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

Vaishali Katju* and Ulfar Berghthorsson
Department of Biology, University of New Mexico, Albuquerque, NM, USA

INTRODUCTION

The publication of Ohno's "Evolution by Gene Duplication" is fit-
tingly 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 therefore he also placed a great significance on whole-genome duplications. 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

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. Periodic 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.
phenotype were unstable. Although the original experiments on the \textit{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-\textit{bar} mutation from \textit{bar} flies was very high, on the order of approximately $10^{-3}$ per generation (Sturtevant, 1923). These early experiments with the \textit{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 biosinformatic analyses of the age-distribution of extant duplicates in the first sequenced genomes had suggested (Lynch et al., 2008; Lipinski et al., 2011; Schilder 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 on 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 superfluos gene expression (Papp et al., 2003; Veitia, 2004). Empirical estimates of this cost in Salmoherella 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 sub-functions 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 (Spotfiord, 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; Bergbromson et al., 2007). There are certain similarities between some of these proposed mechanisms of selective retention of duplicates. For example, hypotheses (iii), (iv), and (v), 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 eroded 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 rates 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 \( D(D + 1) \), 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, 1923). 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 affecting 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 recently 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 |
|------------------|----------------------------------|
| **Prokaryotes**  |                                  |
| *S. enterica*    | 2.0 x 10^{-3} (LacI)^a(x) 3.2 x 10^{-3} – 5.8 x 10^{-3} duplications per locus b(x) |
|                  | 3.0 x 10^{-4} (LacZ)^c |
|                  | 4.6 x 10^{-4} (PyrD)^c |
| **Multicellular eukaryotes** |                                    |
| *D. melanogaster*| 1.6 x 10^{-3} (BarX)^x 1.7 x 10^{-4} (ReaX)^y |
|                  | 2.7 x 10^{-4} (Marion- |
|                  | like)^a |
|                  | 4.0 x 10^{-4} (Body-and |
|                  | eye-color)^a |
| *H. sapiens*     | 1.7 x 10^{-4} (PMP22)^b(x) 2.6 x 10^{-4} (u-globin)^b(x) 1.0 x 10^{-4} (II-4)^b(x) |

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(Anderson et al. 2003)
- b(Anderson and Roth 1998) across 38 loci in overnight culture
- c(Goldberg and Chovnick 1979)
- d(Shapira and Finnerty 1986)
- e(Reams et al. 2010)
- f(Van Ommen 2005)
- g(Shapiro and Finnerty 1986)
- h(Perkins et al. 2006)
- i(Lam and Jeffreys 2007)
- j(Yampolsky and Stoltzfus 2001)
- k(Lin Ommer 2003)
lac operon in *Escherichia coli* suggested spontaneous duplications rates on the order of 10^{-6} to 10^{-5} 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^{-7} (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^{-2} 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 \times 10^{-3} to 4.6 \times 10^{-6} duplications/gene/generation after carefully controlling for selection and spontaneous duplication loss (Beams 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 ros, were 2.7 \times 10^{-6} and 1.7 \times 10^{-6} duplications/focus/generation, respectively (Gelbart and Chovnick, 1979; Shapira and Finnerty, 1996). 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 \times 10^{-5} and 6.2 \times 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 \times 10^{-5} to 8.7 \times 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 duplication rates would equal the deletion rates. Based on this assumption, the per gene duplication rate was estimated to be 4 \times 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 1/2N = fix, 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/s. Using this approach, Van Ommen (2005) calculated the rate of new gene duplications in the X-linked human dystrophin gene leading to Duchenne Muscular Dystrophy (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^{-8} (Table 1). Subgenic duplications account for 9% of these mutations and the rate of duplication was therefore estimated to be \sim 10^{-3} 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/gene/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 \times 10^{-6} 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 genome 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 5.5, 21, and 3.3 × 10^{-15} mutations/base pair/generation (Haag-Liaudard et al., 2007; Lynch et al., 2008; Denver et al., 2009; Knightley 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 \times 10^{-11} in *D. melanogaster*, *C. elegans* and *S. cerevisiae*, respectively. These calculations assume that synonymous site changes are neutral, and
in the event that there is some negative selection on synonymous sites, the per generation duplication rates would be underestimated. However, it was noted that the duplication rates inferred from the age distribution of gene duplicates might be underestimated 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). This particular analysis did not include paralogs in gene families possessing single-copy genes (Lynch and Conery, 2003). (ii) This particular 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 (Gao 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. Gao 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, Gao 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 sequence 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 0.00105 to 0.00149 and from 0.00123 to 0.00423 duplications/gene/million years in humans and mouse, respectively. Bensasson et al. (2005) 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 Finney, 1986; Lam and Jeffreys, 2007; Watanabe et al., 2009). However, per-locus measures of the duplication/deletion rate may not be

