frontiers in
GENETICS

REVIEW ARTICLE
published: 10 December 2013
doi: 10.3389/fgene.2013.00273

Copy-number changes in evolution: rates, fitness effects and adaptive significance

Vaishali Katju* and Ulfar Bergthorsson

Department of Biology, University of New Mexico, Albuquerque, NM, USA

Edited by:
Frederic JJ Chain, Max Planck Institute for Evolutionary Biology, Germany

Reviewed by:
Frederic Guy Brunet, Ecole Normale Supérieure de Lyon, France
Ben J Evans, McMaster University, Canada

****Correspondence:***
Vaishali Katju, Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA
e-mail: vkatju@unm.edu

Gene copy-number differences due to gene duplications and deletions are rampant in natural populations and play a crucial role in the evolution of genome complexity. Per-locus analyses of gene duplication rates in the pre-genomic era revealed that gene duplication rates are much higher than the per nucleotide substitution rate. Analyses of gene duplication and deletion rates in mutation accumulation lines of model organisms have revealed that these high rates of copy-number mutations occur at a genome-wide scale. Furthermore, comparisons of the spontaneous duplication and deletion rates to copy-number polymorphism data and bioinformatic-based estimates of duplication rates from sequenced genomes suggest that the vast majority of gene duplications are detrimental and removed by natural selection. The rate at which new gene copies appear in populations greatly influences their evolutionary dynamics and standing gene copy-number variation in populations. The opportunity for mutations that result in the maintenance of duplicate copies, either through neofunctionalization or subfunctionalization, also depends on the equilibrium frequency of additional gene copies in the population, and hence on the spontaneous gene duplication (and loss) rate. The duplication rate may therefore have profound effects on the role of adaptation in the evolution of duplicated genes as well as important consequences for the evolutionary potential of organisms. We further discuss the broad ramifications of this standing gene copy-number variation on fitness and adaptive potential from a population-genetic and genome-wide perspective.

Keywords: copy-number variants, deletion, duplication, fitness effect, spontaneous rate

INTRODUCTION

The publication of Ohno's "*Evolution by Gene Duplication*" is fittingly viewed as a milestone in the study of gene duplications (Ohno, 1970). In addition to collating evidence for duplications in evolution, it also presented several hypotheses that have since been undergoing robust testing and analyses. For example, Ohno perceived that segmental duplications would be associated with problems with gene dosage balance and genetic instability, and therefore he also placed a great significance on whole-genome duplications. Additionally, he viewed the duplicate copy of a gene as an initially passive element in the evolution of new genes. A duplicated gene was seen as superfluous and therefore not under selection after duplication, that is, not until subsequent mutations conferred novel beneficial functions. Therefore, Ohno predicted that in the majority of instances, a gene duplicate would be lost or degenerate into a pseudogene.

The first characterized segmental gene duplication was the *bar* mutation in *Drosophila melanogaster* (Sturtevant, 1925). Soon after the discovery of the *bar* mutation, Bridges (1935, 1936) suggested that the duplication of genes provided a mechanism for increasing chromosome length and providing material for subsequent functional changes. This potential borne by gene duplication for evolutionary change was further emphasized by early geneticists and evolutionary biologists like Haldane, Müller, and Huxley (Haldane, 1933; Müller, 1935, 1936; Huxley, 1942). The *bar* mutation also serves as an illustration of several general features that should be emphasized about duplications. First, although it is "simply" a duplication of previously existing material that is expected to increase "redundancy" in the genome, the duplication has a striking phenotype. Gene duplication theory often treats duplications as having no immediate consequences after conception under the general assumption that gene duplicates must endure a passive existence in the genome until subsequent mutational events shape their eventual fate toward nonfunctionalization, subfunctionalization, or neofunctionalization. Thus, the immediate phenotypic and fitness consequences of duplications have not received the same degree of attention. Second, the fitness consequences of the *bar* mutation are most likely deleterious (Geer and Green, 1962). Although there is abundant evidence of beneficial duplications, particularly in the context of stressful or perturbed environmental conditions (Maroni et al., 1987; Theodore et al., 1991; Brown et al., 1998; Evgen'ev et al., 2004; Hemingway et al., 2004; Gonzalez et al., 2005; Deng et al., 2010; Nasvall et al., 2012; among others), changes in gene copy-number are usually deleterious (Lupski, 1998; Inoue and Lupski, 2002; Botstein and Risch, 2003; Bailey and Eichler, 2006; Sebat et al., 2007). Before the recent advances in detecting copy-number changes, an estimated 29% of human genetic diseases were thought to result from gene copy-number changes, with 22 and 7% stemming from gene deletions and duplications, respectively (Botstein and Risch, 2003). Lastly, Sturtevant and Morgan (1923) discovered that the segmental duplications that gave rise to the *bar*

phenotype were unstable. Although the original experiments on the *bar* mutation do not provide an estimate of the rate of duplication, the frequency of reversions due to duplication loss and the frequency of double-*bar* mutation from *bar* flies was very high, on the order of approximately 10^{-3} per generation (Sturtevant, 1925). These early experiments with the *bar* mutation therefore showed that gene copy-number changes can occur at much higher rates than point mutations.

The study of structural genetic variation is undergoing an epochal resurgence. The reasons for this increased interest are largely technical. The explosive increase in the number of sequenced genomes has made it abundantly clear that the primary source of new genes is gene duplication, as previously advanced by Ohno (1970). Complementarily, high-throughput screens of structural variation in natural populations have demonstrated that there is abundant genetic variation in gene copy-number variation that we were previously unable to detect on a genome-wide scale (Iafrate et al., 2004; Sebat et al., 2004; Maydan et al., 2007; Emerson et al., 2008; among others). Finally, direct measurements of mutation rates have shown that structural genetic variation arises much more frequently than bioinformatic analysis of the age-distribution of extant duplicates in the first sequenced genomes had suggested (Lynch et al., 2008; Lipinski et al., 2011; Schrider et al., 2013). The high frequencies of spontaneous genome rearrangements and gene copy-number variants (CNVs) have important implications for the evolution of novel genes, speciation and hereditary disease. Much of the recent work in gene duplication has focused on gene copy-number polymorphisms in natural populations, and testing hypotheses of functional divergence between paralogs. Here, however, we review recent developments on two related topics regarding gene duplications, namely the spontaneous rate of segmental gene duplications and deletions, and their fitness consequences.

THE FATE OF DUPLICATED GENES IN POPULATIONS

Although genomes can provide a rich record of the history of gene duplications in a particular lineage, the population-genetic dynamics and selection pressures on duplicated genes remain poorly understood. The frequency of gene copy-number polymorphisms in populations is determined by a combination of the spontaneous duplication/deletion rate and the preservation or elimination of these changes by natural selection and/or random genetic drift.

The fixation of a gene duplicate in a population faces multiple obstacles. First, there is a high probability that the duplicated gene is lost from the population by random genetic drift. Moreover, most gene duplications are probably detrimental to organismal fitness. They can perturb optimal dosage balance between genes contained in the duplicated regions with genes elsewhere in the genome, and increased gene dosage can be costly because of superfluous gene expression (Papp et al., 2003; Veitia, 2004). Empirical estimates of this cost in *Salmonella* was found to be substantial (3–16%; Reams et al., 2010). In addition to reducing fitness, many gene duplications are inherently unstable, particularly if they are in tandem orientation or flanked by repeat elements (Anderson and Roth, 1981). Lastly, given that most mutations are degenerative, a duplicated gene is much more likely to end up as a pseudogene than to acquire a function that is distinct from the ancestral gene and actively maintained by natural selection. Loss of one copy, either due to deletion or mutational inactivation is the fate of the overwhelming majority of duplicated genes (Haldane, 1933; Lynch and Conery, 2000). How redundant gene copies get to be fixed and subsequently maintained in a population has emerged as an important issue in the population-genetic theory of evolution by gene duplication (Force et al., 1999).

Several mechanisms have been proposed that would facilitate retention of a duplicated gene in a genome. (i) Redundancy could be beneficial because it protects the genome from the immediate deleterious effects of degenerative mutations (Clark, 1994). (ii) Degenerative mutations can lead to loss of different subfunctions in the two copies of a gene in such a way that both copies would be required to perform what was originally the role of a single ancestral locus (DDC, Duplication-Degeneration-Complementation; Hughes, 1994; Force et al., 1999). (iii) If there is a heterotic interaction (or overdominance) between alleles at a locus, the same beneficial interaction between alleles at two loci can maintain the duplication through natural selection (Spofford, 1969). (iv) Natural selection can result in functional divergence (neofunctionalization) between alleles prior to gene duplication and different alleles can then be preserved at different loci following duplication (Proulx and Phillips, 2006). (v) Although gene duplications create redundant gene copies, many detrimental mutations could still be subject to purifying selection if they interfere with the function of the wild-type copy and this would delay the process of turning one of the gene copies into a pseudogene (Walsh, 2003). However, selection against these detrimental mutations would not protect against the deletion of duplicated genes. (vi) Increase in gene dosage ("more of the same") can be advantageous directly and would result in an increase in gene copy-number (Ohno, 1970). Selection for greater gene dosage does not have to be for the gene's primary activity. When a promiscuous side-function of a gene becomes biologically valuable, selection for increase in gene dosage would help the spread and maintenance of a duplicated gene in the population until subsequent beneficial mutations result in a novel gene (Roth et al., 1996; Hendrickson et al., 2002; Hooper and Berg, 2003; Bergthorsson et al., 2007). There are certain similarities between some of these proposed mechanisms of selective retention of duplicates. For example, hypotheses (iii), (iv), and (vi), depend on natural selection for functions that are already present in the population prior to duplication.

