, C., Houk, A., Antoniewski, C., et al. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* 20, 2985–2995. doi: 10.1101/gad.1482006

Vandegehuchte, M. B., and Janssen, C. R. (2013). Epigenetics in an ecotoxicological context. *Mutat. Res*. doi: 10.1016/j.mrgentox.2013.08.008. [Epub ahead of print].

Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M., and Poeschla, E. (2005). Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J. Cell Sci.* 118, 1733–1743. doi: 10.1242/jcs.02299

Vanitharani, R., Chellappan, P., and Fauquet, C. M. (2005). Geminiviruses and RNA silencing. *Trends Plant Sci.* 10, 144–151. doi: 10.1016/j.tplants.2005.01.005

Vos, L. J., Famulski, J. K., and Chan, G. K. (2006). How to build a centromere: from centromeric and pericentromeric chromatin to kinetochore assembly. *Biochem. Cell Biol.* 84, 619–639. doi: 10.1139/o06-078

Wang, B., Li, Y., Shao, C., Tan, Y., and Cai, L. (2012). Cadmium and its epigenetic effects. *Curr. Med. Chem.* 19, 2611–2620. doi: 10.2174/092986712800492913

Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C., and Bushman, F. D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* 17, 1186–1194. doi: 10.1101/gr.6286907

Wang, H., Chai, Y., Chu, X., Zhao, Y., Wu, Y., Zhao, J., et al. (2009). Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. *BMC Plant Biol.* 9:63. doi: 10.1186/1471-2229-9-63

Wang, N., Wang, H., Wang, H., Zhang, D., Wu, Y., Ou, X., et al. (2010). Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. *BMC Plant Biol.* 10:190. doi: 10.1186/1471-2229-10-190

Wang, T., Chen, L., Zhao, M., Tian, Q., and Zhang, W. H. (2011). Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *BMC Genomics* 12:367. doi: 10.1186/1471-2164-12-367

Wang, X., Wu, R., Lin, X., Bai, Y., Song, C., Yu, X., et al. (2013a). Tissue culture-induced genetic and epigenetic alterations in rice pure-lines, F1 hybrids and polyploids. *BMC Plant Biol.* 13:77. doi: 10.1186/1471-2229-13-77

Wang, Z. H., Zhang, D., Bai, Y., Zhang, Y. H., Liu, Y., Wu, Y., et al. (2013b). Genomewide variation in an introgression line of rice-Zizania revealed by whole-genome re-sequencing. *PLoS ONE* 8:e74479. doi: 10.1371/journal.pone.0074479

Watanabe, T., Nozawa, T., Aikawa, C., Amano, A., Maruyama, F., and Nakagawa, I. (2013). CRISPR regulation of intraspecies diversification by limiting IS transposition and intercellular recombination. *Genome Biol. Evol.* 5, 1099–1114. doi: 10.1093/gbe/evt075

Weber, B., Heitkam, T., Holtgrawe, D., Weisshaar, B., Minoche, A. E., Dohm, J. C., et al. (2013). Highly diverse chromoviruses of Beta vulgaris are classified by chromodomains and chromosomal integration. *Mob. DNA* 4:8. doi: 10.1186/1759-8753-4-8

Weiner, A., Chen, H. V., Liu, C. L., Rahat, A., Klien, A., Soares, L., et al. (2012). Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol.* 10:e1001369. doi: 10.1371/journal.pbio.1001369

Williams, R. S., Williams, J. S., and Tainer, J. A. (2007). Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. *Biochem. Cell Biol.* 85, 509–520. doi: 10.1139/O07-069

Winkler, D. D., and Luger, K. (2011). The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization. *J. Biol. Chem.* 286, 18369–18374. doi: 10.1074/jbc.R110.180778

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D. H., Shi, J., et al. (2009). Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet.* 5:e1000743. doi: 10.1371/journal.pgen.1000743

Wu, R., Wang, X., Lin, Y., Ma, Y., Liu, G., Yu, X., et al. (2013). Interspecies grafting caused extensive and heritable alterations of DNA methylation in solanaceae plants. *PLoS ONE* 8:e61995. doi: 10.1371/journal.pone.0061995

Xiao, J., Song, C., Liu, S., Tao, M., Hu, J., Wang, J., et al. (2013). DNA methylation analysis of allotetraploid hybrids of red crucian carp (*Carassius auratus* red var.) and common carp (*Cyprinus carpio* L.). *PLoS ONE* 8:e56409. doi: 10.1371/journal.pone.0056409

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol. Cell. Biol.* 21, 6606–6614. doi: 10.1128/MCB.21.19.6606-6614.2001

Xin, M., Wang, Y., Yao, Y., Song, N., Hu, Z., Qin, D., et al. (2011). Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol.* 11:61. doi: 10.1186/1471-2229-11-61

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., et al. (2010). Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 10:123. doi: 10.1186/1471-2229-10-123

Xiong, Z., Gaeta, R. T., and Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7908–7913. doi: 10.1073/pnas.1014138108

Xu, L., Wang, Y., Zhai, L., Xu, Y., Wang, L., Zhu, X., et al. (2013). Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots. *J. Exp. Bot.* 64, 4271–4287. doi: 10.1093/jxb/ert240

Xu, Y., Ayrapetov, M. K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y., and Price, B. D. (2012a). Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. *Mol. Cell* 48, 723–733. doi: 10.1016/j.molcel.2012.09.026

Xu, Y., Huang, L., Fu, S., Wu, J., and Zhou, X. (2012b). Population diversity of rice stripe virus-derived siRNAs in three different hosts and RNAi-based antiviral immunity in Laodelphgax striatellus. *PLoS ONE* 7:e46238. doi: 10.1371/journal.pone.0046238

Xu, Y., and Price, B. D. (2011). Chromatin dynamics and the repair of DNA double strand breaks. *Cell Cycle* 10, 261–267. doi: 10.4161/cc.10.2.14543

Xu, Z., Zhong, S., Li, X., Li, W., Rothstein, S. J., Zhang, S., et al. (2011). Genome-wide identification of microRNAs in response to low nitrate availability in maize leaves and roots. *PLoS ONE* 6:e28009. doi: 10.1371/journal.pone.0028009

Yaakov, B., and Kashkush, K. (2011a). Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. *Genome* 54, 42–49. doi: 10.1139/g10-091

Yaakov, B., and Kashkush, K. (2011b). Methylation, transcription, and rearrangements of transposable elements in synthetic allopolyploids. *Int. J. Plant Genomics* 2011:569826. doi: 10.1155/2011/569826

Yaakov, B., and Kashkush, K. (2012). Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. *Plant Mol. Biol.* 80, 419–427. doi: 10.1007/s11103-012-9957-3

Yang, L., Jue, D., Li, W., Zhang, R., Chen, M., and Yang, Q. (2013). Identification of MiRNA from eggplant (*Solanum melongena* L.) by small RNA deep sequencing and their response to *Verticillium dahliae* infection. *PLoS ONE* 8:e72840. doi: 10.1371/journal.pone.0072840

Yang, N., and Kazazian, H. H. Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 13, 763–771. doi: 10.1038/nsmb1141

Yang, X., Xie, Y., Raja, P., Li, S., Wolf, J. N., Shen, Q., et al. (2011). Suppression of methylation-mediated transcriptional gene silencing by betaC1-SAHH protein interaction during geminivirus-betasatellite infection. *PLoS Pathog.* 7:e1002329. doi: 10.1371/journal.ppat.1002329

Yang, X., Yu, Y., Jiang, L., Lin, X., Zhang, C., Ou, X., et al. (2012). Changes in DNA methylation and transgenerational mobilization of a transposable element (mPing) by the topoisomerase II inhibitor, etoposide, in rice. *BMC Plant Biol.* 12:48. doi: 10.1186/1471-2229-12-48

Yao, Y., Danna, C. H., Zemp, F. J., Titov, V., Ciftci, O. N., Przybylski, R., et al. (2011). UV-C-irradiated Arabidopsis and tobacco emit volatiles that trigger genomic instability in neighboring plants. *Plant Cell* 23, 3842–3852. doi: 10.1105/tpc.111.089003

Yao, Y., Kathiria, P., and Kovalchuk, I. (2013). A systemic increase in the recombination frequency upon local infection of *Arabidopsis thaliana* plants with oilseed rape mosaic virus depends on plant age, the initial inoculum concentration and the time for virus replication. *Front. Plant Sci.* 4:61. doi: 10.3389/fpls.2013.00061

Yao, Y., and Kovalchuk, I. (2011). Abiotic stress leads to somatic and heritable changes in homologous recombination frequency, point mutation frequency and microsatellite stability in Arabidopsis plants. *Mutat. Res.* 707, 61–66. doi: 10.1016/j.mrfmmm.2010.12.013

Yap, K. L., and Zhou, M. M. (2011). Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* 50, 1966–1980. doi: 10.1021/bi101885m

Ye, Y. H., Woolfit, M., Huttley, G. A., Rances, E., Caragata, E. P., Popovici, J., et al. (2013). Infection with a virulent strain of disrupts genome wide-patterns of cytosine methylation in the mosquito. *PLoS ONE* 8:e66482. doi: 10.1371/journal.pone.0066482

Yin, Z., Li, Y., Han, X., and Shen, F. (2012). Genome-wide profiling of miRNAs and other small non-coding RNAs in the *Verticillium dahliae*-inoculated cotton roots. *PLoS ONE* 7:e35765. doi: 10.1371/journal.pone.0035765

Youngson, N. A., Kocialkowski, S., Peel, N., and Ferguson-Smith, A. C. (2005). A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. *J. Mol. Evol.* 61, 481–490. doi: 10.1007/s00239-004-0332-0

Yu, S., Teng, Y., Waters, R., and Reed, S. H. (2011). How chromatin is remodelled during DNA repair of UV-induced DNA damage in *Saccharomyces cerevisiae*. *PLoS Genet.* 7:e1002124. doi: 10.1371/journal.pgen.1002124

Yu, Y., Deng, Y., Reed, S. H., Millar, C. B., and Waters, R. (2013a). Histone variant Htz1 promotes histone H3 acetylation to enhance nucleotide excision repair in Htz1 nucleosomes. *Nucleic Acids Res.* 41, 9006–9019. doi: 10.1093/nar/gkt688

Yu, Y., Yang, X., Wang, H., Shi, F., Liu, Y., Liu, J., et al. (2013b). Cytosine methylation alteration in natural populations of *Leymus chinensis* induced by multiple abiotic stresses. *PLoS ONE* 8:e55772. doi: 10.1371/journal.pone.0055772

Zeng, Q. Y., Yang, C. Y., Ma, Q. B., Li, X. P., Dong, W. W., and Nian, H. (2012). Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. *BMC Plant Biol.* 12:182. doi: 10.1186/1471-2229-12-182

Zhai, L., Liu, Z., Zou, X., Jiang, Y., Qiu, F., Zheng, Y., et al. (2013). Genome-wide identification and analysis of microRNA responding to long-term waterlogging in crown roots of maize seedlings. *Physiol. Plant.* 147, 181–193. doi: 10.1111/j.1399-3054.2012.01653.x

Zhan, X., Wang, B., Li, H., Liu, R., Kalia, R. K., Zhu, J. K., et al. (2012). Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18198–18203. doi: 10.1073/pnas.1216199109

Zhang, W., Gao, S., Zhou, X., Chellappan, P., Chen, Z., Zhou, X., et al. (2011). Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. *Plant Mol. Biol.* 75, 93–105. doi: 10.1007/s11103-010-9710-8

Zhang, W., Lee, H. R., Koo, D. H., and Jiang, J. (2008a). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell* 20, 25–34. doi: 10.1105/tpc.107.057083

Zhang, X., Ge, X., Shao, Y., Sun, G., and Li, Z. (2013). Genomic change, retrotransposon mobilization and extensive cytosine methylation alteration in *Brassica napus* introgressions from two intertribal hybridizations. *PLoS ONE* 8:e56346. doi: 10.1371/journal.pone.0056346

Zhang, Y., and Mager, D. L. (2012). Gene properties and chromatin state influence the accumulation of transposable elements in genes. *PLoS ONE* 7:e30158. doi: 10.1371/journal.pone.0030158

Zhang, Z., Lin, H., Shen, Y., Gao, J., Xiang, K., Liu, L., et al. (2012). Cloning and characterization of miRNAs from maize seedling roots under low phosphorus stress. *Mol. Biol. Rep.* 39, 8137–8146. doi: 10.1007/s11033-012-1661-5

Zhang, Z., Wei, L., Zou, X., Tao, Y., Liu, Z., and Zheng, Y. (2008b). Submergence-responsive MicroRNAs are potentially involved in the regulation of morphological and metabolic adaptations in maize root cells. *Ann. Bot.* 102, 509–519. doi: 10.1093/aob/mcn129

Zheng, X., Chen, L., Li, M., Lou, Q., Xia, H., Wang, P., et al. (2013). Transgenerational variations in DNA methylation induced by drought stress in two rice varieties with distinguished difference to drought resistance. *PLoS ONE* 8:e80253. doi: 10.1371/journal.pone.0080253

