Published in final edited form as:

Annu Rev Genet. 2013 ; 47: 307–333. doi:10.1146/annurev-genet-111212-133301.

New Gene Evolution: Little Did We Know

Manyuan Long^{1,*}, Nicholas W. VanKuren^{1,2}, Sidi Chen³, and Maria D. Vibranovski⁴

¹Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60636
mlong@uchicago.edu

²Committee on Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL 60636
nvankuren@uchicago.edu

³Department of Biology and Koch Institute, Massachusetts Institute of Technology, 77
Massachusetts Avenue, Cambridge, MA 02139 chensidi@mit.edu

⁴Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São
Paulo, São Paulo, Brazil 05508 mdv@ib.usp.br

Abstract

Genes are perpetually added to and deleted from genomes during evolution. Thus, it is important to understand how new genes are formed and evolve as critical components of the genetic systems determining the biological diversity of life. Two decades of effort have shed light on the process of new gene origination, and have contributed to an emerging comprehensive picture of how new genes are added to genomes, ranging from the mechanisms that generate new gene structures to the presence of new genes in different organisms to the rates and patterns of new gene origination and the roles of new genes in phenotypic evolution. We review each of these aspects of new gene evolution, summarizing the main evidence for the origination and importance of new genes in evolution. We highlight findings showing that new genes rapidly change existing genetic systems that govern various molecular, cellular and phenotypic functions.

Keywords

evolutionary patterns; evolutionary rates; phenotypic evolution; brain evolution; sex dimorphism; gene networks

*Corresponding author. Manyuan Long, Ph.D., The University of Chicago, Department of Ecology and Evolution, 1101 East 57th Street, Chicago, IL 60637, Zoology 301, Office: (773) 702-0557, Lab: (773) 702-8927, mlong@midway.uchicago.edu.

DISCLOSURE STATEMENT

The authors declare that they have no competing interests.

DEFINITIONS

Fixation: The process by which a mutation spreads to all individuals in a population

Monophyletic group: A group of taxa which shares a common ancestor

Pseudogenes: Genes which are thought to have lost their ability to code for a full-length protein

MSCI model: X chromosome inactivation during spermatogenesis favors relocation of genes involved in spermatogenesis to autosomes

Neofunctionalization: The process by which a new gene acquires a novel function

RNA interference (RNAi): The use of exogenous short hairpin RNAs to direct the degradation specific of mRNAs, thus reducing gene expression and function

BACKGROUND AND HISTORICAL OVERVIEW

Understanding how genes originate and subsequently evolve is crucial for explaining the genetic basis for the origin and evolution of novel phenotypes and, ultimately, biological diversity. Gene origination is thus a widely interesting, yet difficult problem to study. Perhaps unsurprisingly, the peculiar structures, functions, and evolution of evolutionarily new genes has attracted the interests of pioneers in genetics and evolution since the early 20th century. Sturtevant (125) was one of the first to identify a duplicated gene, the *Bar* duplication in *Drosophila melanogaster*, from which Muller (101) developed the first prevalent model of new gene evolution in 1936. Muller predicted that a new duplicate copy of a gene could acquire a novel function and be preserved in the genome, and further that “there remains no reason to doubt the application of the dictum ‘all life from pre-existing life’ and ‘every cell from a pre-existing cell’ to the gene: ‘every gene from a pre-existing gene.’” This early thinking on single gene and whole-chromosome duplications (54) was greatly expanded in the 1970s. Ohno (110) further developed Muller’s model in 1970, and Gilbert (51) proposed an entirely new model of new gene formation in 1978, whereby pieces of unrelated genes can be recombined into new genes, rather than just strictly duplicated. However, experimental work on new genes did not begin until the early 1990s.

A natural method for experimental studies of new genes was proposed in the early 1990s: In order to understand how new genes are formed and evolve, studies must focus on genes that were recently formed because young genes still carry all the signatures of initial evolutionary forces that shaped their origination and evolution of their new structures and functions (83). As genes age, they accumulate mutations which obscure the structural or evolutionary signals from their early history (52, 77). Genes younger than 10–30 million years have not experienced much sequence evolution and are thus a valid system in which to investigate the early evolution of genes and to understand their properties. This idea was first manifested in the discovery of *jingwei*, a 3 MY old gene in two species of African *Drosophila* (81). *Jingwei* revealed several interesting features of new gene evolution, which are now known to be general: 1) recombination of existing genes, leading to a hybrid gene structure, 2) rapid sequence evolution driven by positive selection, and 3) acquisition of new biochemical functions (145, 158).