THE IMPORTANCE OF THE GENE DUPLICATION RATE IN EVOLUTION

The rate at which copy-number variation is introduced and eradicated from populations is crucial to understanding the early evolutionary dynamics of novel genes and the evolution of complexity. Both the standing levels of genetic variation and the genetic load are expected to be critically dependent on the rates and fitness effects of spontaneous gene duplications and deletions. The resolution of the duplication and deletion rate parameters will also serve to elucidate the role of gene copy-number in the evolution of disease.

The duplication rate is a key parameter in determining the equilibrium frequency of gene copy-number in populations. For neutral duplications, the equilibrium frequency of duplicated

genes is expected to be $D/(D + L)$, with D as the spontaneous duplication rate and L as the rate of spontaneous loss of duplicate gene copies. In the event of deleterious duplications, the equilibrium frequency still depends largely on the duplication rate. The opportunity for mutations that result in the maintenance of duplicate copies, either through neofunctionalization or subfunctionalization, depend on the equilibrium frequency of additional gene copies in the population, and hence on the spontaneous gene duplication (and loss) rate. The duplication rate may therefore have profound effects on the role of adaptation in the evolution of duplicated genes (Ohta, 1988).

Following the rediscovery of Mendel's laws, some geneticists started attributing greater importance to mutations as the driving force in evolutionary change, and de-emphasizing the importance of natural selection (Morgan, 1916, 1925). The importance of mutations and their rate as the greatest determining factor in evolution fell out of favor after it was shown that the mutation rate is, at best, a very weak force in effecting changes in allele frequency (Haldane, 1932, 1933). The neutral theory led to a greater appreciation of mutation rates as an evolutionary force, but primarily for neutral mutations (Kimura, 1983). More recently, theoretical and experimental evidence suggest that differences in mutation rates can have an orienting effect on evolutionary change (Yampolsky and Stoltzfus, 2001; Rokyta et al., 2005). Mutations are, in this view, not simply raw material for evolutionary change, but the differences in the rates of supply of different mutations influences the outcome with respect to adaptive evolutionary change. Given equal mutation rates, the mutations with the highest fitness contributions will, on average, be fixed first (Orr, 2003). However, mutations that are less fit can be fixed in the population earlier than the fittest mutation if the former are more frequent (Yampolsky and Stoltzfus, 2001; Rokyta et al., 2005). Moreover, the influence of the mutation rate on the rate of fixation of beneficial mutations is greater at smaller effective population sizes (Yampolsky and Stoltzfus, 2001). Let us consider the case of selection for increased gene dosage. Both gene duplication and point mutation can result in increased gene expression, and many point mutations might yield higher expression levels than duplications. However, if the gene duplication rate greatly exceeds the per nucleotide substitution rate, duplications will have an opportunity to increase in frequency, and perhaps reach fixation, before the appearance of point mutations in the population with similar or greater effects on gene expression. The rate of gene duplication relative to base substitutions is therefore particularly relevant for the hypothesis that selection for gene dosage is important in the initial preservation of duplicated genes.

ANALYTICAL METHODS USED TO ESTIMATE THE GENE DUPLICATION AND DELETION RATE

Several approaches have been used to estimate the spontaneous gene duplication and deletion rates. These estimates have primarily come from four sources: (i) direct measurements on a single locus where gene copy-number differences resulted in a distinct phenotype or genotype, (ii) analyses of frequencies of duplication polymorphisms in populations, (iii) calculations based on the abundance of evolutionarily recent gene duplications in sequenced genomes, and (iv) direct genome-wide estimates of the duplication/deletion rate from molecular analyses of mutation accumulation (MA) lines evolved experimentally under a regime of minimal natural selection.

Direct estimates at specific loci have yielded the highest gene duplication frequencies. In contrast, analysis of the age distribution of genes in sequenced genomes yields rates that are orders of magnitude lower (Lynch and Conery, 2000, 2003; Gu et al., 2002; Pan and Zhang, 2007). However, the analyses of sequenced genomes assume that the birth and death rates of duplicated genes are constant over long evolutionary periods. This may be unwarranted if most gene duplications are detrimental and removed from the population by natural selection soon after conception.

PER-LOCUS RATES

Per-locus rates of gene duplication have been empirically generated for bacteria, flies and humans (**Table 1**). However, these estimates are often based on a very limited number of loci and may not be representative for these genomes.

PROKARYOTES

Early experiments with phage and bacteria suggested a fairly high duplication rate per gene. For example, experiments with the

Table 1 | Locus-specific duplication rates for prokaryotes and eukaryotes.

Species	Locus-specific duplication rates	
	Locus	**Partial genome**
Prokaryotes		
S. enterica	2.0×10^{-3} (ArgH)[a]	$3.2 \times 10^{-3} - 5.8 \times 10^{-5}$ duplications per locus[b]
	3.0×10^{-4} (LacZ)[a]	
	4.6×10^{-6}(PyrD)[a]	
Multicellular eukaryotes		
D. melanogaster	1.6×10^{-5}(Rosy)[c]	
	1.7×10^{-4}(Rosy)[d]	
	2.7×10^{-6}(Maroon-like)[d]	
	4.0×10^{-7}(Body- and eye-color)[e]	
H. sapiens	1.7×10^{-5}(PMP22)[f]	
	2.6×10^{-5}(α-globin)[g]	
	1.0×10^{-8}(DMD)[h]	

One rate estimate based on 38 loci is included. All rate measurements are in duplications/gene/generation unless otherwise specified. The loci are listed in parentheses.

[a] *Reams et al. (2010)*
[b] *Anderson and Roth (1981); across 38 loci in overnight culture*
[c] *Gelbart and Chovnick (1979)*
[d] *Shapira and Finnerty (1986)*
[e] *Watanabe et al. (2009)*
[f] *Lupski (2007)*
[g] *Lam and Jeffreys (2007)*
[h] *Van Ommen (2005)*

lac operon in *Escherichia coli* suggested spontaneous duplications rates on the order of 10^{-3} to 10^{-4} per gene (Horiuchi et al., 1963; Langridge, 1969; Anderson and Roth, 1977). More generally, the reported frequency of duplication rates in bacteria and phage for a diversity of genes ranged from 10^{-3} to 10^{-5} (Anderson and Roth, 1977; Starlinger, 1977). The first systematic large-scale study of duplication frequency analyzed 38 duplicated loci in stationary phase cultures of *Salmonella* and found frequencies ranging from 10^{-3} to 10^{-5} per gene (Anderson and Roth, 1981). It should be noted that these estimates do not constitute duplication rates per generation as they had accumulated during the growth of the culture where the duplication rate had been countered by both a high rate of spontaneous duplication loss and natural selection. A more recent analysis of the duplication rate at three loci in the *Salmonella* genome found rates ranging from 2×10^{-3} to 4.6×10^{-6} duplications/gene/generation after carefully controlling for selection and spontaneous duplication loss (Reams et al., 2010). The equilibrium frequency of duplications in culture can likewise be quite high, and high-throughput sequencing of *Salmonella* cultures demonstrated that the percentage of cells carrying duplications had reached a steady-state frequency of 20% (Sun et al., 2012).

EUKARYOTES

Direct estimates of duplication rates at two loci in *D. melanogaster*, the *maroon-like* and the *rosy,* were 2.7×10^{-6} and 1.7×10^{-4} duplications/locus/generation, respectively (Gelbart and Chovnick, 1979; Shapira and Finnerty, 1986). More recently, inverse PCR-based methods were used to measure the rates of duplication and deletion of human *α-globin* genes (Lam and Jeffreys, 2006, 2007). The frequencies of spontaneous *α-globin* duplication in sperm were 2.6×10^{-5} and 6.2×10^{-5} in two human males. However, it is possible that the actual duplication rate of *α-globin* genes is in fact higher than reported because the PCR primers used to detect the duplications were designed to detect specific kinds of duplications, and translocated and inverted duplications would not have been detected. Similar methods were used to determine the duplication and deletion rates at four loci in humans and the duplication rate estimates ranged from 1.7×10^{-5} to 8.7×10^{-7} (Turner et al., 2008).

Lastly, Watanabe et al. (2009) screened 1,554 progeny of wild-caught *D. melanogaster* females for spontaneous eye- and body-color mutations and identified five large deletions ranging from 40 to 500 kb. If these deletions originated via unequal crossing-over, the duplications rate should equal the deletion rate. Based on this assumption, the per gene duplication rate was estimated to be 4×10^{-7}/generation, a similar order of magnitude as other empirical per gene duplication rates in *Drosophila* (Watanabe et al., 2009).