---

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

| Species | Bioinformatic | Empirical |
|---------|--------------|-----------|
| **Unicellular eukaryotes** | | |
| E. coli | $1.17 \times 10^{-3}$ per 1% silent-site divergence | $0.028$ duplications/gene/million years |
| P. falciparum | $0.3 \times 10^{-3}$ per 1% silent-site divergence | $0.0014$ duplications/gene/million years |
| S. cerevisiae | $2.5 \times 10^{-3}$ per 1% silent-site divergence | $0.024$ duplications/gene/million years |
| S. pombe | $1.6 \times 10^{-3}$ per 1% silent-site divergence | $0.00123$ to $0.00423$ duplications/gene/million years |

| **Multicellular eukaryotes** | | |
| A. gambiae | $6.2 \times 10^{-3}$ per 1% silent-site divergence | $0.00105$ to $0.00149$ duplications/gene/million years |
| A. thaliana | $3.2 \times 10^{-3}$ per 1% silent-site divergence | $0.028$ duplications/gene/million years |
| C. elegans | $2.6 \times 10^{-3}$ per 1% silent-site divergence | $0.00123$ to $0.00423$ duplications/gene/million years |
| D. melanogaster | $3.5 \times 10^{-3}$ per 1% silent-site divergence | $0.0014$ duplications/gene/million years |
| F. rubripes | $4.3 \times 10^{-3}$ per 1% silent-site divergence | $0.024$ duplications/gene/million years |
| H. sapiens | $4.9 \times 10^{-3}$ per 1% silent-site divergence | $0.00105$ to $0.00149$ duplications/gene/million years |
| M. musculus | $4.9 \times 10^{-3}$ per 1% silent-site divergence | $0.00123$ to $0.00423$ duplications/gene/million years |

estimates are further classified into bioinformatic 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 lineages propagated under strict bottlenecking conditions. All rate measurements are in duplications/gene/generation unless otherwise specified. The loci are listed in parentheses.

**References:**

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 its 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/million 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).
widely applicable at the genome-wide level. Experimental muta-
tion accumulation (MA henceforth) lines in the estimation of
mutation rates and parameters. First, they enable the most accu-
rate estimation of mutation rates without the purging influence
of natural selection. Second, in conjunction with mod-
ern 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 a single ancestral stock
population are allowed to evolve independently of one another
under conditions of extreme bottlenecks each generation. The
repeated bottlenecks severely diminish the efficacy of natural se-
plection, promoting evolutionary divergence due to the accumu-
lation of deleterious mutations by random genetic drift. The vast major-
ity 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 (Miukai, 1964; Ohtsuki, 1977; reviewed in Halligan and
Keightley, 2009).