Zheng, Y., Ao, Z., Jayappa, K. D., and Yao, X. (2010). Characterization of the HIV-1 integrase chromatin- and LEDGF/p75-binding abilities by mutagenic analysis within the catalytic core domain of integrase. *Virol. J.* 7:68. doi: 10.1186/1743-422x-7-68

Zhou, M., Li, D., Li, Z., Hu, Q., Yang, C., Zhu, L., et al. (2013). Constitutive expression of a miR319 gene alters plant development and enhances salt and drought tolerance in transgenic creeping bentgrass. *Plant Physiol.* 161, 1375–1391. doi: 10.1104/pp.112.208702

Zhu, H., and Guo, H. (2012). The role of virus-derived small interfering RNAs in RNA silencing in plants. *Sci. China Life Sci.* 55, 119–125. doi: 10.1007/s11427-012-4281-3

Zhu, J., Gaiha, G. D., John, S. P., Pertel, T., Chin, C. R., Gao, G., et al. (2012). Reactivation of latent HIV-1 by inhibition of BRD4. *Cell Rep.* 2, 807–816. doi: 10.1016/j.celrep.2012.09.008

Zhu, Y., Rowley, M. J., Bohmdorfer, G., and Wierzbicki, A. T. (2013). A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol. Cell* 49, 298–309. doi: 10.1016/j.molcel.2012.11.011

Zhuang, Y., Zhou, X. H., and Liu, J. (2014). Conserved miRNAs and their response to salt stress in wild Eggplant *Solanum linnaeanum* roots. *Int. J. Mol. Sci.* 15, 839–849. doi: 10.3390/ijms15010839

Zucchi, F. C., Yao, Y., and Metz, G. A. (2012). The secret language of destiny: stress imprinting and transgenerational origins of disease. *Front. Genet.* 3:96. doi: 10.3389/fgene.2012.00096

Zvereva, A. S., and Pooggin, M. M. (2012). Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* 4, 2578–2597. doi: 10.3390/v4112578

Examination of exhaustive cloning attempts reveals that *C. elegans* piRNAs, transposons, and repeat sequences are efficiently cloned in yeast, but not in bacteria

*Or Sagy[1], Ron Shamir[2] and Oded Rechavi[1]**

[1] Department of Neurobiology, Wise Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
[2] Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

*Correspondence:
Oded Rechavi, Department of Neurobiology, Wise Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel
e-mail: odedrechavi@gmail.com

Genome sequencing requires insertion of random fragments of the sequenced organism's DNA into a unicellular host, most often *Escherichia coli* bacteria. This manipulation was found in the past to be analogous to naturally occurring horizontal gene transfer, and moreover has proved valuable to understanding toxicity of foreign genetic elements to *E. coli*. Sequencing of the *Caenorhabditis elegans* genome was similarly achieved via DNA transformation into *E. coli*. However, numerous attempts have proven a significant percentage of the genome unclonable using bacteria, although clonable via yeast. We examined the genomic segments that were not clonable in bacteria but were clonable in yeast, and observed that, in line with previous hypotheses, such sequences are more repetitive on average compared with the entire *C. elegans* genome. In addition, we found that these gap-sequences encode significantly more for DNA transposons. Surprisingly, we discovered that although the vast majority of the *C. elegans* genome is clonable in bacteria (77.5%), almost all the thousands of sequences that encode for PIWI-interacting small RNAs, or 21U-RNAs (91.6%) were only clonable in yeast. These results might help understanding why most piRNAs in *C. elegans* are physically clustered on particular loci on chromosome IV. In worms and in a large number of other organisms, piRNAs serve to distinguish "Self" from "Non-Self" sequences, and thus to protect the integrity of the genome against foreign genetic elements, such as transposons. We discuss the possible implications of these discoveries.

Keywords: *C. elegans*, bacteria, lateral gene transfer, piRNAs, PIWI-interacting small RNAs, repetitive elements

INTRODUCTION

During the 1990s and early 2000s, there was a race to sequence the human genome (Collins et al., 2003). The required techniques were developed in a rapid pace, and in parallel utilized for sequencing of other multicellular organism's genomes. The genome of the nematode *Caenorhabditis elegans* (*C. elegans*) was the first one completed (The *C. elegans* Sequencing Consortium, 1998). An unexpected result of the development of the different DNA cloning techniques was the accumulation of negative results, failed cloning attempts that allow gaining insight regarding barriers for genomic information transfer between organisms.

As part of sequencing the *C. elegans* genome, the whole genome of the worm was randomly broken into overlapping fragments, which were transformed into *Escherichia coli* bacteria through the use of very large cloning vectors termed cosmids and fosmids (Coulson et al., 1986; Kim et al., 1992; Perkins et al., 2005). With sufficiently high coverage, sequencing of multiple overlapping fragments should allow in theory the assembly of the corresponding full genome. During the project, researchers soon realized that certain parts of the worm's genome could not be cloned in the bacteria, leaving gaps in the resultant genome (Coulson et al., 1991). Even though there have been numerous attempts over the past 20 years at filling the gaps using cosmids and fosmids [including a

consortium dedicated to creating a library of fosmids that would cover the entire genome (Perkins et al., 2005)], none have been successful in covering these gaps. In total, about 20% of the genome could not be cloned in bacteria in spite of these repeated efforts. Throughout this manuscript we will refer to such sequences as "gap sequences."

In the 1990s, in order to close the gaps and obtain a complete sequence of the *C. elegans* genome, researchers turned to cloning in yeast, by using YACs [Yeast Artificial Chromosomes (Coulson et al., 1991)]. The gap sequences quickly proved clonable in yeast (*Saccharomyces cerevisiae*), and thus YACs that contain sequences covering the gaps were rapidly obtained and facilitated the completion of the worm genome (Wilson, 1999). Even so, the reason for the "non-clonability" of these gaps in bacteria largely remained a mystery.

We found this incompatibility of worm genome fragments with bacterial cloning a valuable resource. *C. elegans* nematodes feed on bacteria both in real-life and in lab conditions, and this close contact facilitates exchange of genetic information between the nematodes and their resident microbes. For example, bacterially-expressed double-stranded RNA transfers from the bacteria to the worm and silences endogenous *C. elegans* genes (Timmons et al., 2001; Mello and Conte, 2004; Liu

et al., 2012; Sarkies and Miska, 2013). Although lateral gene transfer has been most extensively described in microorganisms, recent studies suggest that the process occurs more frequently than previously appreciated in eukaryotes as well (Bruto et al., 2014). DNA sequences were also shown to be laterally exchanged between nematodes and symbionts (Haegeman et al., 2011). This is consistent with the various findings of gene exchange between organisms at all evolutionary distances (Boto, 2010; Koonin, 2010). Thus, a side benefit of the artificial cloning procedure used for sequencing the *C. elegans* genome is that it provides a keyhole into limitations of natural genome transfer between worm and bacteria.

Given that the aforementioned gaps have not been clonable in *E. coli*, a prokaryote, but have been clonable in the eukaryote *S. cerevisiae*, we hypothesized that there would be a genetic factor in these regions causing the adopting prokaryotic host to die. This notion was inspired by Sorek et al.'s (2007) experimental research on inhibition of gene transfer between different bacterial genomes to *E. coli*. Such analysis of clone insertions and cloning gaps has proved valuable to understanding toxicity of bacterially derived genetic elements to *E. coli* in the past (Sberro et al., 2013).

It has been believed that the cause of the lack of clonability of various regions of the *C. elegans* genome was their high degree of repetitiveness (Wilson, 1999). We set out to explore whether there was another underlying property of these regions that, possibly in conjunction with repetitiveness, was actively inhibiting the prokaryotic hosts of these genome sequences from thriving.

MATERIALS AND METHODS

PRECISE ALIGNMENT OF THE CLONES TO THE GENOME

The sequences of *C. elegans* clones were downloaded from NCBI BioProject PRJNA13758, and the precise location of each of the clones was determined by BLASTing the sequence to the *C. elegans* genome (version WBcel235; this step was required as the locations in the NCBI record did not fit the genome version that we utilized). We used clone sequences from WormBase, retrieved from NCBI on April 4th, 2013. The sequences were categorized as "Cosmid," "Fosmid," or "YAC" clones.

The YAC subsequences that we analyzed tended to bridge across gap sequences, with both YAC ends occasionally overlapping with finished cosmid/fosmid sequences. The overlap varied between 0% and 100% (3 YAC sequences were found to be covered fully by cosmids/fosmids – Y119C1C, Y110A2B, Y70C5B). On average ~7% of the YAC sequence's ends overlapped with cosmids/fosmids. To prevent this overlap from influencing our results, we removed the overlapping segments from the YAC sequences before running any analyses. The resulting "YAC-exclusive" sequences were used in our study (hereby referred to as "YACs").

REPEATMASKER ANALYSIS

As a first step, and in accordance with previous notions of repetitiveness as the factor limiting cloning of the YAC sequences, we ran RepeatMasker separately on YACs, cosmids, and fosmids. This was conducted for each of the individual clones. We compared the average, median, minimal, maximal, and non-zero results of each RepeatMasker parameter for every clone type (RepeatMasker was run with the species set as "*C. elegans*"). RepeatMasker lists

retroelements (including sub-classifications), DNA transposons (including sub-classifications), rolling-circles, other unclassified interspersed repeats, small RNA, satellites, simple repeats, and low complexity regions (Smit et al., 1996–2010). As a control test for each clone type we also generated a random library with the same number of clones taken from random segments of the *C. elegans* genome, (hereby "random clones"), of similar sizes as those comprising that clone type (based on the average length and length standard deviation). These libraries were examined using the same RepeatMasker analysis. This analysis was used to assess the significance of our results, in comparison to enrichments detected by chance.

COMPARISON BY GENOMIC LOCATION

It has been observed that the distribution of repeat elements along chromosomes is highly uneven (The *C. elegans* Sequencing Consortium, 1998). We therefore evaluated the difference in repeat element distributions between YACs and cosmids, by taking into account the location of the clones. The purpose of this analysis was to reduce bias caused by different location distributions of the types of clones. It may be hypothesized that these differences were causing the disparity that we found. In this, and the next couple of analyses, we compared YACs to cosmids and ignored fosmids, as there were much fewer fosmids compared to cosmids (107 vs. 2,520). To measure the effect of the clone's location along the chromosome and to obtain robust statistical analysis, we partitioned each chromosome into four equal-size quarters and summarized the statistics of the two extreme quarters together and of the two middle ones together.

COMPARISON BY CLONE LENGTH

Another factor that we took into consideration was the clone's length. Although YACs are generally much longer than cosmids, the YAC segments that were evaluated here were shorter, as they were trimmed (as mentioned above) to contain only the gaps used to bridge between cosmids. The YAC-exclusive segment median length was roughly 23.7 kb, while for the cosmids the median length was approximately 32.5 kb. In this analysis we divided each clone type into two groups – one of clones below the median length of the clone type and the other of clones over the median length. Statistics were collected for each group.

COMPARISON OF GENES ENCODED IN THE CLONES

We categorized the genes in the clones based on the type of product that they encode for (or do not encode for): coding sequences (CDS), non coding RNAs (ncRNA), tRNAs, and rRNAs. This categorization was determined based on the NCBI classification of each gene.

RESULTS

By comparing the sequences that were cloned in YACs, cosmids and fosmids, we observed, consistently with previous reports, that YAC sequences were overall more than twice as repetitive as the sequences cloned in cosmids. On average, 21.42% of the YAC nucleotides vs. 10.15% of cosmid nucleotides were masked by RepeatMasker. The most abundant repeat category, and an interesting result by itself, was enrichment in DNA transposons

of various types in the YAC sequences (See **Table 1**). On average, 11.44% of YAC sequences were found to be DNA transposons vs. 5.69% in cosmids. Of the transposons, the most major difference was in PiggyBac transposons – 1.53% in YACs on average vs. 0.48% in cosmids. Other than the transposons, significant differences were found in unclassified repeats (4.65% vs. 1.58%) and satellites (2.52% vs. 0.51%). The content of Simple repeats also varied – 1.65% vs. 1.06%. The total interspersed repeats (all the repeats that were identified by RepeatMasker, excluding small RNA, satellites, simple repeats and low complexity regions) composed, on average, 17.04% of a YAC sequence vs. 8.27% of a cosmid sequence. Contrary to what might be expected, low complexity repeats did not show a significant difference (0.31% vs. 0.3%).

Comparison with the results that were obtained with the randomly generated clones showed that the differences between the clone types did not originate from their size. For all repeat types the percentage of sequence they covered in the "random YACs" was very similar to the percentages for "random cosmids" and "random fosmids," unlike the results that were obtained with the real clones (Supplementary **Table S7**).