These estimates from single loci yield some of the highest estimates of the duplication rate. This may stem from both a sampling bias toward loci with known high duplication rates, and because some of the examples come from loci that are experiencing unequal crossing-over between related genes. For example, analysis of the duplication rate at the *rosy* locus was undertaken after observing that tandem duplications were occurring at an unusually high frequency (Gelbart and Chovnick, 1979). Similarly, *α-globin* gene copy-number polymorphism was well known and particularly common in populations with high exposure to malaria (Lam and Jeffreys, 2006). The high rate of duplications and deletions found in these systems may therefore not be representative of the genome at large.

ESTIMATES OF THE DUPLICATION RATE BASED ON POPULATION FREQUENCY OF CNVs

The duplication rate can also be estimated using the frequency of gene duplications in a population and population-genetic theory of mutation-selection balance. Haldane (1935) showed that for X-linked genes in equilibrium, the mutation rate can be estimated using $1/3(1 - f)x$, where f is the fertility of affected males relative to unaffected males and x is the frequency of affected males in the population. If the X-linked mutation results in lethality or sterility, the mutation rate is estimated as $x/3$. Using this approach, Van Ommen (2005) calculated the rate of new gene duplications in the X-linked human dystrophin gene leading to Duchenne Muscular Distrophy (DMD). Males with DMD have, until recently, been mostly nonreproductive. The frequency of DMD in male newborns is 1:3,500 and the frequency of mutations leading to DMD is thus $\sim 10^{-4}$ (**Table 1**). Subgenic duplications account for 9% of these mutations and the rate of duplication was therefore estimated to be $\sim 10^{-5}$ duplications/DMD locus/generation. The DMD is very large (2.5 Mb) and extrapolating from this region to the whole genome, the genome-wide duplication rate should be 0.02 duplications/genome/generation. This would be an underestimate if (i) many internal duplications do not result in a DMD phenotype, and/or (ii) if duplications that encompass the whole locus do not result in a DMD phenotype.

CMT1A, a subtype of Charcot-Marie-Tooth (CMT) syndrome, frequently results from a large duplication that includes the *PMP22* gene. Based on the prevalence of CMT1A and the fraction of CMT caused by duplications, the spontaneous duplication rate was estimated to be between 1.7 and 2.6×10^{-5} duplications/*PMP22* locus/generation (Lupski, 2007). This rate is very similar to the rate estimated for DMD and three orders of magnitude higher than the spontaneous point mutation rate in humans.

BIOINFORMATICALLY DERIVED ESTIMATES OF THE DUPLICATION RATE FROM WHOLE GENOME SEQUENCES

Lynch and Conery (2000, 2003) pioneered methods for estimating the duplication frequency in sequenced genomes from the age-distribution of duplicated genes based on the synonymous site divergence between gene paralogs. Their analyses found, for example, that duplications arise at a rate of 0.0011, 0.0028, 0.0025 per gene per 1% divergence at synonymous sites in the *D. melanogaster, Caenorhabditis elegans,* and *Saccharomyces cerevisiae* genomes, respectively (Lynch, 2007; **Table 2**). The spontaneous base substitution rate in these species has been measured as 55, 21, and 3.3×10^{-10} mutations/base pair/generation (Haag-Liautard et al., 2007; Lynch et al., 2008; Denver et al., 2009; Keightley et al., 2009; Schrider et al., 2013). If we utilize these rates to convert the historical gene duplication rate to frequency per gene per generation, the duplication rate would be 60.5, 58.8, and 8.25×10^{-11} in *D. melanogaster, C. elegans* and *S. cerevisiae,* respectively. These calculations assume that synonymous site changes are neutral, and

Table 2 | Genome-wide estimates of the duplication rates for prokaryotes and eukaryotes.

Species	Genome-Wide Gene Duplication Rate	
	Bioinformatic	***Empirical***
Unicellular eukaryotes		
E. cuniculi	11.7×10^{-3} per 1% silent-site divergence[a,b]	–
P. falciparum	0.3×10^{-3} per 1% silent-site divergence[a,b]	–
S. cerevisiae	2.5×10^{-3} per 1% silent-site divergence[a,b]	3.4×10^{-6}[d]
	1.0×10^{-11}/gene/year[c]	
S. pombe	1.6×10^{-3} per 1% silent-site divergence[a,b]	–
Multicellular eukaryotes		
A. gambiae	6.2×10^{-3} per 1% silent-site divergence[a,b]	–
A. thaliana	3.2×10^{-3} per 1% silent-site divergence[a,b]	–
C. elegans	2.8×10^{-3} per 1% silent-site divergence[a,b]	3.4×10^{-7}[e]
D. melanogaster	1.1×10^{-3} per 1% silent-site divergence[a,b]	3.7×10^{-7}[f]
F. rubripes	4.3×10^{-3} per 1% silent-site divergence[a,b]	–
H. sapiens	4.9×10^{-3} per 1% silent-site divergence[a,b]	–
	1.1×10^{-9}/gene/year[g]	
M. musculus	3.0×10^{-3} per 1% silent-site divergence[a,b]	–

Estimates are further classified into bioinoformatic versus empirical estimates. Bioinformatic estimates are based on the distribution of evolutionarily young gene duplicates in the genomes of laboratory strains or natural isolates. Empirical estimates are derived from mutation accumulation (MA) experiments involving experimental lines propagated under strict bottlenecking conditions. All rate measurements are in duplications/gene/generation unless otherwise specified. The loci are listed in parentheses.

[a] *Lynch and Conery (2003)*
[b] *Lynch (2007)*
[c] *Gao and Innan (2004)*
[d] *Lynch et al. (2008)*
[e] *Lipinski et al. (2011)*
[f] *Schrider et al. (2013)*
[g] *Cotton and Page (2005)*

in the event that there is some negative selection on synonymous sites, the per generation duplication rates would be overestimated. However, it was noted that the duplication rates inferred from the age distribution of gene duplicates might be underestimates for several reasons. (i) The assembly of whole genome sequences following shotgun sequencing may erroneously assume evolutionarily recent gene duplicates for redundant sequences of single-copy genes (Lynch and Conery, 2003). (ii) This particular analysis did not include paralogs in gene families possessing more than five members. The rates of spontaneous duplication and deletion might increase with the size of a gene family due to greater abundance of regions of high sequence identity that could serve as targets for copy-number changes by unequal exchange.

Gene conversion between duplicate gene copies lowers nucleotide sequence divergence between them, making them appear evolutionarily younger than they actually are (Teshima and Innan, 2004; Katju and Bergthorsson, 2010; Rane et al., 2010). If gene conversion between duplicated genes is common, the number of recent gene duplications in genomes is overestimated under the approach used by Lynch and Conery (2000, 2003). This in turn would lead to an inflated gene duplication rate. Using the genome of *S. cerevisiae* and six of it relatives, Gao and Innan (2004) calculated the gene duplication rate in yeast by a method that does not depend on synonymous site divergence between duplicate copies in a genome. They found strong evidence for gene conversion between duplicate gene copies, and estimated gene duplication rates to be 0.01–0.06 duplications/gene/billion years, two orders of magnitude lower than the previous estimate of Lynch and Conery (2000). However, *S. cerevisiae* with its large effective population size ($N_e = \sim 3.3 \times 10^7$; Lipinski et al., 2011) typically characteristic of unicellular eukaryotes is subject to a strong intensity of natural selection. Hence, the observed number of extant gene duplicates in a sequenced genome may grossly underestimate the gene duplication rate as many gene paralogs may have been purged from the genome in their infancy leaving no signature of their brief existence (Katju et al., 2009; Watanabe et al., 2009; Lipinski et al., 2011; Katju, 2012).

Codon usage bias due to selection for optimal codon use might also confound analyses of gene duplication rates with methods that rely on DNA sequence divergence at synonymous sites (Gu et al., 2002). The rate of molecular evolution in genes that are subject to natural selection against synonymous mutations in preferred codons is slower than at sites where nucleotide substitutions are selectively neutral. Duplicated genes that are experiencing selection for codon usage would therefore appear evolutionarily younger than they are. Gu et al. (2002) therefore suggested comparing DNA sequence divergence at synonymous sites in duplicated genes to sequence divergence in their introns and flanking sequences to exclude genes that appear to have undergone gene conversion or natural selection for codon usage bias. After "cleaning" their database of genes experiencing gene conversion or selection at synonymous sites, Gu et al. estimated the gene duplication rates in *S. cerevisiae*, *D. melanogaster,* and *C. elegans* to be 0.028, 0.0014, and 0.024 duplications/gene/million years, respectively. These results are qualitatively similar to the results of Lynch and Conery (2000, 2003).

More recently, Pan and Zhang (2007) estimated the gene duplication rates in mouse and humans, using synonymous site divergence as a proxy for the age of duplicated genes as some of the previous analyses, and attempting to distinguish between tandem duplications by unequal crossing over and retrotransposition. Their estimates of the overall gene duplication rate ranged from 0.0005 to 0.00149 and from 0.00123 to 0.00423 duplications/gene/million years in humans and mouse, respectively. Bensasson et al. (2003) arrived at similar rates as Lynch and Conery (2000, 2003) based on the number of duplicated mitochondrial genes that have been transferred to the nucleus (NUMTs).