Today it is clear that new gene origination is a general process in evolution and that species-specific or lineage-specific genes exist in many, if not all, organisms. Gigantic databases of genomic sequences from thousands of species reveal that genomes contain huge numbers and diversity of protein coding genes. For example, the plant *Glycine max* genome encodes more than 50,000 protein coding genes, while the bacterial genome of *Candidatus* *Hodgkinia cicadicola* contains only 189 genes. In addition, the abundance and diversity of non-protein-coding genes is only now beginning to be realized. Even genomes with similar gene numbers can have very different, unrelated genes. These recent data reveal a widespread process of birth and death of genes in organisms, in which new genes enter the genome and old genes are lost. What mechanisms and forces dictate gene birth and death? Particularly, how are new genes and novel functions added to genomes?

In the two decades since the discovery of *jingwei* there have been several hundred additional publications reporting various interesting and significant observations of new genes and new gene functions in many different organisms. Regrettably, we can only choose a few representative publications to sketch several lines of observation that can provide insight into an emerging, global picture of new gene evolution. We will follow the growth of scientific information and underlying ideas and concepts in new gene evolution, beginning by discussing the methods for identifying new genes and mechanistic processes of new gene formation. We will then describe the rates and patterns of new gene origination and evolution that may indicate some rules governing these processes, and discuss the evolutionary forces that act on new genes. Finally, we will review the rapid growth of studies of the phenotypic effects of new genes and their impact on phenotypic evolution.

THE CONCEPT OF NEW GENE ORIGATION

To understand various basic properties of new gene evolution we need to have some conception of the process of new gene origination and an operational definition for the process. This definition will help us explore methods for new gene identification.

The process of new gene origination

New gene origination is a microevolutionary process. A protogene structure is first generated by a mutation in a single germ cell genome. This protogene structure must then spread through the population until it is **fixed**. Various evolutionary forces, such as natural selection and genetic drift, govern the spread of the protogene through the population, thus making protogene fixation a population genetic process. Both before and after fixation, the protogene accumulates mutations that confer on it new structures and beneficial, sometimes novel, functions which are acted on by natural selection. From the point that the protogene carries an optimized function and is fixed in the genome, it will be essentially the same as most other, older genes in the genome and can be considered a new gene. New gene studies typically focus on these first two stages (fixation process and acquisition of a beneficial function) and the consequences of accepted mutations on the sequence, structure and function of the new gene. As the last section of this review will show, these microevolutionary changes produce macroevolutionary changes in traits such as development and brain function.

Interest in new gene origination has raised several general problems. What molecular mechanisms generate new gene structures? What are evolutionary forces that drive the origination of new genes? How often are new genes fixed in a species? Are there any rules or patterns of new gene origination? What are the roles of new genes in phenotypic evolution? This review will provide an overview of efforts to understanding these problems.

Approaches to identify new genes

All new gene identification methods are based on comparative analysis of the structures of genes and genomes. Within a group of closely related species, we can define new genes as those that are present in all members of a **monophyletic** group but absent from all outgroup species (Figure 1). Early studies often serendipitously identified new genes by analyzing the

phylogenetic distribution of genomic DNA Southern blot signals or via characterization of small genomic regions, (e.g. 81, 106). Microarrays (41, 43, 44) and especially next-generation sequencing (162, 163) have made recent searches for new genes more purposeful efforts.

Multiple genomes—Syntenic alignments (Figure 1) of genomes can be used to identify new genes from related species for which we know the phylogenetic relationships. Syntenic alignments of each gene in each species allow one to identify genes that are present or absent in one genome relative to another (Figure 1). In these comparisons, a gene can be defined as a new gene candidate if it is present in a certain clade or single species and absent in all outgroup species (Figure 1). Additionally, the orthologous genes that flank the new gene candidate appear in all species under consideration. This strategy has been used with great success in *Drosophila* and mammals (34, 162, 163, 168). New genes formed by different mechanisms also have correspondingly different structural features that can be used to infer the mechanism of new gene formation and the ancestral and derived characters.

Single genomes—Duplicate genes within a single genome can be identified using exhaustive pairwise comparisons between all annotated genes in that genome. Most mechanisms to form new gene structures (see the next section) result in certain structural changes in the new gene. For example, new genes created by RNA-based duplication (retroposition, retrogenes) most often lack introns, add a stretch of adenine nucleotide at their 3' end, and contain a pair of short flanking direct repeats (these signals fade with evolutionary time). Betrán *et al.* (10), Bai *et al.* (3), Meisel *et al.* (98) and Wang *et al.* (144) took advantage of these new structures to identify new retrogenes in fruit flies; Emerson *et al.* (42), Marques *et al.* (92) and Vinckenbosch *et al.* (140) in primates and humans specifically. Divergence between the new retrogene and the original gene from which the retrogene was derived can be used to define the age of the new genes based on molecular clock. However, both strategies that we have discussed so far can depend on the current annotations, which are biased against the newest genes, so caution must be taken when making claims about the presence/absence of genes in different genomes (161).