THE RELATIVELY HIGH REPETITIVENESS OF THE SEQUENCES IN YACs PERSISTS EVEN WHEN CONTROLLING FOR THE CHROMOSOMAL LOCATION

As previously reported, we observed a significant difference in repetitiveness and transposon-frequency between the different chromosome regions (both $p < 0.0001$, t-test). The extreme quarters of chromosomes were more repetitive than the middle quarters of the chromosomes (bases masked – 18.69% vs. 6.99%), and were made up more of transposons (10.12% vs. 3.96%). We also found a noteworthy difference between chromosomal location distributions of the clone types. The vast majority of the YACs, a little over three quarters (396) were in the extreme quarters and the minority in the middle quarters (119), while the cosmids were mostly in the middle quarters (979 vs. 1,540 – a ratio of ~2:3). We then ran RepeatMasker while taking into account this difference – comparing the differentiated groups between YACs and cosmids and within each clone type. Histograms of the locations of YACs and cosmids and the locations of repeats of any type are shown in **Figures 1–2**. A similar histogram for transposons, portraying the deviation more finely can be found in **Figure 3**. To test the effect of the chromosomal location, we reran Repeat-Masker separately on the extreme and on the middle quarters. In both cases the conclusions that were drawn from the previous analysis persisted. YAC-contained sequences were more repetitive, and were composed more of transposons than cosmids, although clones in the first and last quarters were overall more repetitive than their second and third quarter counterparts (Bases masked: YACs – 22.83 and 16.71%, cosmids – 15.69 and 6.62%, composed of transposons: YACs – 12.78 and 7.01%, cosmids – 8.49 and 3.91%).

DIFFERENCES IN THE LENGTH OF YACs AND COSMIDS DO NOT ACCOUNT FOR THEIR REPETITIVENESS DIFFERENCE

There were no outstanding differences in repeat frequencies between the clones from the same type above and below the median length. Hence, the differences in repeat frequency between the clone types could not be explained by length variation.

18 YACs HAD NO REPETITIVE ELEMENTS

We examined the 18 YACs for which RepeatMasker did not find any repetitive elements. Out of these YACs, after removing overlaps with cosmids and fosmids, 12 were found to be smaller than 1 kb in length, 3 between 1 and 1.5 kb, and only 3 over 1.5 kb in length – Y68G5A (3.4 kb, after removing a 0.2 kb overlap; no genes in the non-overlapping portion), Y2C2A (8.5 kb, after removing an 8.5 kb overlap; part of the gene tag-80 in the non-overlapping portion) and Y53C12C (23.4 kb, after removing a 0.2 kb overlap; the gene *eyg-1* is enclosed in the non-overlapping portion).

YACs ARE SIGNIFICANTLY ENRICHED FOR 21U piRNAs

An outstanding difference between the densities of protein-coding genes in YACs and cosmids was not found (a slightly higher density was found in cosmids: 182 genes per MB in YACs vs. 238 genes per MB in cosmids). tRNAs and rRNA are overrepresented in YACs – 94 YAC tRNA genes vs. 114 cosmid tRNA genes (4.2 vs. 1.5 genes per MB), and 10 YAC rRNA genes vs. 12 cosmid rRNA genes (0.45 vs. 0.15 genes per MB). Results for the CDS and ncRNA genes can be found in **Table 2**.

The most striking result of our analysis was that ncRNAs of the type 21U piRNAs were dramatically overrepresented in YACs. A roughly similar absolute amount of ncRNA was found in YACs and cosmids despite the fact that YACs take up about a quarter of the length that cosmids take in the genome (10,799 vs. 11,108 ncRNA genes, respectively – corresponding with densities of 484 vs. 143 ncRNAs/MB). Further investigation revealed that the vast majority of the YAC ncRNAs are PIWI-interacting small RNAs, or 21U-RNAs (about 91%), while they are less than half of the cosmid ncRNAs (about 42.4%). Moreover, in terms of absolute quantities, YACs include more than twice as many 21U-RNAs as cosmids – 4,713 vs. 9,879 flip order to maintain consistency.

The majority of the DNA sequences that give rise to 21U piR-NAs reside in two regions in chromosome IV (and also in another small region on chromosome IV) that are called "piRNA clusters" (Ruby et al., 2006). We analyzed these regions for repetitiveness and they turned out to be overall less repetitive compared to other regions, both on chromosome IV and on other chromosomes (the two major piRNA clusters had 8.15 and 11.51% of the bases masked as repetitive and the third cluster had 8.11% of the bases masked, compared to 11.89% for chromosome IV as a whole and 14.9% for chromosome I for example). Since the regions that encode for abundant YAC-specific piRNAs are relatively not repetitive, it raises the possibility that they are unclonable in *E. coli* for other reasons, perhaps due to toxicity (see Discussion). The longest four YACs, and seven of the longest ten YACs, were on chromosome IV (see **Table 3**). The longest four YACs alone hold 3,754 ncRNAs (roughly 35% of all ncRNA genes found in YACs), with the 6th longest holding 1,083 more. The longest 4 YACs also mapped specifically to one of the two piRNA clusters on chromosome IV.

Table 1 | RepeatMasker results for YACs, cosmids, and fosmids (average per clone for each clone type).

|	YACs			Cosmids			Fosmids		
Number of clones	515			2,520			107		
Total coverage (bp)	22,327,861			77,668,781			1,885,889		
Approximate portion of genome covered	22%			77%			2%		
Bases masked	21.42%			10.14%			12.14%		
	Number of elements	Length occupied (bp)	Percentage of sequence (%)	Number of elements	Length occupied (bp)	Percentage of sequence (%)	Number of elements	Length occupied (bp)	Percentage of sequence (%)
Retroelements	0.65	370.6	0.94	0.47	276.12	1.01	0.27	159.37	1.71
DNA transposons	22.48	4982.2	11.44	8.73	1743.45	5.69	5.75	1252.3	6.43
Unclassified	9.31	2306.1	4.65	2.57	510.22	1.58	1.99	469.55	1.97
Total interspersed repeats	32.44	7659	17.04	11.77	2529.79	8.27	8.01	1881.22	10.11
Small RNA	0.15	11.2	0.03	0.19	16.98	0.06	0.08	6.59	0.06
Satellites	3.11	942.26	2.52	0.93	156.76	0.51	0.75	143.05	0.74
Simple repeats	11.34	729.05	1.65	6.49	324.9	1.06	3.8	186.04	1.03
Low complexity	2.83	140.12	0.31	1.93	92.55	0.3	1.19	54.61	0.3
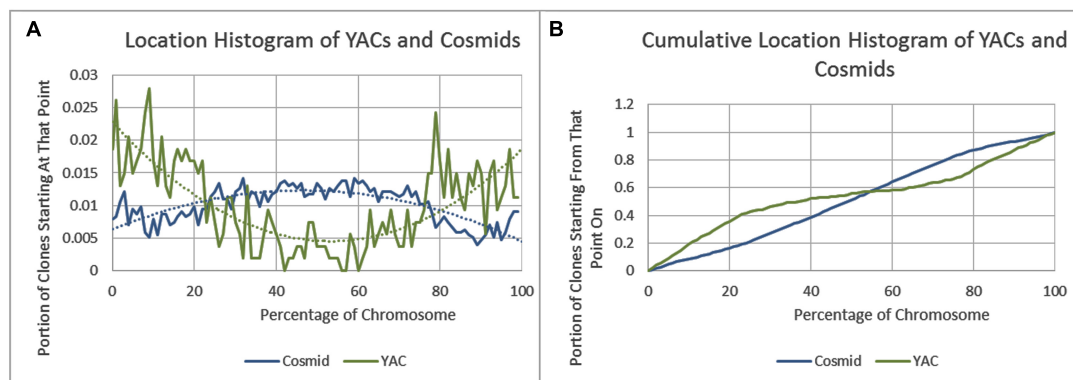
FIGURE 1 | Cosmid and YAC locations. (A) The fraction of cosmids and YACs starting at each percent point of their chromosomes on average. **(B)** The cumulative probability distributions of cosmids and YACs on average. For each percentage point the fraction of clones starting at that point or further is given.
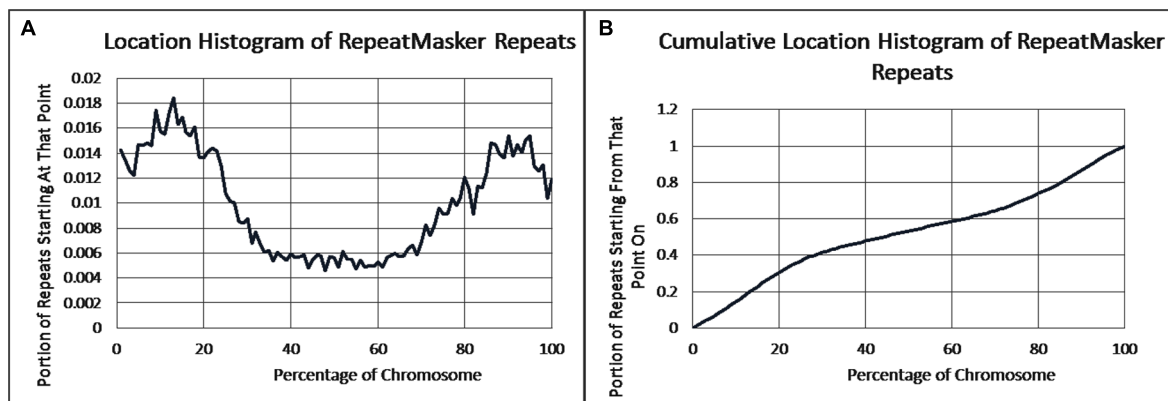
FIGURE 2 | Repeat locations. (A) The fraction of repeats starting at each percent point of their chromosomes on average, as identified by RepeatMasker. **(B)** The cumulative probability distributions of repeats on average. For each percentage point the fraction of clones starting at that point or further is given.
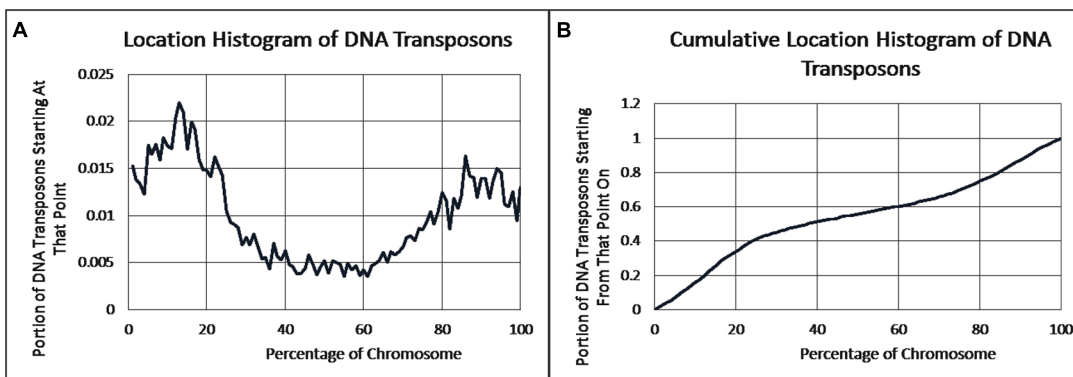
FIGURE 3 | DNA transposon locations. (A) The fraction of DNA transposons starting at each percent point of their chromosomes on average, as identified by RepeatMasker. **(B)** The cumulative probability distributions of DNA transposons on average. For each percentage point the fraction of clones starting at that point or further is given.

Table 2 | Number and length statistics of protein-coding genes and ncRNAs in cosmids, fosmids and exclusive to YACs.

Clone and gene type	Number	Average length	Median length	Length SD
Cosmid genes	18,071	1,119	948	1,090
Cosmid ncRNAs	11,093	102	83	138
Fosmid genes	325	1,193	999	1,143
Fosmid ncRNAs	817	43	21	59
YAC genes	3,882	1,058	894	913
YAC ncRNAs	10,728	31	21	58

Table 3 | Longest 15 YACs, their locations, and lengths.

YAC Segment	Chromosome	Start position	End position	Length (kb)
Y73F8A	IV	15,226,338	15,549,111	323
Y105C5B	IV	15,863,724	16,185,028	321
Y105C5A	IV	15,548,990	15,863,827	315
Y57G11C	IV	14,637,373	14,950,845	313
Y75B8A	III	12,069,105	12,367,511	298
Y54G2A	IV	2,750,222	3,036,901	287
Y105E8A	I	14,333,265	14,610,766	278
Y51H4A	IV	16,471,836	16,741,455	270
Y116A8C	IV	16,900,192	17,160,330	260
Y111B2A	III	12,492,890	12,750,488	258
Y39B6A	V	18,958,057	19,204,189	246
Y53F4B	II	14,951,249	15,178,754	228
Y73B6BL	IV	6,286,329	6,502,271	216
Y47D3A	III	11,136,151	11,335,864	200
Y46G5A	II	12,680,715	12,878,522	198

While chromosome IV contains the vast majority of the YAC ncRNAs (93.3% vs. 44.1% of cosmid ncRNAs), it is not over-represented for YACs overall (21.8% of the YAC exclusive bases are on chromosome IV, which constitutes 17.4% of the nematode genome). When removing chromosome IV from the analysis, the discrepancy between YACs and cosmids on ncRNA content was flipped: instead of the ncRNAs being divided roughly equally between YACs and cosmids in terms of absolute numbers, when excluding chromosome IV, there were about 8.8 as many ncRNAs in cosmids as in YACs (6,182 vs. 702).

DISCUSSION

We used a bioinformatic approach to probe DNA sequences that could be cloned in a eukaryote (yeast), but not in a prokaryote (bacteria). We found several major characteristics in such sequences, including high repetitiveness and enrichment for DNA

transposons and 21U piRNAs. If these characteristics are typical to sequences of other organisms, it could in theory shed light on the barriers to genetic information transfer that exist between prokaryotes and eukaryotes, and should be interesting to validate experimentally.