DIRECT GENOME-WIDE ESTIMATES OF THE SPONTANEOUS DUPLICATION RATE FROM MA EXPERIMENTS

Direct empirical analyses of individual loci where gene copy-number differences result in a distinct phenotype or genotype have provided the highest estimates of the gene duplication and deletion rates (Anderson and Roth, 1977, 1981; Shapira and Finnerty, 1986; Lam and Jeffreys, 2007; Watanabe et al., 2009). However, per-locus measures of the duplication/deletion rate may not be

widely applicable at the genome-wide level. Experimental mutation accumulation (MA henceforth) lines in the estimation of mutation rates and parameters. First, they enable the most accurate estimation of mutation rates without the purging influence of purifying natural selection. Second, in conjunction with modern genome-wide techniques of analyses, they serve to directly quantify genome-wide mutation rates with minimal bias. The underlying principle behind MA experiments is straightforward; multiple replicate lines derived from an inbred ancestral stock population are allowed to evolve independently of one another under conditions of extreme bottlenecking each generation. The repeated bottlenecks severely diminish the efficacy of natural selection, promoting evolutionary divergence due to the accumulation of deleterious mutations by random genetic drift. The vast majority of MA studies have maintained the organism at a constant minimal N_e for the purpose of drastically reducing the efficacy of selection and enabling the accumulation of the vast majority of mutations (Mukai, 1964; Ohnishi, 1977; reviewed in Halligan and Keightley, 2009).

The advancement of molecular technologies such as high-throughput genome sequencing and oligonucleotide array comparative genome hybridization (oaCGH henceforth) have enabled genome-wide analyses of DNA content of MA lines to generate the first empirical measures of the spontaneous gene duplication and deletion rate in a handful of model organisms (**Table 2**). Lynch et al. (2008) conducted pulse-field gel electrophoresis (PFGE) and oaCGH on eight *S. cerevisiae* MA lines passaged through 200 bottleneck generations and estimated the spontaneous duplication rate to be 3.4×10^{-6} per gene/generation. This spontaneous duplication rate in *S. cerevisiae* is four orders of magnitude greater than the spontaneous base-substitution rate of 0.33×10^{-9} per site/generation in this species. Moreover, this spontaneous duplication rate vastly exceeds previous estimates arrived at from bioinformatic analyses (Lynch and Conery, 2000; Gao and Innan, 2004) of the originally sequenced *S. cerevisiae* genome (Goffeau et al., 1996). Additionally, the yeast genome originally sequenced by Goffeau et al. (1996) has an extremely low incidence of extant paralogs with low synonymous divergence that originated from small-scale duplication events (Katju et al., 2009). Of this already limited number of paralogs, a substantial number are likely of older evolutionary origin given the high incidence of selection for codon usage bias in conjunction with ectopic gene conversion within this species (Gao and Innan, 2004; Lin et al., 2006). So where are these new paralogs that are spawned at astoundingly high rates? One hypothesis is that most duplicates have, at the minimum, mildly deleterious fitness effects that renders them amenable to rapid purging from the genome in a unicellular eukaryotic species such as *S. cerevisiae* with a high N_e (Katju et al., 2009; Lipinski et al., 2011; Katju, 2012). As such, genome sequences of isolates/strains that have been subject to some degree of natural selection will invariably underestimate the spontaneous rate of duplication.

Lipinski et al. (2011) provided the first empirical, genome-wide estimates of the spontaneous rate of duplication and deletion in a multicellular eukaryote, the nematode *C. elegans*. As in the preceding study with *S. cerevisiae*, long-term MA lines formed the focus of this study to ensure unbiased estimates of the spontaneous rates of gene duplication with minimal influence of natural selection. Ten *C. elegans* MA lines subjected to single-worm bottlenecks for an average of 432 generations were assayed using oaCGH. In total, 14 duplicated segments that comprised *complete* and/or *partial* gene duplications were detected and verified independently via quantitative PCR. These duplicated segments encompassed 30 genes, giving a spontaneous rate of gene duplication of 3.4×10^{-7} per gene/generation for *partial* or *complete* duplications. If only *complete* gene duplicates were considered, the spontaneous rate of gene duplication was 1.25×10^{-7} per gene/generation. The authors argued that this estimate is downwardly biased for two reasons, namely (i) the number of adjacent microarray probes signaling gene copy-number changes may not be sufficiently dense for the detection of duplication events with small duplication spans, and (ii) the oaCGH DNA microarrays were restricted to unique probes only and duplications of genes in recently duplicated regions, for instance by unequal crossing over, may not have been detected. Despite the possibility that this rate is an underestimate, it is two orders of magnitude greater than the *C. elegans* spontaneous base-substitution rate of $\sim 10^{-9}$ per site/generation (Denver et al., 2009). Additionally, this empirical spontaneous duplication rate estimate is two orders of magnitude greater than the estimate calculated from bioinformatic analyses of the frequency distribution of extant paralogs of varying evolutionary age (Lynch and Conery, 2000) in the originally sequenced genome of the *N2* laboratory strain of *C. elegans* (*C. elegans* Sequencing Consortium, 1998).

More recently, Schrider et al. (2013) sequenced the genomes of eight sublines derived from two ancestral lines of a long-term MA experiment in *D. melanogaster*. Despite the use of vastly different technologies for the estimation of the spontaneous duplication rate in *C. elegans* (oaCGH) and *D. melanogaster* (Illumina paired-ends sequencing), the duplication rate estimates are surprisingly similar. Schrider et al. (2013) generated the following rates for *D. melanogaster*: 3.75×10^{-7} per gene/generation for *partial* or *complete* duplications and 1.25×10^{-7} per gene/generation if only *complete* duplications were considered.

ESTIMATES OF THE DELETION RATE

The frequency of gene copy-number polymorphisms in genomes is determined by a combination of the spontaneous duplication/deletion rate and the preservation or elimination of these changes by natural selection. Hence, in conjunction with other evolutionary forces such as selection and genetic drift, the net difference in the spontaneous rates of duplication and deletion has important consequences for the evolution of genome size. Furthermore, duplications and deletions may work in concert with one another. For example, aneuploidy and duplications were common in a collection of random yeast deletion mutants (Hughes et al., 2000). The duplicated regions often contained genes that were related to the deleted genes suggesting that the duplications were compensating for the deletions even though the primary functions of the deleted and duplicated genes are not identical (Hughes et al., 2000). There exists ample evidence that loss-of-function mutations, for example due to gene deletions, can often be suppressed or compensated for by multiple copies, or increased transcription of another gene in the genome (Berg et al., 1988; Bender and Pringle, 1989; Trempy and Gottesman, 1989; Ueguchi and Ito,

1992; Yamanaka et al., 1994; Serebrijski et al., 1995; Timms and Bridges, 1998; Menez et al., 2001; Miller and Raines, 2004; Patrick et al., 2007; Patrick and Matsumara, 2008). This phenomenon is known as "multicopy suppression" and typically results from side-functions of a multicopy gene that go unnoticed when it exists as a single copy in the genome (Berg et al., 1988). On the flip side, deletion events subsequent to duplications can occur commonly and pervasively at the genome-wide level, leading to the "diploidization" of polyploids and the evolution of reproductive incompatibilities (Wolfe, 2001; Kashkush et al., 2002; Langkjaer et al., 2003; Brunet et al., 2006; Scannell et al., 2006; Albertin and Marullo, 2012). Internal deletions of segmental duplications can also play a role in the eventual fate of duplications. Experiments with selected gene amplifications in S*almonella* have revealed that large duplications are frequently followed by internal deletions that appear to facilitate further amplification, by reducing the fitness cost associated with amplification of genes that are not under selection for increase in gene dosage (Kugelberg et al., 2006, 2010).

The gene deletion frequency in bacteria is generally lower than the duplication rate, and ranges from 10^{-4} to 10^{-8} (Starlinger, 1977). Using a combination of sequential bottlenecking of colonies which reduces effective population size and PFGE, experiments in *Salmonella* found the deletion rate to be 0.5×10^{-8} (Nilsson et al., 2005). This is probably an underestimate because there is still selection against deleterious deletions and the PFGE approach only detects relatively large deletions (Nilsson et al., 2005). If many deletions resulted in the loss of essential genes, they would not be represented in this estimate. However, if spontaneous gene deletion rates are indeed lower than gene duplication rates in bacteria, then what is keeping bacterial genomes lean? One contributing factor is adaptive gene loss (discussed below). We further need to take into consideration that the evolutionary dynamics of duplications are different from deletions in that duplications are prone to loss through recombination. Hence, the instability of segmental duplications relative to deletions likely serves as a factor in maintaining streamlined bacterial genomes. Lastly, natural selection in large bacterial populations is also expected to be more efficient in eliminating slightly deleterious duplications relative to multicellular eukaryotes with smaller effective population sizes.