Predicting functionality of new genes—It is desirable to predict whether candidate new genes are functional before beginning more laborious functional and phenotypic analyses. Comparisons of substitution rates between nonsynonymous and synonymous sites (K_a vs. K_s), polymorphism and divergence (59, 95), open reading frame length, and transcription of new gene candidates are often used to predict whether the new gene is functional. A K_a/K_s ratio significantly lower than 1 (for single genome data, $K_a/K_s < 0.5$ in a comparison between the new gene and its parental copy), for example, indicates functional constraint acting on the new gene, which we would expect if disruptive mutations were being prevented from accumulating in new protein coding genes by natural selection. These methods are widely used as the first step to predict if a new gene is likely functional (e.g. 3, 10, 42, 144, 162, 163).

MECHANISMS TO FORM NEW GENE STRUCTURES

How are new gene structures formed? Mutation toward a new gene structure is the first step of new gene evolution, and at least a dozen distinct molecular processes are known that contribute to the formation of new genes. These mechanisms are covered in depth elsewhere (64, 84), so we will only briefly touch on them here. We highlight several examples in Figure 2.

Gene duplication

Gene duplication is thought to contribute most to the generation of new genes. A single or few new gene structure(s) can be formed at one time by DNA-based duplication (the copying and pasting of DNA sequence from one genomic region to another) or retroposition. While DNA-based duplications are often tandem (130), retroposed genes most often move to a new genomic environment (13, 14, 64, 168), where they must acquire new regulatory elements or risk becoming processed **pseudogenes**. An important gene duplication mechanism is whole genome duplication (WGD), which has occurred multiple times in eukaryote evolution, particularly in plants (122). Hundreds to thousands of duplicate genes are formed by a WGD event, and the vast majority of duplicates are quickly lost. However, estimates of duplicate gene retention after WGDs in teleost fishes (~15 % after 350 MY; 15), yeast (~12% after 80 MY; 67), and *Arabidopsis* (~30 % after 80 MY; 12) all suggest that large fractions of duplicated loci can be retained. We will show below that there are a variety of ways that new gene structures can subsequently acquire new functions (1, 32,60,76,154, 166). McLysaght *et al.* (96) showed that WGD may more easily generate new paralogs.

Alteration of existing gene structures

New gene structures can be generated by modifying existing exons or domains. Gilbert (51) proposed that exons and domains could be recombined to produce new chimeric genes (Figure2A and 2B). Chimeric proteins formed by gene recombination have been found in many organisms since their discovery in the LDL receptor gene (82, 126), including yeast (129), *Drosophila* (81, 116, 117), *Caenorhabditis elegans* (66), mammals (92), and plants (147), and are estimated to have contributed ~19% of new exons in eukaryotes (see 72 and references therein). In addition, retroposed sequences may jump into or near existing genes and recruit existing exons, or be recruited into an existing coding sequence (164). Conversely, new gene structures may be formed by splitting existing genes. Wang *et al.* (146), for example, found that gene duplication is an intermediate stage in an evolutionary process leading to gene fission (Figure 2C). Okamura *et al.* (111) demonstrated that frameshift mutations often generate new coding sequences, finding that 470 human gene duplicates that had done so. Xue *et al.* (153) found that Epstein-Barr virus contains an early gene which undergoes frequent frameshifts, probably to combat host immunity. In addition, divergence in alternative splicing patterns between duplicate genes can generate distinct transcripts that produce noncoding RNAs or polypeptides with slightly or entirely different functions and rapidly alter duplicate gene structures and functions (50, 56,68, 159, 169).

De novo genes

New gene structures may arise from previously non-coding DNA (Figure 2D). Chen *et al.* (23) were the first to show that antifreeze proteins, which bind and halt the growth of ice crystals in the blood of some polar fishes, were created by amplification of microsatellite DNA. Since then, a number of *de novo* genes originating from non-coding regions have been identified in *Drosophila* (5, 25, 73, 162, 168), humans (70, 149, 151, 163), primates (133), murine rodents (102), Protozoa (155), yeast (16, 20), rice (150) as well as virus (118). Similar to strict *de novo* gene origination, horizontal gene transfer (HGT), the exchange of genes between genomes from distantly related taxa, can immediately add new genes and functions to a genome (Figure 2F). HGT is a major mechanism for the addition of new genes to prokaryotic genomes (71, 109), but has also been reported in a number of eukaryotic organisms including plants (7, 157), insects (100) and fungi (55; Figure 2F).