The inability to efficiently clone repetitive sequences in bacteria is perhaps not biologically relevant, as it could simply stem from technical considerations. On the other hand, and while our results do not prove it in any way, it is possible that the inability to clone DNA transposons or piRNAs in bacteria has physiological importance.

Acquisition of vectors (cosmids or fosmids in this case) that carry mobile elements could be detrimental, if such genomic parasites are allowed to colonize or disrupt the genome of the host. It would be interesting to examine whether the type of transposons that we found to be enriched in YACs "jump" in *E. coli*, but not in *S. cerevisiae*, and if such transposition, should it be shown to occur, compromises the bacteria's viability. Transposons that are not dependent on host factors could in theory transpose and confer damage, even upon lateral gene transfer.

piRNA-mediated RNA interference has a role in genome surveillance against foreign sequences in multiple organisms across the tree of life (animals and protists). One fascinating scenario that could be envisioned based on our data is that piRNAs might act *in trans*, to eliminate bacteria, *E. coli* in this case, with which *C. elegans* intimately interacts. Although *C. elegans* worms are not natural hosts of *E. coli*, the two organisms have been grown together in the lab for half a century. *C. elegans* and *E. coli* could in theory adapt to interact by these mechanisms, as piRNAs and piRNA pathway genes evolve rapidly, due to an "arms race" with transposons (Yi et al., 2014), and since piRNAs target rapidly evolving genes (Weick et al., 2014). The promoters of most piRNAs, which are autonomous transcriptional units in *C. elegans*, contain a consensus motif for binding of transcription factors (Weick et al., 2014). Thus in theory, possession of just one appropriate transcription factor could allow a host (such as *E. coli*) to transcribe most piRNAs. In flies, piRNA clusters act as "traps" in which transposons land (Malone and Hannon, 2009). It is not understood why most piRNAs in *C. elegans* are clustered together on chromosome IV. Perhaps the inability to transfer a large cluster of piRNAs from *C. elegans* to *E. coli* is one of the reasons.

Characterization of DNA sequences that are lethal to bacteria but not to yeast could be applicative (Sorek et al., 2007; Kimelman et al., 2012). Given further research and validation, it could be possible in the future to combine in gene therapy, for example, piRNA-encoding DNA sequences, and thus to affect bacteria, while sparing the eukaryote hosts with which they interact.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00275/abstract

Table S1 | RepeatMasker results for the various groups of clones, averaged per clone for each clone type.

Table S2 | RepeatMasker results for the various groups of clones, median per clone for each clone type.

Table S3 | RepeatMasker results for the various groups of clones, summed per clone for each clone type.

Table S4 | RepeatMasker results for the various groups of clones, number of clones for each clone type that are non-zero.

Table S5 | RepeatMasker results for the various groups of clones, the minimal result per clone for each clone type and the clones with that value.

Table S6 | RepeatMasker results for the various groups of clones, the maximal result per clone for each clone type and the clones with that value.

Table S7 | RepeatMasker results for the randomly-selected genomic segments with the same size distributions as YACs, cosmids, and fosmids (averaged per segment for each clone type).

REFERENCES

Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* 277, 819–827. doi: 10.1098/rspb.2009.1679

Bruto, M., Prigent-Combaret, C., Luis, P., Moënne-Loccoz, Y., and Muller, D. (2014). Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proc. Biol. Sci.* 281, 1471–2954. doi: 10.1098/rspb.2014.0848

Collins, F. S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science* 300, 286–290. doi: 10.1126/science.1084564

Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J., and Waterston, R. (1991). YACs and the C. *elegans* genome. *Bioessays* 13, 413–417. doi: 10.1002/bies.950130809

Coulson, A., Sulston, J., Brenner, S., and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* 83, 7821–7825. doi: 10.1073/pnas.83.20.7821

Haegeman, A., Jones, J. T., and Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol. Plant Microbe Interact.* 24, 879–887. doi: 10.1094/MPMI-03-11-0055

Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B., and Simon, M. I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* 20, 1083–1085. doi: 10.1093/nar/20.5.1083

Kimelman A., Levy A., Sberro H., Kidron S., Leavitt A., Amitai G., et al. (2012). A vast collection of microbial genes that are toxic to bacteria. *Genome Res.* 22, 802–809. doi: 10.1101/gr.133850.111

Koonin, E. V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11, 209. doi: 10.1186/gb-2010-11-5-209

Liu, H., Wang, X., Wang, H.-D., Wu, J., Ren, J., Meng, L., et al. (2012). Escherichia coli noncoding RNAs can affect gene expression and physiology of *Caenorhabditis elegans*. *Nat. Commun.* 3, 1073. doi: 10.1038/ncomms2071

Malone, C. D., and Hannon, G. J. (2009). Small RNAs as guardians of the genome. *Cell.* 136, 656–668. doi: 10.1016/j.cell.2009.01.045

Mello, C. C., and Conte, D. (2004). Revealing the world of RNA interference. *Nature* 431, 338–342. doi: 10.1038/nature02872

Perkins, J., Wong, K., Warren, R., Schein, J., Stott, J., Holt, R., et al. (2005). A *Caenorhabditis elegans* fosmid library. *Int. Worm Meet.*

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., et al. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. *elegans. Cell* 127, 1193–1207. doi: 10.1016/j.cell.2006.10.040

Sarkies, P., and Miska, E. A. (2013). Is there social RNA? *Science* 341, 467–468. doi: 10.1126/science.1243175 doi: 10.1126/science.1243175

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., and Sorek, R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol. Cell.* 50, 136–148. doi: 10.1016/j.molcel.2013.02.002

Smit, A., Hubley, R., Green, P., RepeatMasker Open-3.0. (1996–2010). Available at from: www.repeatmasker.org

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. doi: 10.1126/science.1147112

The C. *elegans* Sequencing Consortium. (1998). Genome sequence of the nematode C. *elegans*: a platform for investigating biology. *Science* 282, 2012–2018. doi: 10.1126/science.282.5396.2012

Timmons, L., Court, D. L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans. Gene* 263, 103–112. doi: 10.1016/S0378-1119(00)00579-5

Weick, E. M., Sarkies, P., Silva, N., Chen, R. A., Moss, S. M., Cording, A. C., et al. (2014). PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in C. *elegans. Genes Dev.* 28, 783–796. doi: 10.1101/gad.238105.114

Wilson, R. K. (1999). How the worm was won: the C. *elegans* genome sequencing project. *Trends Genet.* 15, 51–58. doi: 10.1016/S0168-9525(98)01666-7

Yi, M., Chen, F., Luo, M., Cheng, Y., Zhao, H., Cheng, H., et al. (2014). Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. *Genome Biol. Evol.* 6, 1393–1407. doi: 10.1093/gbe/evu105

Insights into the epigenomic landscape of the human malaria vector *Anopheles gambiae*

Elena Gómez-Díaz[1]*, Ana Rivero[2], Fabrice Chandre[2] and Victor G. Corces[1]

[1] Department of Biology, Emory University, Atlanta, GA, USA
[2] Maladies Infectieuses et Vecteurs: Écologie, Génétique, Évolution et Contrôle (UM1-UM2-CNRS 5290-IRD 224), Centre IRD, Montpellier, France

The epigenome of the human malaria vector *Anopheles gambiae* was characterized in midgut cells by mapping the distribution and levels of two post-translational histone modifications, H3K27ac and H3K27me3. These histone profiles were then correlated with levels of gene expression obtained by RNA-seq. Analysis of the transcriptome of *A. gambiae* midguts and salivary glands led to the discovery of 13,898 new transcripts not present in the most recent genome assembly. A subset of these transcripts is differentially expressed between midgut and salivary glands. The enrichment profiles of H3K27ac and H3K27me3 are mutually exclusive and associate with high and low levels of transcription, respectively. This distribution agrees with previous findings in *Drosophila* showing association of these two histone modifications with either active or inactive transcriptional states, including Polycomb-associated domains in silenced genes. This study provides a mosquito epigenomics platform for future comparative studies in other mosquito species, opening future investigations into the role of epigenetic processes in vector-borne systems of medical and economic importance.

Keywords: H3K27ac, H3K27me3, histone post-translational modifications, chromatin immunoprecipitation, gene expression regulation, mosquito-borne diseases, ChIP-seq, RNA-Seq

INTRODUCTION

Host-parasite interactions are among the most plastic and dynamic systems in nature. In these systems, epigenetic modifications can provide an accessory source of fast-acting, reversible and readily available phenotypic variation that can be directly shaped by both host and parasite selection pressures (Bonduriansky and Day, 2009; Gómez-Díaz et al., 2012). Mosquitoes are important disease vectors worldwide. *Anopheles gambiae* is major vector of malaria in Africa, a disease that affects more than 300 million people and causes around 650,000 deaths each year (Who, 2013). The genome of *A. gambiae* was sequenced more than 10 years ago (Holt et al., 2002). Since then, the roadmap for malaria control strategies has been mostly centered on the study of vector genetic variation (Severson and Behura, 2012). Yet, genetic mechanisms alone are not sufficient to explain natural phenotypic variation in terms of vector competence (Lambrechts, 2010). Thus, it is critical to expand our current understanding of mosquito-parasite interactions into an integrated view that includes both genetic and epigenetic dimensions. Although a great deal of progress has been made in deciphering the epigenetic code of *Plasmodium* parasites (see Hoeijmakers et al., 2012 for a review), knowledge of epigenetic processes in mosquitoes is limited.

Post-translational modification (PTM) of histones (acetylation, methylation, phosphorylation, sumoylation, and ubiquitinylation) is an important regulatory mechanism of transcriptional control that acts by altering chromatin structure (Kouzarides, 2007; Bonasio et al., 2010). In *Drosophila*, epigenomic profiling of enrichment or depletion patterns for specific histone modifications has been established genome-wide and this information has been correlated with the location of regulatory elements and functional domains (Kharchenko et al., 2011; Negre et al., 2011). For example H3K4me3 and H3K27ac are associated with active promoters, H3K4me1 and H3K27ac are present in active enhancers, and H3K27me3 associates with Polycomb-mediated silencing. Surprisingly, no information is available in mosquitoes on histone PTMs or regulatory elements involved in chromatin-based epigenetic processes, such as the Polycomb and trithorax complexes, insulators, and enhancers.

Available technologies now allow the analysis of whole epigenomes in a straightforward and cost-effective manner. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) allows the mapping of protein-DNA interactions genome-wide. In this method DNA-protein complexes containing a specific protein of interest are immunoprecipitated from cross-linked, sonicated chromatin. DNA is purified from the enriched pool and used to generate a library for subsequent whole genome sequencing. The sequence reads for the enriched DNA are computationally aligned to the reference genome to define sharp peaks or broad blocks of modified histones. Since its development in 2007 (Robertson et al., 2007), ChIP-seq has been extensively used to survey the genomic profiles of histones and their modifications, transcription factors (TFs), DNA and histone modifying enzymes, the transcriptional machinery, and other chromatin associated proteins in various model organisms, including the fruit fly *Drosophila melanogaster* (Roy et al., 2010). One downside of this technique is, however, that its application in non-model

species is still limited by the availability of reference genomes. On the other hand, in cases where the annotation of the genome is still incomplete, such as *A. gambiae*, the combination of histone modification maps and gene expression information constitutes a powerful approach that allows the functional interpretation of chromatin marks as well as *de novo* prediction of regulatory elements, genes and splice variants (Ernst and Kellis, 2010; Cheng et al., 2011). Previous studies have analyzed the transcriptome of *A. gambiae* using microarrays (Dana et al., 2005; Marinotti et al., 2005; Koutsos et al., 2007; Baker et al., 2011; Maccallum et al., 2011), but this approach gives only semi-quantitative results due to the small dynamic range of the microarray signal. A few studies have applied RNA-seq in this species on whole bodies or chemosensory appendages (Pitts et al., 2011; Vannini et al., 2014). To our knowledge no study has specifically targeted RNA-seq analyses to tissues such as the midgut or salivary glands that play a key role in the development of *Plasmodium* parasites within the mosquito, and are thus directly implicated in malaria transmission.

Using the technical and theoretical knowledge accumulated from chromatin studies in *Drosophila*, here we carry out an analysis of the transcriptome and histone PTMs in the human malaria vector *A. gambiae*. We characterize the distribution and levels of H3K27ac and H3K27me3, two key histone modifications that have been shown to play an important role in gene regulation, by performing ChIP-seq analyses in the midguts of adult mosquito females. The midgut is key tissue for *Plasmodium* development because the obligate passage of the parasite through this tissue results in large losses in parasite numbers, which may explain the frequent failure of the parasite to complete its life cycle in the mosquito (Blandin et al., 2008). We then correlate the histone profiles with levels of genome-wide gene expression obtained by RNA-seq, in order to infer functional states and predict putative regulatory elements in the mosquito. This integrative analysis allowed us to link enrichment or depletion of active and repressive histone modifications to their target genes. The result is the first platform for mosquito comparative epigenomics that can serve as a basis for future studies on the biology of mosquitoes and mosquito-borne diseases.