The advancement of molecular technologies such as high-
throughput genome sequencing and oligonucleotide array com-
parative genome hybridization (oaCGH henceforth) have enabled
gene copy-number changes in evolution (Berg et al., 1988; Bender and
Pringle, 1989; Trempy and Gottesman, 1989; Ueguchi and Ito,
2000). The duplicated regions often contained genes that were
related to the deleted genes suggesting that the duplications were
important consequences for the evolution of genome size. Further-
more, 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 muta-
tions, 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,
2000).
1992; Yamana et al., 1994; Serberjiski et al., 1995; Timms and Bridges, 1998; Menzel et al., 2001; Miller and Raines, 2004; Patrick et al., 2007; Patrick and Matsushara, 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 may also play a role in the eventual fate of duplications. Experiments with selected gene amplifications in S. cerevisiae 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 $10^{-6}$ (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 $0.5 \times 10^{-5}$ (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 (Fült et al., 1986). The frequencies of spontaneous α-globin deletions in the sperm of two human males were $1.6 \times 10^{-7}$ and $6.8 \times 10^{-7}$. 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 $2.2 \times 10^{-5}$ to $9.3 \times 10^{-5}$, 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 cardiovascular syndromes, Williams-Beuren syndrome and Smith-Magen syndrome have been used to estimate the spontaneous deletion rate in humans. The estimated rates range from $2 \times 10^{-5}$ to $1.25 \times 10^{-4}$ 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 $10^{-7}$ (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 (Schrøder et al., 2013). The spontaneous deletion rates were $2.1 \times 10^{-5}$, $2.2 \times 10^{-7}$, and $9.37 \times 10^{-7}$/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 (AMYL) that have enabled adaptation to a high-starch diet (Perry et al., 2007) and copy-number increase of the CCL11 gene that is associated with lowered susceptibility to HIV infection (Gonzalez et al., 2003). Interestingly, the domestication of dogs by humans too has resulted in a copy-number increase in the canine 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; beams and Needle, 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, Yamana et al., 1994; Serberjiski et al., 1995; Timms and Bridges, 1998; Menzel 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). Bioinformatic-
ically 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 duplica-
tions 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 recom-
bination rates, which is consistent with the notion that greater
efficacy of natural selection associated with higher recombi-
nation 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 (Itaara 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 adap-
tive. 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 allelics to occur in lower frequencies than duplication alle-
les (Conrad et al., 2006; Locke et al., 2006). This is suggestive of
strong purifying selection weeding out deletions. Furthermore,
a deficit of genomic deletions has been observed in humans (Con-
rad et al., 2006, 2010; Redon et al., 2006) and D. melanogaster
(Emerson et al., 2008; Langley et al., 2012), implying that dele-
tions in coding sequences are more deleterious than dupli-
cations of these sequences, and therefore more likely to be
purged by purifying selection. Conrad et al. (2009) com-
pared 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 (Con-
rad 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 adapta-
tion. For example, a recurrent deletion of an enhancer for Ptc1 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 (Koskimies 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 pat-
terns 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 Methylbacterium, Lee and Marx (2012) found
that repeated gene loss was adaptive, and the benefit from the dele-
tions 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 Ns IN DICTATING CNV LOSS OR FIXATION

The loss or fixation of CNVs and their consequences for pop-
ulation fitness depend upon both (i) the selection coefficients
associated with individual duplications/deletions, and (ii) the
effective population size (N_e) for the species. The fate of dupli-
cations/deletions with selection coefficients much less than the
reciprocal of the N_e [1/d < 1/(2N_e) for diploids] are expected to be
dictated entirely by random genetic drift. Conversely, the dynamics
of duplications/deletions with 1/d > 1/(2N_e) are governed by nat-
ural selection. Deleterious duplications and deletions with very
large deleterious effects will be rapidly eradicated from the popu-
lation and unlikely to reach fixation; those with very small effects
would be effectively neutral. Although the effect of any muta-
tion is dependent on the N_e, the prevailing opinion is that the
detrimental class of mutations influencing long-term popu-
lation 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, 1999; 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 dupli-
cates 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 esti-
mates of the gene duplication rate is influenced by differences in
the efficacy of selection in species due to their varying
effective population size (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