Noncoding RNAs

Not all new genes code for proteins. Noncoding RNAs were found to play an important role in neuronal functions in the early 1990s (132). A large number of functional RNAs from noncoding regions have been reported to play vital roles in a wide variety of organisms (6, 78). MicroRNAs appear to turn over rapidly, but can be strongly influenced by positive selection (87, 88, 107). Strikingly, Dai *et al.* (33) showed that a new long non-coding RNA influences courtship behavior in *D. melanogaster*. Pseudogenes are conventionally thought as dead genes which play no functional roles (40), but may evolve functions in regulating expression of related genes. Zheng *et al.* (167) recently found that many mammalian pseudogenes are transcribed and thus may still function. McCarrey and Riggs (94) predicted that pseudogenes may regulate their parental genes, similar to long non-coding RNAs or miRNAs. An explicit mechanistic model of the use of pseudogene transcripts as decoys for cross-regulating expression of target genes was actually proposed and tested by Marques *et al.* (91) and Marques *et al.* (90).

New gene regulatory systems

New genes must acquire a specific transcription regulatory system to ensure certain temporal and spatial expression patterns. Betrán and Long (9) investigated the origin of the male-specific expression of *Dntf-2r*, a retroposed gene in the *D. melanogaster*-*D. simulans* clade. The new retrogene did not contain the parental promoter, but had acquired a new β 2-tubulin-like promoter by recruiting a novel 5' regulatory sequence. This regulatory sequence drives testis-specific expression of β 2-tubulin, and appears to still do so for *Dntf-2r*. In addition, the new retrogene *Xcbp1* recruited existing neuron promoters present at its site of integration (28). This co-opted mode of promoter recruitment is also observed in human retrogenes (140) and may be a general mode for retrogene promoter gain (64). Additionally, Ni *et al.* (105) observed that eight new genes essential for *Drosophila* development evolved binding sites for the CTCF insulator under positive selection, ensuring the delineation of the regulatory domains of these genes.

Transposable elements

TEs can contribute to functional divergence between duplicate genes in several ways, all similar to those described above (11). For instance, TEs can mediate gene recombination by carrying coding sequences from one part of the genome to another (62, 154), and can even themselves be incorporated into existing coding sequences (45, 86, 104). In addition, TEs were recently found to be a source of micro-RNAs, major components of post-transcriptional regulation of expression (114).

While we still have a developing picture of the contributions of each of these mechanisms to new gene formation in different taxa, work in humans and *Drosophila* suggests that ~80% of genes are formed by DNA-based duplication, 5% - 10% by *de novo* duplication, and ~10% by retroposition (162, 163). And while these mechanisms may generate the initial gene structures, many new structures (in a large variety of taxa) undergo radical structural renovation to change exon-intron structure, or even recruit new or existing coding sequence into the new locus (29, 48, 147, 168).

ABUNDANCE AND ORIGATION RATES OF NEW GENES

The advent of whole genome sequences for many organisms allowed identification of many new DNA-based and RNA-based duplicate genes (e.g. 10, 42). With more genome sequences available, especially in closely-related groups like the twelve *Drosophila* species (31), it became possible to investigate the rates of new gene origination in particular lineages. We will review these findings in *Drosophila*, mammals, and plants. New genes generated by four main mechanisms have been examined: DNA-based duplication, retroposition, *de novo* origination and exon/domain shuffling. There have been no reports of new gene origination rates for mechanisms other than DNA-based duplication, RNA-based duplication, *de novo* origination, and gene recombination. Thus the rates of new gene origination we highlight should be viewed as serious underestimates.

Drosophila

The first estimate of the rate of new gene origination was made for retrogenes in *Drosophila* in 2002 by Betrán *et al.* (10), who identified ~150 retrogenes in *D. melanogaster* (3, 10) that arose after the divergence of the *Drosophila* and *Sophophora* subgenera, ~50 MYs ago. Their estimate of 3 new retrogenes per MYs in the lineage leading to *D. melanogaster* was corroborated by an independent estimation of ~1.5 new retrogenes per MY based on cDNA hybridization against salivary polytene chromosomes in species in the *D. melanogaster* subgroup (~25 MY old; 154). Zhou *et al.* (168) computationally estimated new gene origination rates in the *D. melanogaster* subgroup via DNA-based duplication, retroposition, *de novo* origination and gene recombination to be 5–11 new genes per million years, and found different rates for the four mechanisms. In particular, about 80% of new genes added to the *D. melanogaster* lineage genome were generated by DNA-based duplication. More extensive and detailed analyses of DNA-based and RNA-based duplicates were conducted by Vibranovski *et al.* (138), Meisel *et al.* (98), and Zhang *et al.* (162). Zhang *et al.* (162) analyzed the 12 *Drosophila* genomes and estimated that ~17 duplicate genes per MYs arose

in the *Drosophila* genome. Figure 3A shows the distribution of these new genes on the *Drosophila* phylogeny.