Inverse-PCR methods in humans found that the duplication and deletion rates of *α-globin* were very similar. The frequency of deletions in *α-globin* genes can be common in areas where malaria is endemic, and polymorphism for the number of *α-globin* genes is probably maintained by balancing selection involving increased resistance to malaria (Flint et al., 1986). The frequencies of spontaneous *α-globin* deletions in the sperm of two human males were 1.6×10^{-5} and 6.8×10^{-5}. More recently, similar methods were used to determine the duplication and deletion rates at four hotspots in human sperm and the deletion rate estimates ranged from 2.2×10^{-5} to 9.5×10^{-6}, with all deletion rate estimates exceeding the duplication rates by 2.1 to 4.1 fold (Turner et al., 2008). The population frequency of CNVs resulting in DiGeorge-Velo cardiofacial syndrome, Williams-Beuren syndrome and Smith-Magenis syndrome have been used to estimate the spontaneous deletion rate in humans. The estimated rates range from 2×10^{-5} to 1.25×10^{-4} deletions/locus/generation (Lupski, 2007). Loss of gene duplication occurs generally at a higher rate than the duplication rate. For example, loss of the *bar* duplication in *D. melanogaster* may occur at a rate as high as 10^{-3} (Sturtevant, 1925).

Genome-wide estimates of the spontaneous deletion rates are currently available for three species: *S. cerevisiae* (Lynch et al., 2008), *C. elegans* (Lipinski et al., 2011) and *D. melanogaster* (Schrider et al., 2013). The spontaneous deletion rates were 2.1×10^{-6}, 2.2×10^{-7}, and 9.37×10^{-7}/gene/generation in *S. cerevisiae*, *C. elegans,* and *D. melanogaster*, respectively. In *S. cerevisiae* and *C. elegans*, there appears to be a slight excess of duplications relative to deletions when considered on a gene-by-gene basis, whereas the deletion rate exceeded the duplication rate in the *D. melanogaster* experiment. However, deletions tend to be smaller than duplications and the net change in base pairs is positive in all three experiments. That is, nucleotides added by duplications exceed those deleted.

FITNESS EFFECTS OF CNVs

The scientific literature is replete with descriptions of gene duplications that are either beneficial or detrimental to the fitness of their carriers. On the beneficial side, some of the most striking examples in humans include the copy-number increase of the human salivary amylase gene (*AMY1*) that have enabled adaptation to a high-starch diet (Perry et al., 2007) and copy-number increase of the *CCL3L1* gene that is associated with lowered susceptibility to HIV infection (Gonzalez et al., 2005). Interestingly, the domestication of dogs by humans too has resulted in a copy-number increase in the canid *amylase* gene, enabling dogs to benefit from a high-starch diet that is distinctly human and contrasting from their wolf ancestors (Axelsson et al., 2013). Copy-number increases are also implicated in adaptation to novel or resource-limited environments in microbial laboratory populations (Sonti and Roth, 1989; Reams and Neidle, 2003), insecticide resistance (Newcomb et al., 2005) or metal tolerance (Maroni et al., 1987) in natural insect populations, drug resistance in parasites (Nair et al., 2007), increased vertebrate resistance to bacterial pathogens (Jackson et al., 2007) and as a compensatory response to loss-of-function mutations (Berg et al., 1988; Bender and Pringle, 1989; Trempy and Gottesman, 1989; Ueguchi and Ito, 1992; Yamanaka et al., 1994; Serebrijski et al., 1995; Timms and Bridges, 1998; Menez et al., 2001; Miller and Raines, 2004; Patrick et al., 2007).

However, most gene duplications are probably deleterious. The detrimental consequences of duplications can come from a variety of sources: (i) dosage imbalance between the duplicated genes and other genes in the genome that remain in single copy, (ii) inappropriate expression of gene duplicates that are under the control of a different regulatory system, and (iii) the cost of superfluous expression. From the perspective of the deleterious nature of gene duplications, increases in gene copy-number are implicated in increased susceptibility to a wide range of human diseases (Lupski, 1991, 1998; Inoue and Lupski, 2002 and references therein; Botstein and Risch, 2003; Sebat et al., 2007). Several additional lines of evidence support the notion that gene duplications are, on average, deleterious. First, the large discrepancy in empirical (from MA experiments) and bioinformatics-based estimates of

the gene duplication rate is best explained by selection against new duplications (Katju et al., 2009; Lipinski et al., 2011). Bioinformatically based methods to determine the duplication rate from the age distribution of genes in a sequenced genome assume a constant loss rate for duplicate genes. However, if selection against duplicate copies in their infancy removes most detrimental gene duplicates before they can diverge at the DNA sequence level, the loss rate may appear to be constant, and yet result in an underestimate of the spontaneous duplication rate. Second, population variation in gene copy-number also suggests that duplications are generally detrimental. In natural populations of *D. melanogaster*, the allele frequencies of duplications are lower than expected if the duplications are neutral (Langley et al., 2012), although not all studies can reject the null hypothesis of no fitness consequences of completely duplicated genes (Emerson et al., 2008). Third, there is a negative correlation between allele frequencies of duplicates and recombination rates, which is consistent with the notion that greater efficacy of natural selection associated with higher recombination rates is eradicating duplicates at a greater rate from regions of high recombination relative to regions of low recombination (Langley et al., 2012). A significant negative association between the length of the duplicated segment and gene density with allele frequencies in humans and *Drosophila* (Itsara et al., 2010; Langley et al., 2012) suggests that duplications encompassing more genes are more deleterious than those spanning fewer genes. This is expected if dosage imbalance plays a large role in determining the fitness cost of duplications.

Deletions, like duplications, can be either detrimental or adaptive. Examples of adaptive deletions are more limited relative to adaptive duplications and it is generally assumed that deletions are, on average, more detrimental than duplications. Several genome-wide studies of copy-number variation in humans have found deletion alleles to occur in lower frequencies than duplication alleles (Conrad et al., 2006; Locke et al., 2006). This is suggestive of strong purifying selection weeding out deletions. Furthermore, a deficit of genic deletions has been observed in humans (Conrad et al., 2006, 2010; Redon et al., 2006) and *D. melanogaster* (Emerson et al., 2008; Langley et al., 2012), implying that deletions in coding sequences are more deleterious than duplications of these sequences, and therefore more likely to be purged by purifying selection. Conrad et al. (2010) compared the relative frequencies of deletions in two additional genomic regions, namely intronic and intergenic. Intergenic deletions outnumbered intronic deletions, suggesting stronger selection against the latter, given their central role in the maintenance of accurate intronic sequence for splicing (Conrad et al., 2010). This might also explain why the frequency of spontaneous deletions appears lower than duplications in MA experiments in yeast and *C. elegans* (Lynch et al., 2008; Lipinski et al., 2011). Although MA experiments can capture a wide range of deleterious mutations, mutations with severe fitness consequences are still less likely to be fixed than mutations with minor and moderate fitness costs.

Nonetheless, deletions have played an important role in adaptation. For example, a recurrent deletion of an enhancer for *Pitx1* in sticklebacks is associated with adaptive pelvic reduction (Chan et al., 2010). Adaptive deletions might be more common than we assume. In experiments with *Salmonella*, a surprisingly high proportion of deletions resulted in increased growth rate, which suggests that many bacterial genes are not necessary, and indeed a burden, in a specific laboratory environment (Koskiniemi et al., 2012). Parallel patterns of gene loss have been seen in bacteria, for example, during infection or host adaptation and although it is tempting to ascribe these to adaptive gene loss, these patterns can, in principle, also be explained by relaxation of selection on the lost genes (Feng et al., 2011; Rau et al., 2012). However, many studies of bacterial genome evolution suggest that gene loss is often adaptive. For example, the removal of pseudogenes from *Salmonella* genomes occurs at a faster rate than expected if the gene loss is purely neutral (Kuo and Ochman, 2010). The question of whether deletions are beneficial or neutral is easiest to address in an experimental setting rather than by retrospective analysis. In experiments with *Methylobacterium*, Lee and Marx (2012) found that repeated gene loss was adaptive, and the benefit from the deletions was not due to a shorter genome *per se*. The frequent and parallel patterns of gene loss in bacterial genomes recently inspired the Black Queen Hypothesis, which suggests that the evolution of dependencies in microbes resulted from selection against genes whose products can be acquired from other organisms (Morris et al., 2012).

THE ROLE OF N_e IN DICTATING CNV LOSS OR FIXATION

The loss or fixation of CNVs and their consequences for population fitness depend upon both (i) the selection coefficients (s) associated with individual duplications/deletions, and (ii) the effective population size (N_e) for the species. The fate of duplications/deletions with selection coefficients much less than the reciprocal of the N_e [$|s| \ll 1/2N_e$ for diploids] are expected to be dictated entirely by random genetic drift. Conversely, the dynamics of duplications/deletions with $|s| \gg 1/2N_e$ are governed by natural selection. Deleterious duplications and deletions with very large deleterious effects will be rapidly eradicated from the population and unlikely to reach fixation; those with very small effects would be effectively neutral. Although the effect of any mutation is dependent on the N_e, the prevailing opinion is that the most detrimental class of mutations influencing long-term population fitness includes mutations with small selection coefficients, also referred to as slightly deleterious or nearly neutral mutations (Ohta, 1992). Such mutations would be eradicated via purifying selection at high N_e, but can behave in an "effectively neutral" fashion and reach fixation by genetic drift at low N_e (Lynch and Gabriel, 1990; Lande, 1994).