An advanced understanding of how the structural resemblance only available for three species, S. cerevisiae, D. melanogaster and C. elegans with estimated $N_e$ of $3.3 \times 10^6$, $1.15 \times 10^6$ and 80,000, 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 would 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; 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 sub-functions 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 Chian for valuable comments. This work was supported by a National Science Foundation (NSF) grant DEB-0952524 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

Alberts, W., and Marullo, P. (2012). Polyphyletic fungi: evolution after whole-genome duplication. Proc. R. Soc. B 279, 2507–2518. doi: 10.1098/rspb.2012.0634
Anderson, R. P., and Roth, J. R. (1977). Tandem genetic duplications in phage and bacteria. Ann. Rev. Microbiol. 31, 475–505. doi: 10.1146/annurev.mi.31.100177.002353
Anderson, P., and Roth, J. (1983). Spontaneous tandem genetic duplications in Saccharomyces cerevisiae arise by unequal recombination between DNA (ori) cistrons. Proc. Natl. Acad. Sci. U.S.A. 78, 3113–3117. doi: 10.1073/pnas.78.5.3113
Anderson, S., Kamakar, A., Arndt, M.-J., Magoshi, K., Wittenberg, M. T., Perlodis, 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
Bakker, L. A., and Eshkabilov, E. E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. Nat. Rev. Genet. 7, 552–564. doi: 10.1038/nrg1895
Benda, 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 RSH1. Proc. Natl. Acad. Sci. U.S.A. 86, 9976–9980. doi: 10.1073/pnas.86.24.9976
Bennison, D., Feldman, M. W., and Peter, D. A. (2003). Rates of DNA duplication and microsatellite DNA insertion in the human genome. J. Mol. Evol. 57, 345–354. doi: 10.1007/s00239-003-2485-7
Berg, C. M., Wang, M. D., Varsak, N. R., and Liu, L. (1998). Acquisition of new metabolic capabilities: multicopy suppression by cloned transaminase genes in Escherichia coli K-12. Gene 205, 195–202. doi: 10.1016/0378-1119(98)00438-8
Bergthorsson, U., Anderson, 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, 17084–17089. doi: 10.1073/pnas.0707581104
Bentzen, 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/ng1080
Bridges, C. B. (1935). Salivary chromosome maps with a key to the banding of the chromosomes of Drosophila melanogaster. J. Hered. 26, 60–64.
chromosomal segments. Mol. Microbiol. 47, 1291–1304. doi: 10.1046/j.1365-2958.2003.04562.x

Rudin, R., Ishikawa, S., Frish, K. R., Fooks, L., Perry, G. H., Androwos, T. D., et al (2006). Global variation in copy number in the human genome. Nature 444, 444–454. doi: 10.1038/nature05229

Rokitis, D. R., Joyce, P., Caule, S. B., and Wellsman, H. A. (2005). An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet. 37, 461–444. 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 application,’ in Escherichia coli and Salmonella: Cellular and Molecular Biology, ed F. C. Neunhardt (Washington, DC: American Society for Microbiology Press), 2256–2276.

Scannell, D. R., Byrne, K. P., Wong, S., and Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in prokaryotic yeasts. Nature 440, 341–345. doi: 10.1038/nature04682

Schrider, D. R., Houle D., Lynch, M., and Hahn, M. W. (2013). Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. Genetics 199, 397–394. doi: 10.1534/genetics.113.157670

Selv, J., Lakhotia, B., Maliketa, D., Troje, I., Lose Martin, C., Wulfs, T., et al. (2007). Strong association of de novo copy number mutations with autism. Science 316, 445–448. doi: 10.1126/science.1138659

Selv, J., Lakhotia, B., Troje, I., Alexander, J., Irving, J., Luminet, P., et al. (2008). Large-scale copy number polymorphisms in the human genome. Science 320, 525–528. doi: 10.1126/science.1180918

Seshagiri, S., Wicaksono, D., Boyo, O., and Loeb, G. (1995). Multicopy suppression by a deletion and osmotic stress-dependent complementation by heterologous proA in proK mutants. J. Bacteriol. 177, 7233–7236.