Mammals

Emerson *et al.* (42) and Marques *et al.* (92) identified ~120 retrogenes in the human genome, yielding an estimated retrogene origination rate of 1 retrogene per MYs in the lineage leading to humans. Zhang *et al.* (163) and Zhang *et al.* (160) systematically identified new genes in vertebrates, especially in primates, and showed that the rates of new gene origination are variable in different evolutionary stages of vertebrates (Figure 3B), although 25–30 genes generated by DNA-based and RNA-based duplication and *de novo* arise per MY. Interestingly, this rate is much higher on branches closer to human (65 new genes per MYs in the human lineage alone; 160).

Plants

In contrast to flies and mammals, Zhang *et al.* (165) reported 0.6 retrogenes per MY arose in the *Arabidopsis thaliana* genome, a rate comparable to *Populus* (170), while a microarray-based study in *Arabidopsis* identified 94 new genes created by DNA-based duplication and retroposition (44). Surprisingly, Wang *et al.* (147) found a very high rate of retrogene and chimeric gene origination was found in rice: over 1000 retrogenes were identified in the rice genome, 380 of which evolved chimeric gene structures by recruiting previously existing genes into their gene structures. These authors determined the rate of chimeric gene origination to be 7 per MYs in grass genomes in the lineage leading to rice, 50 times the origination rate of chimeric genes in humans (140), and the highest rate of chimeric gene origination known. In addition, Jiang *et al.* (62) identified over 3000 gene recombinants in rice mediated by Pack-MULE transposable elements. These results suggest a huge potential for protein diversity plant genomes.

Besides these extensive studies in *Drosophila*, mammals, and plants, there have been many valuable investigations of chimeric genes and retrogenes in *Caenorhabditis elegans* (65), fish (24, 48), silkworm (144), and chicken (61).

Copy number variation

Inexpensive whole genome analysis has also made possible the identification of genes at the very earliest stages of their evolution, before fixation. Abundant copy number variation (CNV) has been detected in *Drosophila* (39, 41, 120), humans (46), mouse (53), and *C. elegans* (79). Dopman and Hartl (39), Emerson *et al.* (41), Cardoso-Moreira and Long (19), and Cardoso-Moreira *et al.* (18) identified over 1000 partial and 100 complete gene duplications/deletions in just 15 strains of *D. melanogaster* relative to the reference genome using microarray hybridization. In addition, next-generation sequencing and microarrays have identified over 1200 partial and 600 complete gene duplications/deletions in 179 individual human genomes relative to the reference (99, 121). The recent sequencing of 43 genomes in two *D. melanogaster* populations detected more CNVs including 2588 duplications and 3336 deletions (72). The large number of new genes segregating in populations is just now beginning to be appreciated and investigated further. An active area

of research will be to perform functional and statistical analyses of these new genes to understand their earliest stages of evolution.

In all, these studies have shown that new gene origination rates can differ between taxa, yet are appreciable in all groups studied. These results further strengthen the conclusion that new gene origination is a general evolutionary process.

PATTERNS OF NEW GENE ORIGINATION

Gene traffic in *Drosophila*, humans and other organisms

With the large number of new genes identified in various organisms, researchers were able to investigate statistical patterns of new gene characteristics to explore the mechanistic and evolutionary forces that impact the formation, origination, and evolution of new genes. Betrán *et al.* (10) examined the chromosomal distribution of retrogenes and their parental copies in *D. melanogaster* (Figure 4A). Surprisingly, these authors found a significant excess of autosomal retrogenes derived from X-linked parental genes (XA), and a significant deficiency of retrogenes formed in the opposite direction (AX) or between autosomes (AA). Dai *et al.* (33) further revealed that retrogenes derived from autosomal parental copies tend to locate to the same chromosome as the parental copies. However, 42 out of the 43 retrogenes exhibited XA movement; only one retrogene moved X→X. These two observations clearly reveal a striking pattern of new gene origination in flies: retrogenes derived from X-linked genes prefer to copy into autosomes. This directional movement of new genes is called “gene traffic” (42). These results hold in the 12 sequenced species of *Drosophila* (98, 138) and in *Anopheles gambiae* (4, 134). Interestingly, 90% of X→A retrogenes in *D. melanogaster* are expressed in testis, a significantly higher proportion of testis-expressed genes than average (10), suggesting that the retrogene’s function (in this case, male-beneficial function) can influence its relocation. The symmetric pattern was observed in silkworm, which has ZW sex determination (females are ZW and males ZZ), whereby genes retroposed from Z→A tend to be ovary-expressed (144). Gene traffic appears to be general in *Drosophila* for different mechanisms of new gene formation, as Vibranovski *et al.* (138) also showed that new genes created by DNA-based duplication exhibit the same X→A movement and testis expression. Moreover, the neo-X chromosome, an autosomal chromosome arm that fused to the ancestral X chromosome in the *Drosophila* genus evolution, also shows the same excess of gene traffic (86, 118).