EXPERIMENTAL PROCEDURES
MOSQUITO REARING AND DISSECTION
Experiments were performed using an isogenic strain of *A. gambiae* (strain Kisumu) maintained at the MIVEGEC insectarium under standard rearing conditions ($27 \pm 1°C$, $70 \pm 10\%$ RH and 16L: 8D photoperiod). Blood-fed adult females were allowed to lay eggs. On the day of hatching, larvae were seeded into plastic trays ($25 \times 35 \times 7$ cm) containing one liter of mineral water at a density of 300 individuals per tray. Larvae were provided with 200 mg of TetraMin® fish flakes the day after hatching, and, from then on, 400 mg TetraMin every 2 days until pupation. On pupation, trays were placed inside an emergence cage ($27 \times 40 \times 35$ cm) and provided with an *ad libitum* source of 10% sugar solution for the emerged adults. Dissection of the midguts and the salivary glands was performed on adult females aged 6–8 days. Tissues were maintained in ice-cold Schneider's insect culture medium (Sigma-Aldrich) to avoid degradation, and

fresh tissues were immediately processed for chromatin and RNA analyses.

RNA ISOLATION AND RNA-SEQ LIBRARY PREPARATION
Total RNA was extracted from fresh mosquito tissues ($\sim$25 midguts and $\sim$ 50 salivary glands) using the *mir*Vana™ RNA Isolation Kit (Ambion®) according to the manufacturer's protocol. RNA concentration was quantified using a Qubit® 2.0 Fluorometer, and RNA integrity was determined with an Agilent 2100 Bioanalyzer. Illumina libraries were prepared and sequenced at the HudsonAlpha Institute for Biotechnology, using an Illumina HiSeq2000 sequencer. Standard directional RNA-seq library construction, 50 bp paired end reads with ribosomal reduction (RiboMinus™ Eukaryote Kit, Ambion®).

CHROMATIN ISOLATION, IMMUNOPRECIPITATION, AND SEQUENCING
Antibodies to histone modifications H3K27ac (Abcam ab4729) and H3K27me3 (Millipore 07-449) used in this study have been previously tested and shown to specifically recognize the appropriate epitopes (Kharchenko et al., 2011; Landt et al., 2012). The epitope sequences are conserved between *Drosophila* and *A. gambiae*.

For immunoprecipitation experiments, mosquito tissues (15–20 midguts) were suspended in 100 μl of Schneider's medium, fixed by adding 37% formaldehyde to a final concentration of 1%, and incubated at room temperature for 10 min. The reaction was stopped by adding 1/10 volumes of 1.25 M glycine solution, incubated for 5 min before centrifugation at 4000 rpm for 3 min, and washed twice using cold 1X PBS. Tissues were lysed in 200 μl of cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40), and protease inhibitor cocktail (complete protease inhibitor cocktail tablets, Roche), then homogenized with a pestle mixer and incubated on ice for 15 min. After centrifugation at 4000 rpm for 8 min, the nuclei were resuspended in 200 μl of nuclear lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA.Na2, 1% SDS, and protease inhibitors), and incubated on ice for 20 min. Then, 100 μl of IP dilution buffer (0.01% SDS, 1.1% Trition X-100, 1.2 mM EDTA.Na2, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl) were added. Sonication was performed using a Diagenode BioRuptor using 40 cycles of 30 s ON/ 30 s OFF, resulting in fragments around 300–500 bp, followed by centrifugation at 4°C for 10 min. Quality control of fixed chromatin was performed by visualization of seared fragments by agarose electrophoresis prior to proceeding with the chromatin IP. For the preclear step, 50 μl of protein A magnetic beads (Dynabeads® Protein A, Novex®) were pre-coated using anti-rabbit IgG in 5% BSA. Beads were then added to each sample diluted 5X in IP dilution buffer, and incubated overnight (OVN) at 4°C in a rocker. Seventy five μl (5%) were taken as a control input from this solution and subjected to the same experimental procedure than the test sample, except for the antibody-binding step. For the test sample, 3 μl of Ab (1 μg/μl) pre-bound to 50 μl of protein A beads were added to the pre-cleared chromatin and incubated OVN at 4°C. After discarding the supernatant by magnetic separation, antibody-bound chromatin was subjected to 3 washes in low salt buffer (0.1% SDS, 1% Trition X-100, 2 mM EDTA.Na2 pH 8.0, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), 2 washes in high

salt buffer (0.1% SDS, 1% Trition X-100, 2 mM EDTA.Na2 pH 8.0, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), 2 washes in LiCl buffer (1 mM EDTA.Na2 pH 8.0, 10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 1% NP40, 1% SDS), and one final wash in TE. Elutions were then performed by adding $2 \times 200\,\mu l$ of IP Elution Buffer (0.1 M NaHCO$_3$, 1% SDS). Crosslinks were reversed by incubating the ligated chromatin at 65°C OVN with the addition of 0.25 M NaCl, 10 mM EDTA, and 40 mM Tris. Extraction was performed by adding 8 μl of proteinase K (10 mg/ml) and incubation at 50°C for 2 h. The DNA was extracted following a standard phenol-chloroform procedure.

Sequencing libraries were prepared following the protocol described by Bowman et al. (2013) optimized for low quantity DNA. Before library preparation, ChIP DNA was pre-cleaned using a $1.8 \times$ ratio of SPRI beads (Agencourt AMPure XP, Beckman Coulter). End preparation was performed by end repair (End-It DNA End Repair Kit, Epicenter Cat# ER0720) and addition of "A" base to 3′ ends (Klenow 3′-5′exo, NEB Cat# M0212S). Adaptor ligation was performed using a universal adaptor sequence and T4 ligase (NEB Cat# M0202S), and the ligated products were SPRI cleaned ($1.6 \times$ beads ratio). Amplification of ChIP libraries was performed by qPCR using an Applied Biosystems 7500 Fast Real-Time PCR System. Reactions (50 μl) consisted of 0.2 μM universal primer, 0.2 μM barcoded primer and 25 μl of KAPA SYBR FAST qPCR Master Mix (KapaBiosystems). Real time kinetics was monitored and library amplifications were stopped at the end of the extension after SYBR green reported reaction kinetics in the log phase (usually 15 cycles). A final SPRI purification with $1.0 \times$ beads ratio was performed on the ChIP DNA libraries to remove adapter/primer dimers (120 bp) from the rest of the library (250 bp and above). ChIP libraries were sequenced at the HudsonAlpha Institute for Biotechnology using an Illumina HiSeq2000 sequencer.

DATA ANALYSES
RNA-seq data analysis
RNA raw sequencing data were quality checked using FastQC, and trimmed on both ends, based on the quality estimates for each sequence, by using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). For the analysis of RNA paired directional reads we used the Tuxedo suite of programs as described by Trapnell et al. (2012). First, TopHat v2.0.10 (Trapnell et al., 2009) was used to map reads to the GTF annotation file of the most recent version of the *A. gambiae* genome, assembly AgamP4 PEST strain, available at VectorBase (https://www.vectorbase.org/). Reads were aligned using default parameters, except the option of library type set as first-strand for directional RNA-seq, and the inner mate distance parameter r, which was estimated using the CollectInsertSizeMetrics tool implemented in the Picard's tools suite (http://picard.sourceforge.net/index.shtml). The Cufflinks v2.1.1 package (Roberts et al., 2011), which includes Cuffcompare, Cuffmerge and Cuffdiff; was used to quantify transcript abundance in terms of Fragment Per Kilobase Million (FPKM), allowing the discovery of new transcripts. Parameters were set as default using the GTF option, bias correction, and applying the total hits norm. We then compared the new transcriptomes to existing annotated transcripts using Cuffcompare. We ran Cuffdiff first on the conserved, known transcripts, and then on the combined transcriptome, which includes both previously annotated and new transcripts, to examine tissue specific differences in gene expression between midguts and salivary glands. The clusters of differently expressed genes were visualized using the R/Bioconductor cummeRbund package (http://compbio.mit.edu/cummeRbund/). Significant expression differences between tissues was defined as having both a *P*-value < 0.01 and a false discovery rate (FDR) of <0.05.

Chip-seq data analyses
Quality metrics statistics of Illumina reads were obtained using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). We used Bowtie v.1.0.0 (Langmead and Salzberg, 2012) to map reads from control and ChIP libraries to a custom genome index for *A. gambiae*; built using the most recent AgamP4 reference assembly using the bowtie-*build* tool. We ran Bowtie using the—m option to filter repetitive sequences. To identify significantly enriched regions for H3K27ac we performed peak calling analysis using MACS v1.4.2 (Zhang et al., 2008; Feng et al., 2012). Parameters were set to consider equal numbers of unique reads for input and ChIP samples, remove redundant reads (PCR duplicates), and a *p*-value cutoff of $1 \times 10e-5$. To analyze the distribution of H3K27me3 we used SICER, which is optimized for the analysis of broad ChIP-enrichment regions (Zang et al., 2009; Xu et al., 2014) characteristic for this histone modification mark. Parameters were set to a window size of 200, gap size 600 and FDR cutoff of $1 \times 10e-5$, and the default value for the redundant rate cutoff.

We used BEDTools2.19.0 (Quinlan and Hall, 2010) to obtain genome coverage for each histone modification mark, which is the percentage of the genome in bp that shows significant enrichment of H3K27ac and H3K27me3, as well as the intersects of H3K27ac and H3K27me3 enrichment peaks with various genomic features (promoters, TSSs and genes). For these analyses we define promoters as regions located within a distance of 200 bp upstream and downstream of the transcription start site (TSS).

To examine the association between H3K27ac and H3K27me3 profiles genome-wide, we first normalized control and ChIP samples based on read counts and calculated the difference using DeepTools (Ramírez et al., in press). The resulting enrichment files were binned into 10 bp windows covering the entire *A. gambiae* genome reference sequence. We then referenced the 10 bin normalized and subtracted ChIP enrichments to gene coordinates of the AgamP4 gene set and calculated average enrichments per gene across a region that includes the gene body scaled to 500 bp, flanked by 200 bp before the TSS and after the Transcription Termination Site (TTS), respectively. The same type of analysis was performed by referencing ChIP data to gene coordinates from the "custom" gene set generated by CuffCompare, which merges novel and conserved transcripts (see above). We then applied K-means clustering to the matrix of enrichment values for H3K27me3 to organize genes based on their similarities into 3 clusters using Cluster v3.0. Genes within each cluster were ordered by decreasing mean enrichment values. Groups of H3K27me3 enrichment were then used as anchors to order the

matrix of mean values of H3K27ac and gene expression. Clustered data was then visualized as a heatmap using deepTools.

Combined analyses of ChIP and RNA-seq data were performed in R 3.1.0 (http://www.r-project.org/) using Bioconductor (http://www.bioconductor.org). We used BedTools v2.19.0 and SamTools v1.4 (http://samtools.sourceforge.net) for file manipulation and conversion. Coverage analyses on bed and bam files were performed using BedTools and DeepTools suites.

Functional analysis of genes

Functional analyses were performed on genes obtained from the clustering approach corresponding to the various enrichment profiles (see results), and that intersect with H3K27ac and H3K27me3 peaks identified by MACS or SICER. Sequences for the various gene sets were retrieved by mapping VectorBase gene IDs against the UniProtKB database (http://beta.uniprot.org/). We employed Blast2Go (http://www.blast2go.de/) to conduct Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways annotations. Fisher's exact test was used to retrieve significantly enriched GO terms for genes marked with H3K27ac and H3K27me3 using Blast2Go. GO functional categories are defined as those containing at least five genes and a minimum enrichment score of 1.3 (p-value < 0.05).

REAL-TIME PCR VALIDATION

Validation by ChIP-qPCR was performed by targeting two housekeeping genes, actin5C and TER94, which typically show significant enrichment of H3K27ac in *Drosophila* (http://www.modencode.org/). For this assay we also targeted an intergenic region located 2 Kb upstream of these two genes as a negative control. Reactions were performed using SYBR Green I Master kit (Applied Biosystems) in an Applied Biosystems Lightcycler. The PCR parameters are: 1 cycle of 10 min at 95°C, 40 cycles of 10 s at 95°C, 10 s at 60°C, and 20 s at 72°C. PCR primer sequences are listed in Table S1.