Empirical estimates of the spontaneous duplication rate, be they locus-specific or genome-wide from MA studies, invariably exceed estimates from analyzing the age distribution of gene duplicates in sequenced genomes. What may explain this discrepancy, with empirical estimates exceeding bioinformatically based ones by two to four orders of magnitude? We have previously proposed that the degree of discrepancy in bioinformatic and empirical estimates of the gene duplication rate is influenced by differences in the efficacy of selection in species due to their varying N_e (Katju et al., 2009; Lipinski et al., 2011; Katju, 2012). Specifically, slightly deleterious CNVs will be efficiently weeded out in species with large N_e but are more likely to survive the onslaught of purifying

selection in species with small N_e. Currently, bioinformatic and spontaneous empirical estimates of the gene duplication rate are only available for three species, *S. cerevisiae*, *D. melanogaster* and *C. elegans* with estimated N_e of 3.3×10^7, $1.15 \times 10^6$and 80,000 individuals, respectively (Lipinski et al., 2011; Katju, 2012 and references therein). The empirical estimates of the duplication rate exceed the bioinformatic estimates by 36,000-, 660-, and 340-fold for *S. cerevisiae*, *D. melanogaster*, and *C. elegans*, respectively. This discrepancy correlates positively with the species N_e as we have previously predicted (Lipinski et al., 2011). A more robust test of this hypothesis will require greater sampling of the empirical genome-wide duplication rates across more species.

CONCLUDING REMARKS

Gene CNVs are of fundamental importance for genetic variation in populations, genome evolution and the evolution of genes with novel functions. When the first genome-wide estimates of the spontaneous duplication rate were bioinformatically determined from sequenced genomes, they were reported as being similar to the point mutation rates (Lynch and Conery, 2000). These rates were hailed as being "astronomical" (Pennisi, 2000). Direct empirical estimates of spontaneous duplication rates derived from experimental MA lines have been demonstrated to be orders of magnitude higher. The discrepancy between the bioinformatically derived and empirical duplication rates suggests that the vast majority of gene duplications are deleterious and rapidly eradicated from genomes before being afforded any opportunity to impart a genomic signature of their all too brief existence. This discrepancy between bioinformatically and empirically derived estimates of the duplication rate also appears to be positively correlated with the species N_e. Prokaryotes and unicellular eukaryotes with large N_e and greater efficacy of selection are expected to rapidly purge even mildly deleterious duplicates. Conversely, in organisms with small N_e such as many multicellular eukaryotic species, genetic drift is expected to play an integral role in the accumulation of gene duplicates leading to the eventual preservation of duplicates following functional divergence.

The last decade or so has witnessed a revolution in the cataloging of structural variants in species, both at the population- and genomic-level. Structural variants, however, present multiple challenges in the analysis of their dynamics in populations and the evolutionary forces responsible for their ultimate fate in genomes. Whereas standard population-genetic theory is well-equipped to analyze the frequency of alleles or base substitutions in populations, CNVs of particular genes can have breakpoints in different locations, and duplicated genes can have additional variation with respect to genomic location and transcriptional orientation, all of which can differentially influence their function. In this review, we have not tackled issues relating to the structural complexity of CNVs. Gene duplicates, for example, exhibit varying degrees of structural resemblance to their progenitor loci (Katju and Lynch, 2003, 2006; Katju, 2012). An advanced understanding of how the structural resemblance between paralogs influences their eventual fate (pseudogenization, subfunctionalization, or neofunctionalization) must precede and is germane to elucidating the full contribution of CNVs to genome evolution.

Although most CNVs appear to be selected against, we need more information about their distribution of fitness effects, and what particular aspects of their genomic and molecular structure underlie these phenotypic fitness costs/gains. Are duplication and deletion rates species-specific and if so, do these show a dependence on the structural features of a genome, say the fraction of repetitive sequences within a genome? Furthermore, how do these high rates of duplication influence the fate of duplicated genes in populations via natural selection or genetic drift. One consequence of a high duplication rate is that adaptive variation in gene dosage can frequently arise by duplications. One of the important questions regarding the evolution of novel genes is how often this kind of selection for higher gene dosage results in functional divergence, for example, because of adaptive enhancement of subfunctions or promiscuous activity. Or is selection for gene dosage just a temporary response to ephemeral environmental challenges and do duplicates revert back to existence in single-copy form when these challenges no longer exist?

ACKNOWLEDGMENT

The authors would like to thank reviewers Frederic Brunet and Ben Evans and guest Editor Frederic Chain for valuable comments. This work was supported by a National Science Foundation (NSF) grant DEB-0952342 to Ulfar Bergthorsson and Vaishali Katju

AUTHOR CONTRIBUTIONS

Vaishali Katju and Ulfar Bergthorsson contributed equally to all aspects of designing and writing this manuscript.

REFERENCES

Albertin, W., and Marullo, P. (2012). Polyploidy in fungi: evolution after whole-genome duplication. *Proc. R. Soc. B.* 279, 2497–2509. doi: 10.1098/rspb.2012.0434

Anderson, R. P., and Roth, J. R. (1977). Tandem genetic duplications in phage and bacteria. *Annu. Rev. Microbiol.* 31, 473–505. doi: 10.1146/annurev.mi.31.100177.002353

Anderson, P., and Roth, J. (1981). Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. *Proc. Natl. Acad. Sci. U.S.A.* 78, 3113–3117. doi: 10.1073/pnas.78.5.3113

Axelsson, E., Ratnakumar, A., Arendt, M.-J., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. doi: 10.1038/nature11837

Bailey, J. A., and Eichler, E. E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* 7, 552–564. doi: 10.1038/nrg1895

Bender, A., and Pringle, J. R. (1989). Multicopy suppression of the cdc24 budding defect in yeast by CDC42 and three newly identified genes including the ras-related gene RSR1. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9976–9980. doi: 10.1073/pnas.86.24.9976

Bensasson, D., Feldman, M. W., and Petrov, D. A. (2003). Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J. Mol. Evol.* 57, 343–354. doi: 10.1007/s00239-003-2485-7

Berg, C. M., Wang, M. D., Vartak, N. B., and Liu, L. (1988). Acquisition of new metabolic capabilities: multicopy suppression by cloned transaminase genes in *Escherichia coli* K-12. *Gene* 65, 195–202. doi: 10.1016/0378-1119(88)90456-8

Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17004–17009. doi: 10.1073/pnas.0707158104

Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat. Genet.* 33, 228–238. doi: 10.1038/ng1090

Bridges, C. B. (1935). Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J. Hered.* 26, 60-64.

Bridges, C. B. (1936). The bar "gene" – a duplication. *Science* 83, 210–211. doi: 10.1126/science.83.2148.210

Brown, C. J., Todd, K. M., and Rosenzweig, R. F. (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* 15, 931–942. doi: 10.1093/oxfordjournals.molbev.a026009

Brunet, F. G., Crollius, H. R., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., et al. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol. Biol. Evol.* 23, 1808–1816. doi: 10.1093/molbev/msl049

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

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal G. Jr., Shapiro, M. D., Brady, S. D., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305. doi: 10.1126/science.1182213

Clark, A. G. (1994). Invasion and maintenance of a gene duplication. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2950–2954. doi: 10.1073/pnas.91.8.2950

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., and Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38, 75–81. doi: 10.1038/ng1697

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712. doi: 10.1038/nature08516

Cotton, J. A., and Page, R. D. M. (2005). Rates and patterns of gene duplication and loss in the human genome. *Proc. R. Soc. B.* 272, 277–283. doi: 10.1098/rspb.2004.2969

Deng, C., Cheng, C. H. C., Ye, H., He, X., and Chen, L. (2010). Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21593–21598. doi: 10.1073/pnas.1007883107

Denver, D. R., Dolan, P. C., Wilhelm, L. J., Sung, W., Lucas-Lledó, J. I., Howe, D. K., et al. (2009). A genome-wide view of the *Caenorhabditis elegans* base-substitution mutation processes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16310–16314. doi: 10.1073/pnas.0904895106

Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O., and Long, M. (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* 320, 1629–1631. doi: 10.1126/science.1158078

Evgen'ev, M. B., Zatsepina, O. G., Garbuz, D., Lerman, D. N., Velikodvorskaya, V., Zelentsova, E., et al. (2004). Evolution and arrangement of the hsp70 gene cluster in two closely related species of the virilis group of *Drosophila*. *Chromosoma* 113, 223–232. doi: 10.1007/s00412-004-0312-6

Feng, Y., Chen, Z., and Liu, S.-L. (2011). Gene decay in *Shigella* as an incipient stage of host-adaptation. *PLoS ONE* 6:e27754. doi: 10.1371/journal.pone.0027754

Flint, J., Hill, A. V. S., Bowden, D. K., Oppenheimer, S. J., Sill, P. R., Serjeantson, S. W., et al. (1986). High-frequencies of alpha-thalassemia are the result of natural selection by malaria. *Nature* 321, 744–750. doi: 10.1038/321744a0

Force, A., Lynch, M., Bryan Pickett, F., Amores, A., Yan, Y.-L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary degenerative mutations. *Genetics* 151, 1531–1545.