Relative to *Drosophila*, human and mouse revealed similar yet distinct patterns of gene traffic (42). Compared to a neutral expectation based on the chromosomal distribution of processed pseudogenes, which are expected to be evolving neutrally, there is an excess of X→A retrogene movement and most X→A retrogenes exhibit testis expression. However, there is also a significant excess of A→X retrogene movement in human, and these A→X retrogenes exhibit either female expression or unbiased expression. A→A movement is very low in humans (42). The mouse genome shows a very similar pattern. Zhang *et al.* (162) and Zhang *et al.* (160) extended these patterns to all including DNA-based duplicates, retrogenes and *de novo* genes in *Drosophila*, humans and mouse.

Consequences of gene traffic for genome evolution

If gene traffic has been historically important for genome evolution, the majority of testis/male-biased genes should be autosomal, contrary to the previous conclusion that the X was a hotbed for male-biased genes (143). Several microarray-based studies of male-biased genes and their chromosome locations by Ranz *et al.* (115) and Parisi *et al.* (112) in *Drosophila*, Khil *et al.* (69) in mouse, and later by Zhang *et al.* (160) in *Drosophila*, humans and mouse have confirmed this prediction. In *Drosophila*, Zhang *et al.* (162) showed a smooth transition of new male-biased genes from X-linkage to autosomal linkage with evolutionary time.

Models to interpret the causes of gene traffic

In general, models to explain gene traffic, and experimental evaluation of those models, show that natural selection is a major force governing gene traffic, but that mutational processes likely also play a role (35). Meiotic sex chromosome inactivation (MSCI) in the male germline (10, 42, 135, 136), dosage compensation in the heterogametic sex (2, 139), sexual antagonism between male- and female-beneficial genes (21, 124), and meiotic drive (127, 128) have all been implicated in driving gene traffic. The relative role of each of these forces has been hotly debated. MSCI has a strong effect in mammals (69), and experimental evidence for MSCI in *Drosophila* comes from several studies (58, 135, 136). Vibranovski *et al.* (135) showed that genes that are highly expressed in the meiotic phase of spermatogenesis (when the X chromosome is predicted to be inactivated) are significantly enriched on the autosomes. Conversely, genes expressed in the mitotic phases of spermatogenesis are randomly distributed throughout the genome. Other studies suggest reduced expression throughout spermatogenesis, including in the spermatogonia, which also discredits dosage compensation models (97; however, see 137). A clear-cut single cell transcriptome is needed to clarify these issues. Besides the MSCI model, other non-germline-based models, *e.g.* sexual antagonism, are also necessary to interpret the expression of new genes in the male somatic cells although these models need to be rigorously experimentally tested.

Correlation between gene age and expression

Early studies revealed a connection between the expression and the ages of new genes. Betrán and Long (9) showed that *Dntf-2r*, a ~10 MY old gene in the *D. melanogaster* subgroup, is expressed only in testis while its parent *Dntf-2* is expressed ubiquitously. Almost all retrogenes in *Drosophila* appear to have testis expression (33), and to have maintained testis-biased or testis-specific expression independent of age (49). Vinckenbosch *et al.* (140) showed that new human retrogenes are often transcribed in testis and later evolve stronger and more diverse spatial expression patterns, coining the “out of the testis” hypothesis. Whether or not the testis is the starting point for new genes, a general survey of the expression patterns for new genes that originated within vertebrates revealed strong positive correlation with the age in both transcription intensity and spatial expression (161). It is possible that this testis-biased pattern of regene expression is due to our inability to detect genes expressed at low levels in different tissues, but this issue should be resolved soon with advances in next-generation sequencing.

EVOLUTIONARY FORCES ACTING ON NEW GENES

Evolutionary forces such as natural selection and genetic drift operate on both facets of new gene evolution: the fixation of new gene loci and their acquisition of a beneficial function. The two phases of new gene evolution, fixation and acquisition of a beneficial role, may overlap. In this section we will discuss theoretical models developed to describe how new genes arise and acquire novel functions, as well as discuss general approaches to studying new genes and the selective forces that act on them.

Selective models of new gene evolution

Muller (101) was among the first to recognize the potential importance of duplicate genes in evolution. He proposed a simple model whereby new duplicate genes could acquire novel, beneficial functions distinct from those of the original copies. Ohno (110) elaborated on Muller's model and named the fate Muller described as **neofunctionalization**. But Ohno also predicted that duplicate genes are most often inactivated and become pseudogenes. This classic model assumes that the new gene is functional upon duplication and that the new gene subsequently acquires mutations that provide a novel beneficial function. The novel function is then preserved in the genome by natural selection.