RESULTS

GENE EXPRESSION AND TRANSCRIPT DISCOVERY

RNA library processing and quality metrics are shown in **Table 1**. The expression analysis on the assembly of reads obtained from *A. gambiae* female mosquito tissues, midguts and salivary glands, resulted in a total of 13,969 novel transcripts that correspond to newly discovered non-annotated loci (class code "u": Unknown, intergenic transcript); combined GTF and BED files available at GEO GSE59773) as well as transcripts categorized as class code "=" correspond in most cases to extensions of previously annotated genes based on intron matches. Most novel transcripts (6907 in midguts and 6943 in salivary glands) show relative abundances, expressed as Fragments Per Kilobase of exon model per Million mapped fragments (FPKM), above the 95 percentile of expression for the distribution of transcript abundances, which we estimated correspond to a value of 0.26 for midguts and 0.56 for salivary glands (lower bound of the 95% confidence interval (CI) for the distribution of transcript abundance). Thus transcripts that show FPKM values above this threshold can be considered as reliable in terms of novel transcript discovery and gene expression differences. Blasting a subset of these transcripts against non-redundant protein databases confirms that they are novel (data not shown). This supports the notion that there is a large number of genes in the reference genome of *A. gambiae* that remain to be discovered and/or properly annotated.
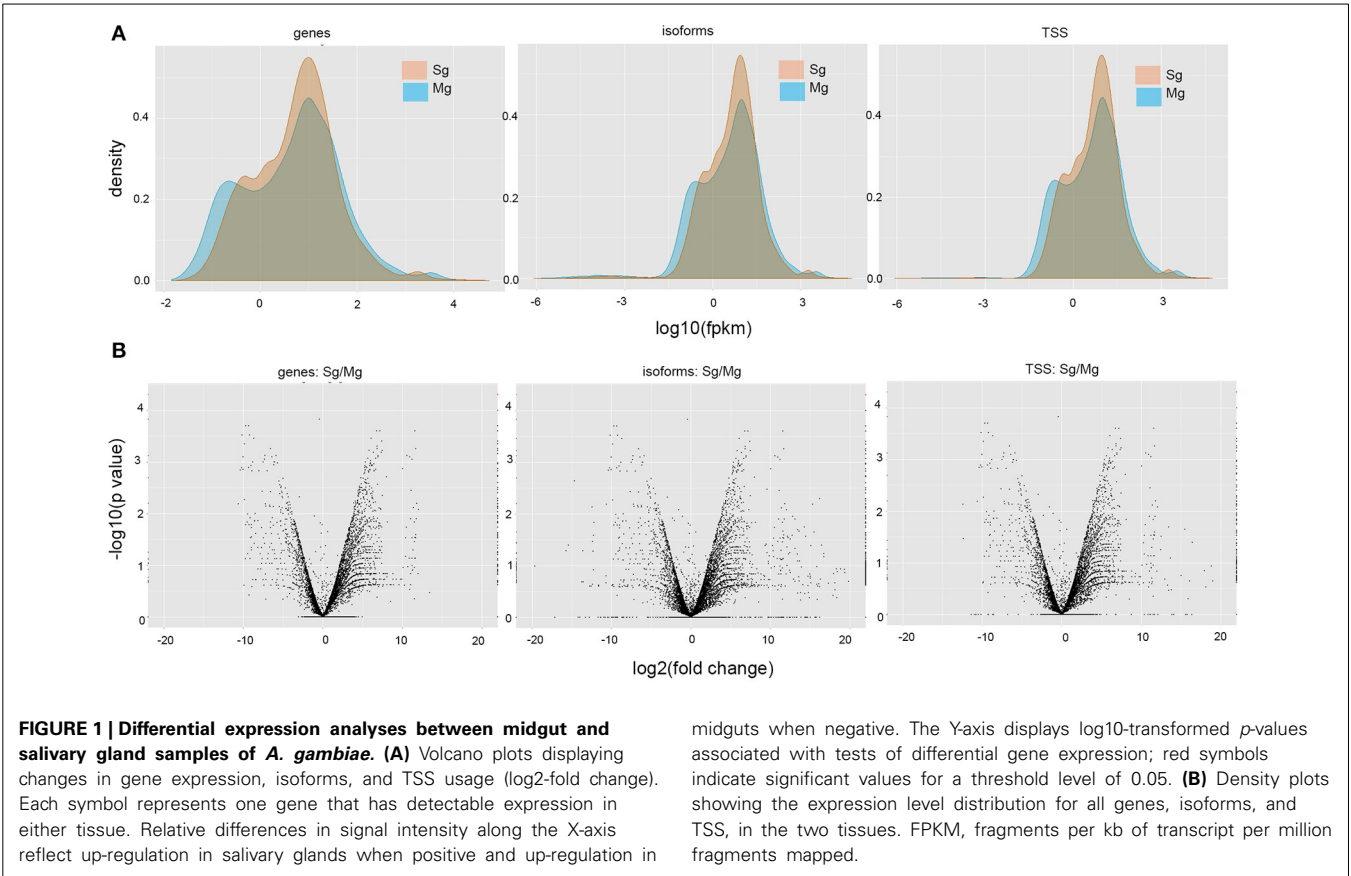
The comparison between transcripts obtained from midguts and salivary glands identified a total of 23 differentially expressed genes, 23 TSS and 21 isoforms switching events (p-value < 0.01, q-value < 0.05) (**Figure 1**). We did not detect significant differences in expression when considering the combined transcriptome, which includes both previously annotated and novel genes (**Figure 1**), probably due to low abundances of novel transcripts, which represent half of the combined data set, and which may have bias quantification results. Note that these are conservative estimates based on the threshold used by Cufflinks.

VALIDATION OF THE ChIP-SEQ APPROACH AND MAPPING OF H3K27ac AND H3K27me3 IN *A. GAMBIAE*

Statistics of data processing are presented in **Table 1**. ChIP and control libraries passed all quality tests (**Figure S1**). The percentage of reads mapped to the mosquito genome was 53.70% for H3K27ac (13,300,823 reads) and 58.19% for H3K27me3 (12,832,729 reads). Of these, 13.73% for H3K27ac and 17.51% for H3K27me3 reads were identified by Bowtie as corresponding to repetitive sequences such as those typically found in transposons and centromeric or telomeric regions. The percentage of mapping is similar between the two target histones and the control. Results of the ChIP-qPCR validation are also consistent with data obtained by the ChIP-seq analysis (data not shown). Taken together, these results indicate that the ChIP-seq data provide an accurate representation of the genome-wide distribution of histone modifications analyzed in *A. gambiae* and offer proof-of-concept validation of the approach.

Table 1 | Statistics of ChIP-seq and RNA-seq data processing and quality checks.

	No. reads	Reads mapped	Reads excluded	Reads analyzed	%GC	Quality score
ChIP-seq						
H3K27ac	24770844	13300823 (53.7%)	3401339 (13.7%)	12286069 (49.6%)	45	38
H3K27me3	22053926	12832729 (58.2%)	3861626 (17.5%)	10249493 (46.5%)	49	38
Input	19456294	9684252 (49.8%)	3235677 (16.6%)	9068488 (50.7%)	42	38
RNA-seq						
Midgut	83046282	60149597 (72.4%)			52	37
Salivary gland	78884200	57346015 (72.7%)			52	37

FIGURE 1 | Differential expression analyses between midgut and salivary gland samples of *A. gambiae*. (A) Volcano plots displaying changes in gene expression, isoforms, and TSS usage (log2-fold change). Each symbol represents one gene that has detectable expression in either tissue. Relative differences in signal intensity along the X-axis reflect up-regulation in salivary glands when positive and up-regulation in midguts when negative. The Y-axis displays log10-transformed *p*-values associated with tests of differential gene expression; red symbols indicate significant values for a threshold level of 0.05. **(B)** Density plots showing the expression level distribution for all genes, isoforms, and TSS, in the two tissues. FPKM, fragments per kb of transcript per million fragments mapped.

Table 2 | Coverage of H3K27ac and H3K27me in the *A. gambiae* genome by chromosome.

Chromosome	H3K27ac		H3K27me3		Size (bp)
	Read depth	% cov	Read depth	% cov	
2L	4051943	8.21	10503751	21.28	49364325
2R	5094054	8.28	15188866	24.68	61545105
3L	2984278	7.11	8646592	20.61	41963435
3R	3645201	6.85	11609840	21.82	53200684
UNKN	111428	0.26	249640	0.59	42389979
X	1661196	6.81	6941572	28.45	24393108
Genome	17548100	6.43	53140261	19.45	273109044

The depth of coverage in bp, the fraction of the chromosome covered, and the total size in bp for each chromosome and the complete genome is indicated.

Coverage analyses of ChIP peaks regions reveal that 6.4% of the genome is occupied by H3K27ac, whereas 22% is occupied by H3K27me3. The fraction of the genome covered by each histone modification across different chromosomes is shown in **Table 2**. By applying a peak calling approach using either MACS or SICER we identified 6639 peaks for H3K27ac and 12,939 peaks for H3K27me3 after input control normalization. Of the peaks identified for H3K27ac, 70.2% (4657) intersect with annotated genes or their promoters (here defined as regions located 200 bp upstream and downstream of TSSs), whereas a small fraction of the remaining peaks (29.8% of total) correspond to intergenic regions. In the case of H3K27me3, only 41% of

the peaks (5414) intersect genes or their promoters and, therefore, most peaks appear to be intergenic. When considering the custom gene set that includes the newly discovered transcripts, the percentage of peaks that overlap to genes is 6156 for H3K27ac and 7813 for H3K27me3, thus indicating that a considerable proportion of ChIP signal maps to non-annotated regions.

GENOME-WIDE PROFILING OF HISTONE AND GENE TRANSCRIPTIONAL STATES

To examine the patterns of histone acetylation and methylation genome-wide, we considered *A. gambiae* genes plus flanking

regions of 200 bp upstream and downstream of the TSS and TTS, respectively. The distribution maps of the two histone modifications at genes revealed a mutually exclusive pattern of H3K27ac and H3K27me3 enrichment genome-wide (**Figure 2**). The same pattern was obtained when using the standard annotated AgamP4 or our custom gene-set as anchors. When examining the average histone modification profiles across the gene regions, results show peak maxima of H3K27ac downstream of the TSSs of genes (**Figure 2B**), whereas H3K27me3 tends to occupy broader regions that cover the entire body of genes (**Figure 2A**). K-means clustering of ChIP-seq data resulted in the classification of genomic regions into three clusters based on their H3K27me3 mean enrichment profiles. Cluster 1 genes, (C1: 12,913 genes) are depleted of H3K27me3 and contain high levels of H3K27ac. This cluster may correspond to genes that are highly transcribed. Cluster 3 genes (C3: 2021 genes) contain high levels of H3K27me3 and they are depleted of H3K27ac. Genes in this cluster probably correspond to Polycomb-silenced genes in *Drosophila*. Interestingly, a third gene cluster identified based on H3K27me3 levels, cluster 2 (C2: 11,976 genes) contains intermediate levels of H3K27me3 and are also generally depleted of H3K27ac but not as dramatically as genes in C3 (**Figure 2D**). Pairwise comparisons between clusters revealed significant differences in mean enrichment values for genes marked with H3K27ac (C1 $= 0.320 \pm 0.013$, C2 $= -0.100 \pm 0.004$, C3 $= -0.198 \pm 0.007$; Kruskal-Wallis, all *p*-values < 0.001), or H3K27me3 (C1 $= -1.281 \pm 0.005$, C2 $= -0.017 \pm 0.004$, C3 $= 2.525 \pm 0.033$; Kruskal-Wallis, all *p*-values < 0.001; Mean values of ChIP enrichment by gene, normalized and input subtracted). It is also noticeable from the heatmaps that at regions where H3K27me3 localize at high levels, the signal often expands toward the 3 and 5′ end of the genes covering most of the flanking region. Likewise, a closer look at genes marked with H3K27ac reveals two sets of genes based on the distribution across the region: genes that have this histone modification at the promoter and coding region and a second group of intermediate enrichment where this mark is centered on promoters (**Figure 2B**).

We next asked how these various histone modification profiles relate to the transcriptional status of genes. Considering genes, annotated and novel, that intersect with H3K27ac and H3K27me peaks, we found significant differences in gene expression between genes containing each of these two histone modification (Kruskal-Wallis chi-squared $= 7138.69$, *df* $= 1$, *p*-value $= 0$). Comparison of histone modification maps and gene expression genome-wide reveals a preferential localization of H3K27ac in active genes and promoters whereas genes that are marked with H3K27me3 are silent or are transcribed at low levels (**Figure 2C**). Indeed, this gene expression pattern is also true in quantitative terms when considering the gene clusters defined based on the levels of enrichment or depletion of histone PTMs (C1: 1.60 ± 0.029, C2: -1.22 ± 0.03, C3: -2.18 ± 0.05; Kruskal-Wallis chi-squared C1–C2 $= 4646.224$, *df* $= 1$; C2–C3 $= 238.38$, *df* $= 1$; C1–C3 $= 2097.31$, *df* $= 1$, all *p*-values < 0.001; Mean gene expression by gene are calculated based on the log2 of FPKM values) (**Figure 2E**).