Shapira, S. K., and Finnerty, V. G. (1986). The use of genetic complementation as an orienting factor in evolution. Genetics 115, 103–126. doi: 10.1086/282611

Simpson, J. R. (1969). Heterosis and the evolution of duplications. An. Nat. 103, 407–432. doi: 10.1096/an69-02613

Stoffel, J. B. (1989). Heterozygosis and the evolution of duplications. An. Nat. 103, 407–432. doi: 10.1096/an69-02613

Stumpf, A. H. (1932). The effects of unequal crossing over at the bar locus in Drosophila. Genetics 17, 147–147.

Sturtevant, A. H., and Morgan, T. H. (1923). Reverse mutations of the bar genes of Drosophila melanogaster. J. Biol. Chem. 2, 159–167. doi: 10.1016/B978-008099910-05010

Sonn, R. V., and Roth, J. R. (1989). Role of gene duplication in the adaptation of Salmonella typhimurium to growth on limiting carbon sources. Genetics 125, 19–28.

Spokos, J. F. (1989). Heterosis and the evolution of duplications. An. Nat. 103, 407–432. doi: 10.1096/an69-02613

Stafstrom, F. (1977). DNA rearrangements in prokaryotes. Annu. Rev. Genet. 11, 103–132. doi: 10.1146/annurev.ge.11.090177.000735

Sturtevant, A. H. (1932). The effects of unequal crossing over at the bar locus in Drosophila. Genetics 17, 147–147.

Sun, S., Kim, 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 consequences of spontaneous mutational events in Drosophila melanogaster. Genetics 167, 1553–1560. doi: 10.1534/genetics.166.3.1553

Theodore, L., Ho, A. S., and Marinis, G. (1991). Recent evolutionary history of the metallothionein gene MinB in Drosophila. Genet. Res. 58, 203–210. doi: 10.1017/S0016672300029935

Timms, A. R., and Bridge, B. A. (1998). Reversion of the tryptophane ochre strain Escherichia coli W3110 under starvation conditions depends on a new gene tas. Genetics 148, 1627–1633

Trempe, J.-E., and Gottman, S. (1989). A suppressor of lon protease mutants in Escherichia coli. J. Bacteriol. 171, 3548–3555.

Turner, D. J., Mottiri, M., Rajen, D., Fugleg, H., Carter, N. P., Blayney, M., et al. (2008). Germline rates of de novo mosaics and duplications causing several genomic disorders. Nature Genet. 40, 96–99. 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, 1436–1441.

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

Vistica, R. A. (2004). Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. Genetics 168, 509–574. doi: 10.1534/genetics.104.029785

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

Yamanaka, K., Ogas, T., Kromov, E. V., Niki, H., and Hiraga, S. (1994). Multicopy suppressors, mood and mood, of an amber mutation of Escherichia coli. Mol. Gen. Genet. 243, 8–16. doi: 10.1007/BF00200370

Yampolsky, L. Y., and Stoltzfus, A. (2001). Bias in the introduction of variation in Drosophila melanogaster. Genetics 157, 1035–1045. doi: 10.1534/genetics.101.003505

Yampolsky, L. Y., and Stoltzfus, A. (2001). Bias in the introduction of variation in Drosophila melanogaster. Genetics 157, 1035–1045. doi: 10.1534/genetics.101.003505

Yampolsky, L. Y., and Stoltzfus, A. (2001). Bias in the introduction of variation in Drosophila melanogaster. Genetics 157, 1035–1045. doi: 10.1534/genetics.101.003505

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; accepted: 30 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. 4:123. doi: 10.3389/fgene.2013.00123

This article is 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.

“fgene-04-00273” — 2013/12/7 — 13:11 — page 12 — #12