However, strictly duplicate genes are redundant, and beneficial mutations are extremely rare. How do new duplicate genes remain in the population long enough to accumulate a beneficial, selected mutation(s)? This problem led to the development of models that predict selective preservation of both copies at all stages of their evolution: Adaptive Radiation (AR), Innovation-Amplification-Divergence (IAD), and Escape from Adaptive Conflict (EAC). The AR model proposes that gene duplication itself is favored, *e.g.* for increased dosage of a gene product, and that the new duplicates then undergo functional radiation (47). Thus, AR posits that novel functions are acquired post-duplication. IAD and EAC, in contrast, propose that ancestral loci develop novel beneficial secondary functions before duplication (8, 35). Under IAD, repeated gene duplication is favored to increase the dosage of the novel secondary function. Different duplicates are then free to optimize the ancestral or function, and only the two best copies are retained in the genome. The increase in the number of duplicate genes under the AR and IAD models also provides additional targets for beneficial mutations, thus increasing the probability and speed of functional improvement. EAC predicts that the bifunctional ancestral gene is subject to selection before gene duplication, adaptive conflict between the ancestral and new function constrains improvement of the selected function(s) before duplication, and that adaptive changes and functional improvement occur in the daughter genes after duplication.

For additional information on duplicate gene evolution, see Conant and Wolfe (32), who suggest that preservation of new genes stems from the co-option of existing functions to serve new purposes, and Walsh (141, 142) for detailed mathematical descriptions of the models and relative probabilities of neofunctionalization and pseudogenization.

Examples of EAC (35), IAD (103), and AR (47) have been published, and each model has specific predictions for what we should observe if a new gene originated by each process (32). However, none of these models can be used as a statistical framework for rigorously

testing the roles of evolutionary forces in new gene origination. Classic molecular population genetic tests based on nucleotide substitution patterns and allele frequency spectra do provide this framework and have been used extensively to detect selection on new genes. These tests, such as the M-K test (95) and the HKA test (59), detect elevated rates of amino acid substitutions (M-K) or reduced effective population size (HKA) at loci. In addition, Thornton (131) introduced a coalescent-based model that can be used to test for selection on copy number variation (CNV). The HKA test and Thornton's test compare measurements of nucleotide variation in genes to a distribution of parameter values derived from neutral coalescent simulations. Thus the M-K, HKA, and Thornton's tests are used to test the classic model.

Each of these 5 models (classic, AR, IAD, EAC, and statistical models) predicts that new genes should experience strong natural selection after they are formed. We will now discuss some of the evidence indicating that this often appears to be the case.

Fixation of new genes within species and populations

The first study to identify signatures of selection on a new gene journeying to fixation was performed by Llopart *et al.* (80), who analyzed a new variant of the *jingwei* gene in *D. teissieri* which lost its second intron. This *D. teissieri*-specific intron presence-absence polymorphism exhibits a significant excess of rare alleles and patterns of nucleotide polymorphism consistent with moderate natural selection driving the polymorphism to fixation. Selection has also been detected on CNV in *D. melanogaster* and other organisms. Emerson *et al.* (41) found a genome-wide pattern consistent with strong purifying selection on CNV except whole-gene duplications, which are under significantly weaker purifying selection. Similarly, Schrider *et al.* (120) and Schrider *et al.* (119) showed a significant excess of fixed versus polymorphic retrogene CNVs originating from the X chromosome in both *Drosophila* and humans, indicating that natural selection governs the patterns of retrogene CNV evolution (Figure 5A). Overall, these studies show that natural selection can play a key role in driving new genes to fixation. In addition, they highlight the use of classic population genetic tests in determining whether selection acts on new genes during their journeys to fixation.

Selection on sequence changes in new genes

In addition to studies of the evolutionary forces governing the fixation of new genes, many studies have investigated the effects of selection and drift on new gene sequences. Long and Langley (81) showed that the new chimeric gene *jingwei* in *D. teissieri* and *D. yakuba* contains a significant excess of nonsynonymous substitutions to nonsynonymous polymorphisms (relative to synonymous substitutions to polymorphisms), indicating that amino acid substitutions were rapidly driven to fixation shortly after the origination of *jgw* (Figure 5B). Similarly, Nurminsky *et al.* (108) showed that a *D. melanogaster*-specific gene family, *Sdic*, involved in sperm motility, rapidly acquired a new exon-intron structure and testis-specific expression (Figure 1). This fusion protein underwent rapid structural renovations, including the conversion of a *Cdic* intron into an exon and an *AnnX* exon and *Cdic* intron into a testis-specific promoter. Low levels of sequence polymorphism,

preservation of coding potential, and the absence of *Sdic* in other closely related species suggests that *Sdic* was rapidly swept to fixation.

These first discoveries sparked searches for general evolutionary patterns in new genes. Jones and Begun (63) searched for common patterns in the evolution of three *Adh*-derived chimeric genes in different lineages of *Drosophila*. All three new genes quickly accumulated a large number of amino acid replacement substitutions in the *Adh*-derived region shortly after they arose, several at identical amino acid sites. Strikingly, Jones and Begun (63) and Shih and Begun (123) showed that different *Adh*-derived fusion genes often accumulate mutations at the same sites, regardless of which other gene they have fused to (Figure 5C). In addition, each of the 4 *Adh*-derived fusion genes exhibits strong signals of accelerated amino acid substitution using classic population genetic statistical tests (e.g. M-K test).