Gao, L. Z., and Innan, H. (2004). Very low gene duplication rate in the yeast genome. *Science* 306, 1367–1370. doi: 10.1126/science.1102033

Geer, B. W., and Green, M. M. (1962). Genotype, phenotype, and mating behavior of *Drosophila melanogaster*. *Am. Nat.* 96, 175–181. doi: 10.1086/282220

Gelbart, W. M., and Chovnick, A. (1979). Spontaneous unequal exchange in the rosy region of *Drosophila melanogaster*. *Genetics* 92, 849–859.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science* 274, 563–567. doi: 10.1126/science.274.5287.546

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307, 1434–1440. doi: 10.1126/science.1101160

Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P., and Li, W.-H. (2002). Extent of gene duplication in the genomes of *Drosophila*, nematode and yeast. *Mol. Biol. Evol.* 19, 256–262. doi: 10.1093/oxfordjournals.molbev.a004079

Haag-Liautard C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Charlesworth, B., et al. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445, 82–85. doi: 10.1038/nature05388

Haldane, J. B. S. (1932). *The Causes of Evolution*. London: Longmans, Green & Co.

Haldane, J. B. S. (1933). The part played by recurrent mutation in evolution. *Am. Nat.* 67, 5–19. doi: 10.1086/280465

Haldane, J. B. S. (1935). The rate of spontaneous mutation of a human gene. *J. Genet.* 31, 317–326. doi: 10.1007/BF02982403

Halligan, D. L., and Keightley, P. D. (2009). Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 40, 151–172. doi: 10.1146/annurev.ecolsys.39.110707.173437

Hemingway, J., Hawkens, N. J., McCarroll, L., and Ranson, H. (2004). The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem. Mol. Biol.* 34, 653–665. doi: 10.1016/j.ibmb.2004.03.018

Hendrickson, H., Slechta, E. S., Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2002). Amplification-mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2164–2169. doi: 10.1073/pnas.032680899

Hooper, S. D., and Berg, O. G. (2003). On the nature of gene innovation: duplication patterns in microbial genomes. *Mol. Biol. Evol.* 20, 945–954. doi: 10.1093/molbev/msg101

Horiuchi, T., Horiuchi, S., and Novick, A. (1963). The genetic basis of hypersynthesis of betagalactosidase. *Genetics* 48, 157–169.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256, 119–124. doi: 10.1098/rspb.1994.0058

Hughes, T. R., Roberts, C. J., Dai, H. Y., Jones, A. R., Meyer, M. R., Slade, D., et al. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.* 25, 333–337. doi: 10.1038/77116

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

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951. doi: 10.1038/ng1416

Inoue, K., and Lupski, J. R. (2002). Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* 3, 199–242. doi: 10.1146/annurev.genom.3.032802.120023

Itsara, A., Wu, H., Smith, J. D., Nickerson, D. A., Romieu, I., London, S. J., et al. (2010). De novo rates and selection of large copy number variation. *Genome Res.* 20, 1469–1481. doi: 10.1101/gr.107680.110

Jackson, A. N., McLure, C. A., Dawkins, R. L., and Keating, P. J. (2007). Mannose binding lectin (MBL) copy number polymorphism in zebrafish (*D. rerio*) and identification of haplotypes to *L. anguillarum*. *Immunogenetics* 59, 861–872. doi: 10.1007/s00251-007-0251-5

Kashkush, K., Feldman, M., and Levy, A. A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160, 1651–1659.

Katju, V. (2012). In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. *Int. J. Evol. Biol.* 2012, Article ID 341932. doi: 10.1155/2012/341932

Katju, V., and Bergthorsson, U. (2010). Genomic and population-level effects of gene conversion in *Caenorhabditis* paralogs. *Genes* 1, 452–468. doi: 10.3390/genes1030452

Katju, V., Farslow, J. C., and Bergthorsson, U. (2009). Variation in gene duplicates with low synonymous divergence in *Saccharomyces cerevisiae* relative to *Caenorhabditis elegans*. *Genome Biol.* 10, R75. doi: 10.1186/gb-2009-10-7-r75

Katju, V., and Lynch, M. (2003). The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* 165, 1793–1803.

Katju, V., and Lynch, M. (2006). On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* 23, 1056–1067. doi: 10.1093/molbev/msj114

Keightley, P. D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S., and Blaxter, M. L. (2009). Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19, 1195–1201. doi: 10.1101/gr.091231.109

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. New York: Cambridge University Press. doi: 10.1017/CBO9780511623486

Koskiniemi, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787. doi: 10.1371/journal.pgen.1002787

Kugelberg, E., Kofoid, E., Andersson, D. I., Lu, Y., Mellor, J., Roth, F. P., et al. (2006). Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17319–17324. doi: 10.1073/pnas.0608309103

Kugelberg, E., Kofoid, E., Reams, A. B., Andersson, D. I., and Roth, J. R. (2010). The tandem inversion duplication in *Salmonella enterica*: selection drives unstable precursors to final mutation types. *Genetics* 185, 65–80. doi: 10.1534/genetics.110.114074

Kuo, C.-H., and Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. *PLoS Genet.* 6:e1001050. doi: 10.1371/journal.pgen.1001050

Lam, K.-W. G., and Jeffreys, A. J. (2006). Processes of copy-number change in human DNA: the dynamics of alpha-globin gene deletion. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8921–8927. doi: 10.1073/pnas.0602690103

Lam, K.-W. G., and Jeffreys, A. J. (2007). Processes of de novo duplication of human α-globin genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10950–10955. doi: 10.1073/pnas.0703856104

Lande, R. (1994). The risk of population extinction from new deleterious mutations. *Evolution* 48, 1460–1469. doi: 10.2307/2410240

Langkjaer, R. B., Cliften, P. F., Johnston, M., and Piskur, J. (2003). Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. *Nature* 421, 848–852. doi: 10.1038/nature01419

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y.-C. G., Schrider, D. R., Pool, J. E., et al. (2012). Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192, 533–598. doi: 10.1534/genetics.112.142018

Langridge, J. (1969). Mutations conferring quantitative and qualitative increases in beta-galactosidase activity in *Escherichia coli*. *Mol. Gen. Genet.* 105, 74–83. doi: 10.1007/BF00750315

Lee, M.-C., and Marx, C. J. (2012). Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet.* 8:e1002651. doi: 10.1371/journal.pgen.1002651.

Lin, Y.-S., Byrnes, J. K., Hwang, J.-K., and Li, W.-H. (2006). Codon-usage bias versus gene conversion in the evolution of yeast duplicate genes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14412–14416. doi: 10.1073/pnas.0606348103

Lipinski, K. J., Farslow, J. C., Fitzpatrick, K. A., Lynch, M., Katju, V., and Bergthorsson, U. (2011). High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Curr. Biol.* 21, 306–310. doi: 10.1016/j.cub.2011.01.026

Locke, D. P., Sharp, A. J., McCarroll, S. A., McGrath, S. D., Newman, T. L., Cheng, Z., et al. (2006). Linkage disequilibrium and heritability of copy-number polymorphisms within duplication regions of the human genome. *Am. J. Hum. Genet.* 79, 275–290. doi: 10.1086/505653

Lupski, J. R. (1991). DNA duplication associated with Charcot-Marie-Tooth disease Type 1A. *Cell* 66, 219–232. doi: 10.1016/0092-8674(91)90613-4

Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* 14, 417–422. doi: 10.1016/S0168-9525(98)01555-8

Lupski, J. R. (2007). Genomic rearrangements and sporadic disease. *Nat. Genet.* 39, S43–S47. doi: 10.1038/ng2084

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

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Lynch, M., and Conery, J. S. (2003). The evolutionary demography of duplicate genes. *J. Struct. Funct. Genomics* 3, 35–44. doi: 10.1023/A:1022696612931

Lynch, M., and Gabriel, W. (1990). Mutation load and the survival of small populations. *Evolution* 44, 1725–1737. doi: 10.2307/2409502

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., et al. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9272–9277. doi: 10.1073/pnas.0803466105

Maroni, G., Wise, J., Young, J. E., and Otto, E. (1987). Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. *Genetics* 117, 739–744.

Maydan, J. S., Flibotte, S., Edgley, M. L., Lau, J., Selzer, R. R., Richmond, T. A., et al. (2007). Efficient high-resolution deletion discovery in *Caenorhabditis elegans* by array comparative genomic hybridization. *Genome Res.* 17, 337–347. doi: 10.1101/gr.5690307

Menez, J., Remy, E., and Buckingham, R. H. (2001). Suppression of thermosensitive peptidyltRNA hydrolase mutation in *Escherichia coli* by gene duplication. *Microbiol.* 147, 1581–1589.

Miller, B. G., and Raines, R. T. (2004). Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. *Biochem.* 43, 6387–6392. doi: 10.1021/bi049424m

Morgan, T. H. (1916). A Critique of the Theory of Evolution. Princeton, NJ: Princeton University Press.