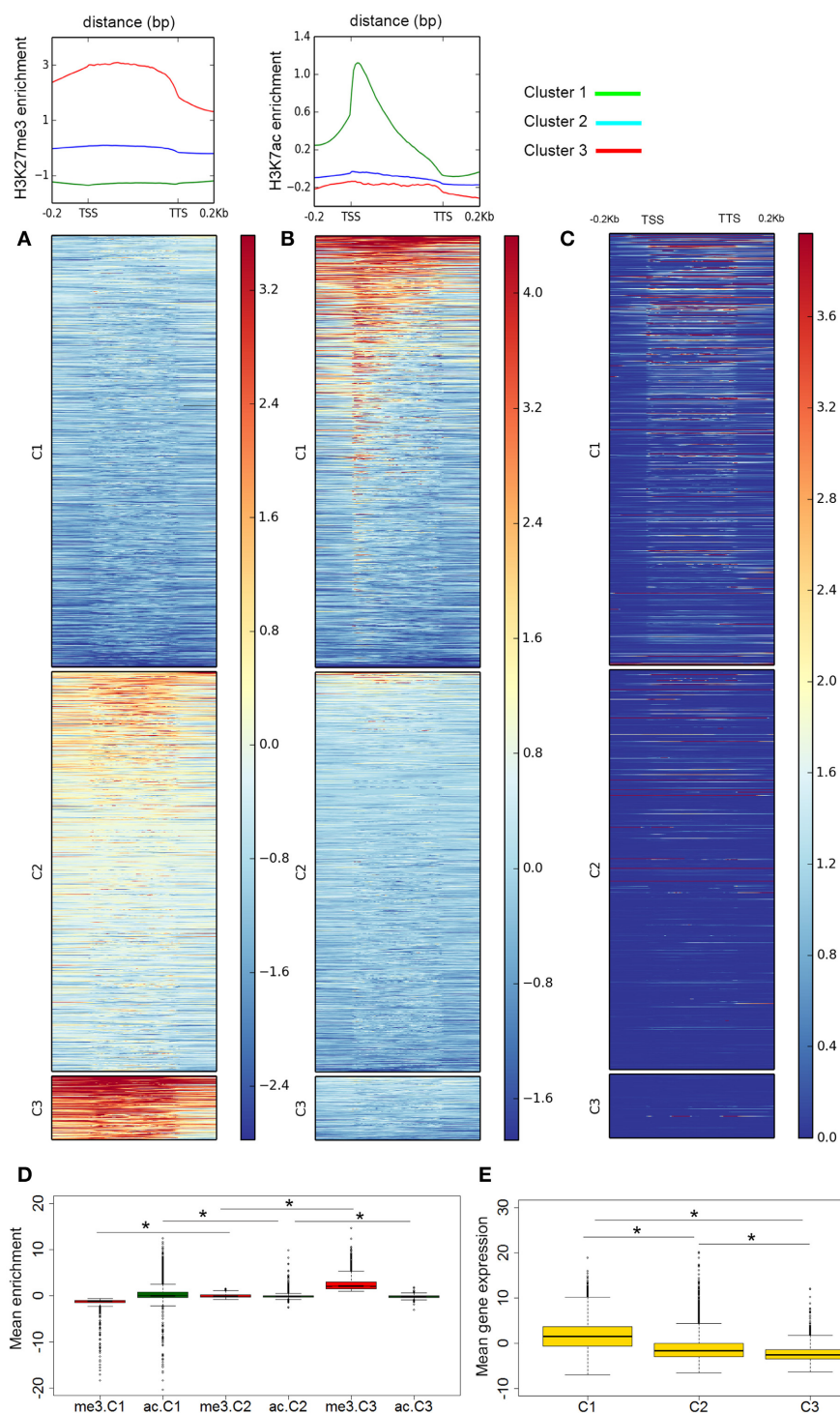## FUNCTIONAL ANNOTATION OF HISTONE MODIFICATION PROFILES

Genes clustered based on their histone modification profile, either marked with H3K27ac or H3K27me3, were assigned to GO accession numbers and classified into functional categories under three major classes (biological process, cellular component, and molecular function). We found multiple GO categories associated with each cluster (Table S2). Genes enriched in either H3K27ac or H3K27me3, were assigned to multiple different KEGG pathways. Of these, pathways involved in metabolism of amino acids and signaling pathways were top ranked hits for genes marked with H3K27ac or H3K27me3 (**Figure S2**). The genes associated to each histone PTM (see **Figure 2** for cluster names) also differ in the relative enrichment of GO terms (**Figure 3**, $p < 0.05$). Particularly, genes in cluster 1, highly enriched in H3K27ac, and cluster 2, containing intermediate levels of H3K27ac and H3K27me3, display a significant differential enrichment in GO terms associated with metabolic processes, immune, and signaling pathways. This is in marked contrast with genes in cluster 3, which display differential enrichment in GO terms related to membrane transporters and developmental processes (**Figure 3**, Table S2). One classic example of developmental genes that fall into this category and are associated with Polycomb and H3K27me3 domains in *Drosophila* are the homeobox genes (Schwartz et al., 2006) (**Figure 4A**, GO:0005634 Table S2). We then asked whether there is an association between the GO terms retrieved for genes differentially marked with H3K27ac and H3K27me3, and GO terms that have been linked to malaria infection and mosquito resistance to insecticides (Table S3). Interestingly, genes marked with these two histone modification marks include immunity related genes such as Toll, Immune Deficiency (IMD) and Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathways; antimicrobial effectors such as defensins and cepropins; and metabolic genes such as the Vitellogenin gene. Examples of histone modification profiles of two of these candidate genes, which are significantly associated with enriched GO terms, are shown in **Figure 4B**, and the histone modification profiles for each candidate gene examined are shown in Table S3.

DISCUSSION

Changes in chromatin state play an important role in a number of biological processes, including responses to stress and environmental conditions. In organisms like *Drosophila* that lack DNA methylation, post-translational modifications of histones are an important layer of gene regulation (Bonasio et al., 2010; Bannister and Kouzarides, 2011). However, chromatin maps and the dynamics of epigenetic control in mosquitoes have remained to date unexplored. In order to fill this gap we describe genome-wide maps of two histone modifications with key roles in gene regulation, H3K27ac and H3K27me3. To establish a link with gene expression, we then correlate the presence of these modifications in specific genes with their transcriptional state in midguts and salivary glands of *A. gambiae* females.

In *Drosophila*, various combinations of active and repressive histone marks have been found to be associated with distinct chromatin functional states (Filion et al., 2010; Roy et al., 2010; Kharchenko et al., 2011; Negre et al., 2011; Yin et al., 2011). Two

**FIGURE 2 | Genome-wide distribution of histone modifications.** Distribution of **(A)** H3K27me3 and **(B)** H3K27ac with respect to gene features in *A. gambiae* midguts. The enrichment or depletion is shown relative to chromatin input. The regions in the map comprise gene bodies flanked by a segment of 200 bp at the 5′ end of TSSs and TTSs. Average profiles across gene regions ±200 bp for each histone modification are shown on top. **(C)** Heatmap of RNA-seq data showing the level gene expression, as read count, profiled along the region. In all heatmaps **(A–C)** genes were organized into 3 clusters based on their level of H3K27me3. For H3K27ac, cluster genes are ordered by descent but independent to H3K27me3. The color bars indicate the range of intensities based on ChIP enrichment, from red to blues for higher to low enrichment values. Boxplots showing the mean enrichment of H3K27ac and H3K27me3 by cluster **(D)**, and mean level of gene expression by gene cluster **(E)**. Significant pairwise comparisons are indicated by asterisks (see text).

FIGURE 3 | Functional analysis of *A. gambiae* genes. The graphs show GO terms significantly associated with genes that show significant enrichment or depletion of H3K27ac and H3K27me3 at high levels (see **Figure 2**). Bars corresponds to number of sequences associated with each GO term. In the case of genes marked with H3K27ac from cluster 1 due to the number of records only most significant GO terms are shown (see Table S2 for the complete list).

of these histone modifications that display contrasting distribution patterns and regulatory functions include acetylation and methylation of lysine 27 of histone 3. Previous studies have shown that H3K27ac is typically enriched at enhancers, TSSs and coding

regions of active genes, whereas H3K27me3 localizes to repressed loci (Tolhuis et al., 2006; Kharchenko et al., 2011; Kellner et al., 2012; Van Bortle et al., 2012; Brown and Bachtrog, in press). Results described here for the mosquito *A. gambiae* agree with
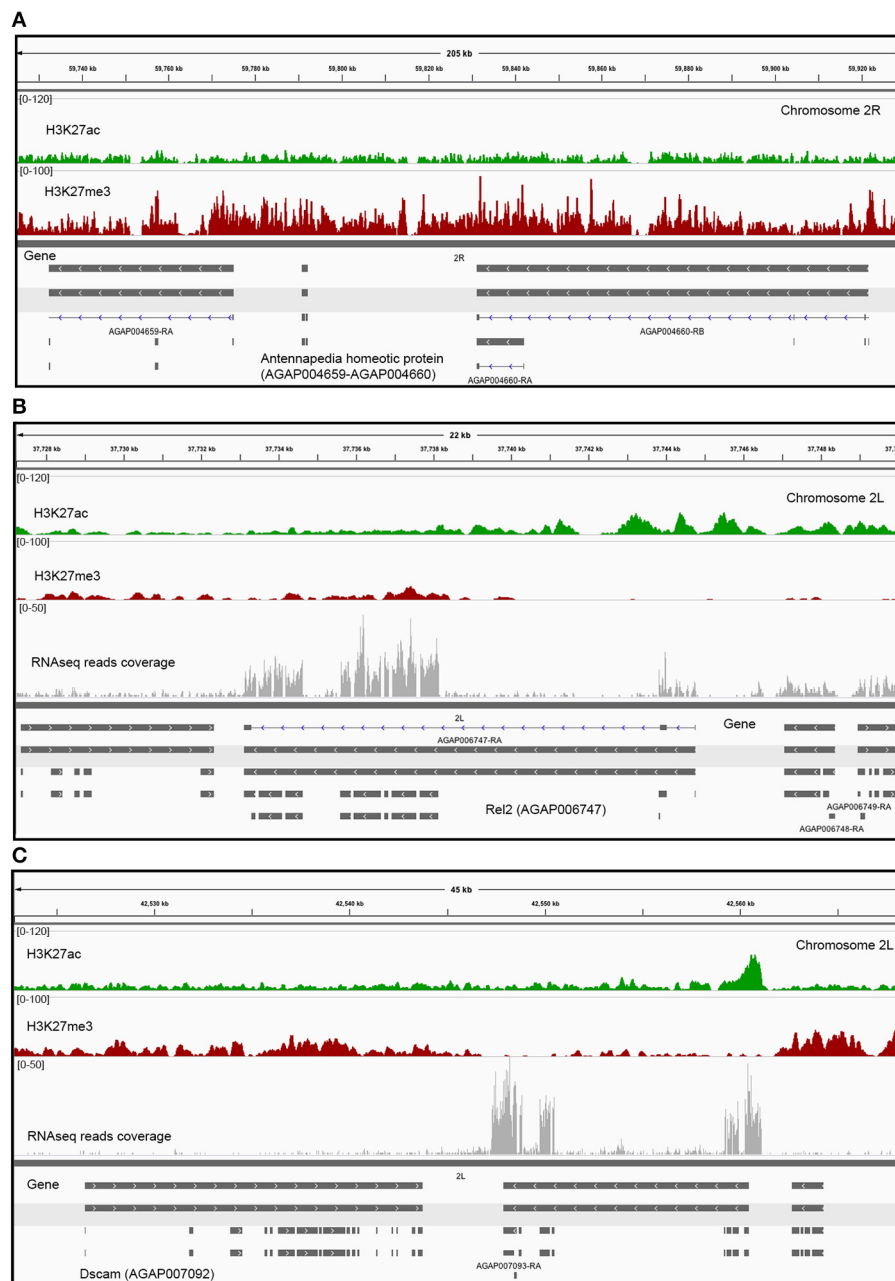
FIGURE 4 | Histone modification levels in selected genes. (A) Enrichment profile of histone PTMs in two *Hox* genes of *A. gambiae* (Antennapedia homeotic proteins, AGAP004659-AGAP004660). These are classic examples of *Drosophila* developmental genes where H3K27me3 is highly enriched and distributed in large domains that encompass both genic and intergenic regions in cells where the *Antennapedia* gene is silenced. Examples of various histone modification profiles, based on the level of enrichment or depletion in H3K27ac or H3K27me3, in two candidate genes: **(B)** *Rel2* (AGAP006747, GO:0005515), and **(C)** *Dscam* (AGAP007092, GO: GO:0005515) (Tables S2, S3). The scale of the tracks is proportional to the number of sequencing reads for each histone modification. Gene expression in terms number of RNA reads mapped to each gene is also displayed.

these observations and suggest a close association between the presence of H3K27ac and H3K27me3, and the transcriptional state of genes.

Genome-wide histone maps in *A. gambiae* reveal that H3K27ac and H3K27me3 cover approximately 6 and 14% of the genome, respectively. These two genome fractions do not overlap and there is a mutually exclusive pattern of enrichment at most genes: H3K27me3-containing promoters and genes show no enrichment for H3K27ac, and, reciprocally, those containing H3K27ac are not enriched for H3K27me3 (**Figure 2**). This is in agreement with previous studies in *Drosophila*, and supports the notion that these two histone modifications also have

opposite biological functions in mosquitoes. Despite a general mutual exclusive pattern in the distribution of methylated and acetylated H3K27 sites, we found some regions of overlap, including both genic and intergenic regions. A possible explanation for this pattern would be promoter bivalency, which occurs when gene promoters contain both active and repressive modifications. These promoters are considered poised for activation (or repression) at a later stage. But this pattern has been described in the context of differentiating cells or during development (Creyghton et al., 2010), and is restricted in principle to promoters containing activating histone H3K4me3 and repressive H3K27me3 marks, not H3K27ac (Voigt et al., 2013). Alternatively, the presence of the two histone modifications can be the result of heterogeneous enrichments across cells as it has been recently been shown using single-cell ChIP-seq technologies (Wang and Bodovitz, 2010). In the case of intergenic regions, co-localization of H3K27ac and H3K27me3 has been associated with the presence of enhancer elements (Creyghton et al., 2010). To validate this hypothesis in mosquitoes it would be necessary to map other histone modifications such as H3K4me1 that are diagnostic for this type of regulatory elements (Visel et al., 2009; Zhu et al., 2013).

The combined analysis of ChIP-seq and RNA-seq data in midgut cells of *A. gambiae* supports a relationship between transcription levels and profiles of histone PTMs. We find that the presence of H3K27ac coincides with actively transcribed genes, whereas H3K27me3 is mostly associated with clusters of repressed genes that show low levels of transcription. Our analyses also suggest that enrichment or depletion of H3K27ac and H3K27me3 correlates not only with the transcriptional state of genes but also with the magnitude of their expression or silencing, which is in agreement with earlier studies in *Drosophila* (Negre et al., 2011). For example, in flies and mammals it has been shown that H3K27me3 can form domains that can expand 100 kb or more and covers blocks of silent genes and intergenic regions, and expressed genes that are preferentially located in regions immediately flanking H3K27me3 BLOCs (Pauler et al., 2009; Young et al., 2011; Hou et al., 2012). Many of these domains can also contain Polycomb (Schwartz et al., 2006; Tolhuis et al., 2006; Schuettengruber et al., 2009). In this study, Polycomb-containing genes would correspond to those of cluster 3 in **Figure 2**. One example of this type is the *Hox* genes of *Drosophila*, which are active in embryonic and larval stages of the life of this organism. These *Hox* genes have their homologous in mosquitoes and they display a pattern of H3K27me3 enrichment as broad peak blocks in differentiated cells of the midgut (**Figure 4**). It has been shown in *Drosophila* that H3K27me3 can be also be present in genes in euchromatic regions as focal peaks, rather than in domains, where there is a dynamic turnover between active and repressive histone marks and where these marks participate in the fine tuning of transcription (Riddle et al., 2011; Young et al., 2011). In the case of *A. gambiae* such a distribution pattern would correspond to genes of cluster 2, which show intermediate levels of H3K27ac and H3K27me3 and medium to low transcription (**Figure 2**). Finally, genes containing H3K27ac show different enrichment levels in this histone PTM. Particularly interesting is the finding of a subset of genes that have this

modification only at promoters and show intermediate levels of transcription, suggesting that RNA Polymerase II may be paused at these regions (**Figure 2**). This has been previously reported by ChIP-seq studies in flies and mammals that mapped RNA-polymerase binding patterns in association with various histone PTMs and regulatory elements genome-wide (Negre et al., 2011; Adelman and Lis, 2012). These findings together support the conclusion that mechanisms of regulation of chromatin structure and function have been conserved between *Anopheles* and *Drosophila*.

The functional analysis of genes differentially enriched in histone PTMs further supports the idea that histone modification profiles correlate with gene expression in *A. gambiae*. This finding, and the similarities found with the mechanisms of epigenetic regulation described in *Drosophila*, has potentially important implications in the context of infection and the response of mosquitoes to environmental stresses. We further explored this issue by examining the enrichment in H3K27 PTMs and the transcriptional status of a set of target genes that have been previously shown to be up-regulated in response to malaria infection as well as genes involved in insecticide resistance (referenced in Table S3). Interestingly, the majority of the candidate genes examined appear to be enriched in one of the two histone modifications, and most fall within clusters 1 and 2, which suggests that they are in a dynamic on/off switch for transcriptional regulation (Table S3).

One limitation of epigenomic studies in non-model organisms is the availability of fully annotated genomes that allow high resolution mapping of epigenetic modifications and regulatory elements to their target genes. This is clearly illustrated in this study, and may be one of the reasons why analysis of epigenetic modifications has remained unexplored in mosquitoes. The analysis of the *A. gambiae* transcriptome in midguts and salivary glands using RNA-seq resulted in the identification of several thousand novel transcripts. The high discovery rate is not surprising considering that the majority of transcriptome studies carried out to date in *A. gambiae* have employed microarrays (Baker et al., 2011; Maccallum et al., 2011). It is important to note that a large fraction of regions enriched for H3K27ac and H3K27me3 map to non-annotated genes. The differential enrichment in histone modification marks and tissue specific signatures of transcription at these genes can be important in predicting potential targets in the response of *A. gambiae* to pathogens and environmental stressors.

CONCLUSIONS AND FUTURE DIRECTIONS

Here we present an integrative approach that combines genome-wide profiling of histone modifications and gene expression that allows the functional interpretation of chromatin modifications as well as *de novo* prediction of regulatory elements and genes. Results support a link between changes in histone modification profiles and transcription levels in *A. gambiae*, and this includes genes that have been previously identified as having important roles in the epidemiology and control of malaria transmission. The patterns obtained are similar to those previously described in *Drosophila*, suggesting the existence of common mechanisms of chromatin regulation in *A. gambiae*. These findings open new

opportunities to study the dynamics of chromatin states including other histone PTMs in various developmental stages and environmental conditions (i.e., infection or exposure to insecticides) as well as the mechanisms by which these states are controlled by regulatory elements such as enhancers and insulators.

DATA ACCESS

ChIP-seq and RNA-seq raw data as well as processed data (wig, bed, FPKM values and GTF files) are deposited in GEO database under accession number GSE59773.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00277/abstract

Figure S1 | Fingerprint plot that shows the distribution of reads, as the cumulative sum of read counts in 10 bp window bins, for H3K27ac, H3K27me3, and input samples. A tight diagonal is expected when reads are equally distributed across the genome, as is the case for the input, whereas the curve becomes more pronounced as the degree of enrichment increases and is more localized in the ChIPs relative to the control sample.

Figure S2 | Piecharts that show the number of sequences assigned to KEGG pathways for each histone modification gene cluster (see text, Figure 2).

REFERENCES

Adelman, K., and Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731. doi: 10.1038/nrg3293

Baker, D., Nolan, T., Fischer, B., Pinder, A., Crisanti, A., and Russell, S. (2011). A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, Anopheles gambiae. *BMC Genomics* 12:296. doi: 10.1186/1471-2164-12-296

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi: 10.1038/cr.2011.22

Blandin, S. A., Marois, E., and Levashina, E. A. (2008). Antimalarial responses in *Anopheles gambiae*: from a complement-like protein to a complement-like pathway. *Cell Host Microbe* 3, 364–374. doi: 10.1016/j.chom.2008.05.007

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* 330, 612–616. doi: 10.1126/science.1191078

Bonduriansky, R., and Day, T. (2009). Non-genetic inheritance and its evolutionary implications. *Annu. Rev. Ecol. Evol. Syst.* 40, 103–125. doi: 10.1146/annurev.ecolsys.39.110707.173441

Bowman, S., Simon, M., Deaton, A., Tolstorukov, M., Borowsky, M., and Kingston, R. (2013). Multiplexed Illumina sequencing libraries from picogram quantities of DNA. *BMC Genomics* 14:466. doi: 10.1186/1471-2164-14-466

Brown, E. J., and Bachtrog, D. (in press). The chromatin landscape of Drosophila: comparisons between species, sexes, and chromosomes. *Genome Res.* doi: 10.1101/gr.172155.114

Cheng, C., Yan, K.-K., Yip, K. Y., Rozowsky, J., Alexander, R., Shou, C., et al. (2011). A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. *Genome Biol.* 12:R15. doi: 10.1186/gb-2011-12-2-r15

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21931–21936. doi: 10.1073/pnas.1016071107

Dana, A. N., Hong, Y. S., Kern, M. K., Hillenmeyer, M. E., Harker, B. W., Lobo, N. F., et al. (2005). Gene expression patterns associated with blood-feeding in the malaria mosquito Anopheles gambiae. *BMC Genomics* 6:5. doi: 10.1186/1471-2164-6-5

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* 28, 817–825. doi: 10.1038/nbt.1662

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* 7, 1728–1740. doi: 10.1038/nprot.2012.101

Filion, G. J., Van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143, 212–224. doi: 10.1016/j.cell.2010.09.009

Gómez-Díaz, E., Jordà, M., Peinado, M. A., and Rivero, A. (2012). Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathog.* 8:e1003007. doi: 10.1371/journal.ppat.1003007

Hoeijmakers, W. A., Stunnenberg, H. G., and Bartfai, R. (2012). Placing the *Plasmodium falciparum* epigenome on the map. *Trends Parasitol.* 28, 486–495. doi: 10.1016/j.pt.2012.08.006

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149. doi: 10.1126/science.1076181

Hou, C., Li, L., Qin, Z. S., and Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* 48, 471–484. doi: 10.1016/j.molcel.2012.08.031

Kellner, W. A., Ramos, E., Van Bortle, K., Takenaka, N., and Corces, V. G. (2012). Genome-wide phosphoacetylation of histone H3 at Drosophila enhancers and promoters. *Genome Res.* 22, 1081–1088. doi: 10.1101/gr.136929.111

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., et al. (2011). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471, 480–485. doi: 10.1038/nature09725

Koutsos, A. C., Blass, C., Meister, S., Schmidt, S., Maccallum, R. M., Soares, M. B., et al. (2007). Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11304–11309. doi: 10.1073/pnas.0703988104

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Lambrechts, L. (2010). Dissecting the genetic architecture of host–pathogen specificity. *PLoS Pathog.* 6:e1001019. doi: 10.1371/journal.ppat.1001019

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831. doi: 10.1101/gr.136184.111

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Maccallum, R. M., Redmond, S. N., and Christophides, G. K. (2011). An expression map for *Anopheles gambiae*. *BMC Genomics* 12:620. doi: 10.1186/1471-2164-12-620

Marinotti, O., Nguyen, Q. K., Calvo, E., James, A. A., and Ribeiro, J. M. (2005). Microarray analysis of genes showing variable expression following a blood meal in *Anopheles gambiae*. *Insect Mol. Biol.* 14, 365–373. doi: 10.1111/j.1365-2583.2005.00567.x

Negre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., et al. (2011). A cis-regulatory map of the Drosophila genome. *Nature* 471, 527–531. doi: 10.1038/nature09990

Pauler, F. M., Sloane, M. A., Huang, R., Regha, K., Koerner, M. V., Tamir, I., et al. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions

and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* 19, 221–233. doi: 10.1101/gr.080861.108

Pitts, R. J., Rinker, D. C., Jones, P. L., Rokas, A., and Zwiebel, L. J. (2011). Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue-and sex-specific signatures of odor coding. *BMC Genomics* 12:271. doi: 10.1186/1471-2164-12-271

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., and Manke, T. (in press). DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* doi: 10.1093/nar/gku365

Riddle, N. C., Minoda, A., Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Tolstorukov, M. Y., et al. (2011). Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* 21, 147–163. doi: 10.1101/gr.110098.110

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27, 2325–2329. doi: 10.1093/bioinformatics/btr355

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657. doi: 10.1038/nmeth1068

Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330, 1787–1797. doi: 10.1126/science.1198374

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., et al. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol.* 7:e13. doi: 10.1371/journal.pbio.1000013

Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X.-Y., Bourgon, R., Biggin, M., et al. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705. doi: 10.1038/ng1817

Severson, D. W., and Behura, S. K. (2012). Mosquito genomics: progress and challenges. *Annu. Rev. Entomol.* 57, 143–166. doi: 10.1146/annurev-ento-120710-100651

Tolhuis, B., De Wit, E., Muijrers, I., Teunissen, H., Talhout, W., Van Steensel, B., et al. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699. doi: 10.1038/ng1792

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J. E., and Corces, V. G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* 22, 2176–2187. doi: 10.1101/gr.136788.111

Vannini, L., Augustine Dunn, W., Reed, T. W., and Willis, J. H. (2014). Changes in transcript abundance for cuticular proteins and other genes three hours after a blood meal in *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* 44, 33–43. doi: 10.1016/j.ibmb.2013.11.002

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858. doi: 10.1038/nature07730

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev.* 27, 1318–1338. doi: 10.1101/gad.219626.113

Wang, D., and Bodovitz, S. (2010). Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* 28, 281–290. doi: 10.1016/j.tibtech.2010.03.002

Who. (2013). *World Malaria Report: 2013.* Geneva: World Health Organization.

Xu, S., Grullon, S., Ge, K., and Peng, W. (2014). "Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to map regions of histone methylation patterns in embryonic stem cells," in *Stem Cell Transcriptional Networks*, ed B. L. Kidder (New York, NY: Springer), 97–111.

Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2011). A high-resolution whole-genome map of key chromatin modifications in the adult *Drosophila melanogaster*. *PLoS Genet.* 7:e1002380. doi: 10.1371/journal.pgen.1002380

Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., et al. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res.* 39, 7415–7427. doi: 10.1093/nar/gkr416

Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952–1958. doi: 10.1093/bioinformatics/btp340

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi: 10.1186/gb-2008-9-9-r137

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res.* 41, 10032–10043. doi: 10.1093/nar/gkt826

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on