Morgan, T. H. (1925). *Evolution and Genetics.* Princeton, NJ: Princeton University Press.

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3, e00036-12. doi: 10.1128/mBio.00036-12

Mukai, T. (1964). The genetic structure of natural populations of Drosophila melanogaster. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50, 1–19.

Müller, H. J. (1935). The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetica* 17, 237–252. doi: 10.1007/BF01985012

Müller, H. J. (1936). Bar duplication. *Science* 83, 528–530. doi: 10.1126/science.83.2161.528-a

Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A. C., et al. (2007). Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.* 24, 562–573. doi: 10.1093/molbev/msl185

Nasvall, J., Sun, L., Roth, J. R., and Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification and divergence. *Science* 338, 384–387. doi: 10.1126/science.1226521

Newcomb, R. D., Gleeson, D. M., Yong, C. G., Russell R. J., and Oakeshott, J. G. (2005). Multiple mutations and gene duplications conferring organophosphorus insecticide resistance have been selected at the Rop-1 locus of the sheep blowfly, *Lucilia cuprina*. *J. Mol. Evol.* 60, 207–220. doi: 10.1007/s00239-004-0104-x

Nilsson, A. I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. D., and Andersson, D. I. (2005). Bacterial genome size reduction by experimental evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12112–12116. doi: 10.1073/pnas.0503654102

Ohnishi, O. (1977). Spontaneous and ethyl methanesulfate-induced mutations controlling viability in *Drosophila melanogaster*. I. Recessive lethal mutations. *Genetics* 87, 519–527.

Ohno, S. (1970). *Evolution By Gene Duplication.* New York: Springer-Verlag.

Ohta, T. (1988). Time for acquiring a new gene by duplication. *Proc. Natl. Acad. Sci. U.S.A.* 85, 3509–3512. doi: 10.1073/pnas.85.10.3509

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286. doi: 10.1146/annurev.es.23.110192.001403

Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics* 163, 1519–1526.

Pan, D., and Zhang, L. (2007). Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. *Genome Biol.* 8, R158. doi: 10.1186/gb-2007-8-8-r158

Papp, B., Pal, C., and Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197. doi: 10.1038/nature01771

Patrick. W. M., and Matsumara, I. (2008). A study in molecular contingency: glutamine phosphoribosylpyrophospate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* 377, 323–336. doi: 10.1016/j.jmb.2008.01.043

Patrick W. M., Quandt, E. M., Swartzlander, D. B., and Matsumara, I. (2007). Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* 24, 2716–2722. doi: 10.1093/molbev/msm204

Pennisi, E. (2000). Twinned genes live life in the fast lane. *Science* 290, 1065–1066. doi: 10.1126/science.290.5494.1065a

Perry, G. H., Dominy, N. J., Claw, K. W., Lee, A. S., Fiegler, H., Redon, R., et al. (2007). Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39, 1256–1260. doi: 10.1038/ng2123

Proulx, S. R., and Phillips, P. C. (2006). Allelic divergence precedes and promotes gene duplication. *Evolution* 60, 881–892.

Rane, H. S., Smith, J. M., Bergthorsson, U., and Katju, V. (2010). Gene conversion and DNA sequence polymorphism in the sex-determination gene fog-2 and its paralog ftr-1 in *Caenorhabditis elegans*. *Mol. Biol. Evol.* 27, 1561–1569. doi: 10.1093/molbev/msq039

Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S., and Jelsbak, L. (2012). Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Env. Microbiol.* 14, 2200–2211. doi: 10.1111/j.1462-2920.2012.02795.x

Reams, A. B., Kofoid, E., Savageau, E., and Roth, J. R. (2010). Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* 184, 1077–1094. doi: 10.1534/genetics.109.111963

Reams, A. B., and Neidle, E. L. (2003). Genome plasticity in Acinetobacter: new degradative capabilities acquired by the spontaneous amplification of large

chromosomal segments. *Mol. Microbiol.* 47, 1291–1304. doi: 10.1046/j.1365-2958.2003.03342.x

Redon, R., Ishikawa, S., Fitch, K. R., Feuk. L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454. doi: 10.1038/nature05329

Rokyta, D. R., Joyce, P., Caudle, S. B., and Wichman, H. A. (2005). An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* 37, 441–444. doi: 10.1038/ng1535

Roth, J. R., Benson, N., Galitski, T., Haack, K., Lawrence, J. G., and Miesel, L. (1996). "Rearrangement of the bacterial chromosome: formation and applications," in *Escherichia coli* and *Salmonella: Cellular and Molecular Biology*, ed. F. C. Neidhardt (Washington, DC: American Society for Microbiology Press), 2256–2276.

Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., and Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440, 341–345. doi: 10.1038/nature 04562

Schrider, D. R., Houle D., Lynch, M., and Hahn, M. W. (2013). Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194, 937–954. doi: 10.1534/genetics.113.151670

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., et al. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445–449. doi: 10.1126/science.1138659

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528. doi: 10.1126/science.1098918

Serebrijski, I., Wojcik, F., Reyes, O., and Leblon, G. (1995). Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants. *J. Bacteriol.* 177, 7255–7260.

Shapira, S. K., and Finnerty, V. G. (1986). The use of genetic complementation in the study of eukaryotic macromolecular evolution: rate of spontaneous gene duplication at two loci of *Drosophila melanogaster*. *J. Mol. Evol.* 23, 159–167. doi: 10.1007/BF02099910

Sonti, R. V., and Roth, J. R. (1989). Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* 123, 19–28.

Spofford, J. B. (1969). Heterosis and the evolution of duplications. *Am. Nat.* 103, 407–432. doi: 10.1086/282611

Starlinger, P. (1977). DNA rearrangements in prokaryotes. *Ann. Rev. Genet.* 11, 103–126. doi: 10.1146/annurev.ge.11.120177.000535

Sturtevant, A. H. (1925). The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics* 10, 117–147.

Sturtevant, A. H., and Morgan, T. H. (1923). Reverse mutations of the bar genes correlated with crossing over. *Science* 57, 746–747. doi: 10.1126/science.57.1487.746

Sun, S., Ke, R.-Q., Hughes, D., Nilsson, M., and Andersson, D. I. (2012). Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. *PLoS ONE* 7:e42639. doi: 10.1371/journal.pone.0042639

Teshima, K. M., and Innan, H. (2004). The effect of gene conversion on the divergence between duplicated genes. *Genetics* 166, 1553–1560. doi: 10.1534/genetics.166.3.1553

Theodore, L., Ho, A. S., and Maroni, G. (1991). Recent evolutionary history of the metallothionein gene Mtn in *Drosophila*. *Genet. Res.* 58, 203–210. doi: 10.1017/S0016672300029955

Timms, A. R., and Bridges, B. A. (1998). Reversion of the tyrosine ochre strain *Escherichia coli* WU3610 under starvation conditions depends on a new gene tas. *Genetics* 148, 1627–1635.

Trempy, J. E., and Gottesman, S. (1989). Alp, a suppressor of lon protease mutants in *Escherichia coli*. *J. Bacteriol.* 171, 3348–3353.

Turner, D. J., Miretti, M., Rajan, D., Fiegler, H., Carter, N. P., Blayney, M., et al. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature Genet.* 40, 90–95. doi: 10.1038/ng.2007.40

Ueguchi, C., and Ito, K. (1992). Multicopy suppression: an approach to understanding intracellular functioning of the protein export system. *J. Bacteriol.* 174, 1454–1461.

Van Ommen, G.-J. B. (2005). Frequency of new copy number variation in humans. *Nat. Genetics* 37, 333–334. doi: 10.1038/ng0405-333

Veitia, R. A. (2004). Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics* 168, 569–574. doi: 10.1534/genetics.104.029785

Walsh, B. (2003). Population-genetic models of the fates of duplicate genes. *Genetica* 118, 279–294. doi: 10.1023/A:1024194802441

Watanabe, Y., Takahashi, A., Itoh, M., and Takano-Shimizu, T. (2009). Molecular spectrum of spontaneous de novo mutations in male and female germline cells of *Drosophila melanogaster*. *Genetics* 181, 1035–1043. doi: 10.1534/genetics.108.093385

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2, 333–341. doi: 10.1038/35072009

Yamanaka, K., Ogura, T., Koonin, E. V., Niki, H., and Hiraga, S. (1994). Multicopy suppressors, mssA and mssB, of an smbA mutation of *Escherichia coli*. *Mol. Gen. Genet.* 243, 9–16. doi: 10.1007/BF00283870

Yampolsky, L. Y., and Stoltzfus, A. (2001). Bias in the introduction of variation as an orienting factor in evolution. *Evol. Dev.* 3, 73–83. doi: 10.1046/j.1525-142x.2001.003002073.x

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

Received: 07 October 2013; paper pending published: 27 October 2013; accepted: 18 November 2013; published online: 10 December 2013.

*Citation: Katju V and Bergthorsson U(2013) Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front. Genet. **4**:273. doi: 10.3389/fgene.2013.00273*

This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics.

Copyright © 2013 Katju and Bergthorsson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.