Some of these observations have recently been borne out by genome-wide studies. Xu *et al.* (152) surveyed structural differences between over 600 paralogous pairs of genes in plants and found that most new genes underwent radical changes in exon/intron content and boundaries, as well as insertion/deletions. And using molecular population genetic tests, Chen *et al.* (29) found that young genes in *D. melanogaster* show strong signals of selection. These authors predicted that ~25% of amino acid substitutions in young essential genes were fixed by natural selection. In addition, this signal of selection diminishes as genes grow older.

Altogether these studies indicate that there are general patterns to new gene evolution: new genes often undergo rapid (or immediate) structural and sequence renovations and expression pattern changes which are driven by strong natural selection.

Analysis of new gene structure and function

In addition to analyses of new gene frequencies and nucleotide changes, many groups have investigated the evolutionary forces acting on new genes by analyzing new gene functions, genomic locations, or expression patterns. This complementary approach has revealed several fundamental patterns of new gene origination. Chen *et al.* (23) and Cheng and Chen (30), for example, investigated the antifreeze proteins found in the blood of several orders of Arctic and Antarctic fish. These proteins independently evolved in the different orders, yet they consist of nearly identical tripeptide repeats. These tripeptide repeats were generated *de novo* by amplification of short nucleotide sequences. These studies showed that similar environmental pressures may favor the generation of genes with similar functions.

In addition, as we showed in the previous section, testis-biased genes are under-represented on the *D. melanogaster* and mammal X chromosome. Diaz-Castillo and Ranz's (37) analysis of the genomic location of genes relative to the position of chromosome domains during spermatogenesis led the authors to alternatively propose that the enrichment of testis-biased retrogenes on the autosomes is caused by an increased availability during spermatogenesis of open chromatin domains that contain testis-expressed genes. This larger target for retrogene integration allows a higher proportion of these retrogenes to acquire testis-biased expression. These general observations of the location of sex-biased genes, and their general

movement off of the X chromosome, indicate that differences in expression alone can dictate where in the genome new genes originate.

Together, these results show that studies of general patterns of extant gene locations, structures, and expressions can be informative of new gene origination and evolution.

PHENOTYPIC EFFECTS OF NEW GENES

Studying the roles of new genes in phenotypic evolution recently became feasible with the advent of sophisticated genetic tools, molecular techniques and significant progress in related areas of important phenotypes in biology. Young genes are often assumed to be dispensable because important functions are thought to require a long evolutionary time to be developed and optimized (74). However, studies in the last decade have found numerous young genes with important, and sometimes essential, functions at the molecular, cellular, and individual level (26).

Biochemical pathways

New genes can generate new biochemical pathways and products if they are enzymes or become enzymes. Zhang *et al.* (158) showed that *jingwei* evolved the capacity to catalyze breakdown of long-chain alcohols in *D. yakuba* and *D. teissieri*, while the parent *Adh* can only act on short-chain alcohols. In *Arabidopsis*, Weng *et al.* (148) and Matsuno *et al.* (93) demonstrated that three recently evolved new gene duplicates from the P-450 family, Cyp98A9 and Cyp98A8 and Cyp84A4 assembled two new biochemical pathways related to phenolic metabolism required for pollen development and α -pyrone synthesis, respectively.

Gene expression networks

New genes can also be quickly integrated into existing gene networks. Chen *et al.* (29) observed that almost all young essential genes have been assimilated into protein-protein physical interaction networks in *Drosophila* and a significant number of these young genes have developed multiple interactions with old genes (Figure 7). Integration appears to be driven by natural selection. Several new genes have become new hubs. Analysis of one new gene, *Zeus*, derived from the DNA-binding protein *Caf40* via retroposition (27), revealed that *Zeus* retained ~30% of *Caf40*'s DNA binding sites. But in a short evolutionary time (4–6 MYs) *Zeus* acquired 193 new binding sites through which it activates or represses hundreds of downstream genes involved in reproduction. This observation indicates that gene expression networks can be rapidly and globally reshaped in evolution by new genes. Li *et al.* (75) showed that a de novo gene in yeast can suppress a previously existing mating type-control pathway thus rewiring the structure of gene networks in the species. Capra *et al.* (17) revealed that new genes in yeast become more integrated into cellular networks over time.

Development

Surprisingly, new genes can quickly acquire essential roles in development. Chen *et al.* (29) identified 59 genes that originated in the last ~35 MYs in *Drosophila* that evolved essential developmental functions. Silencing expression of these young genes causes development

failure in early to late pupa, and in some cases earlier (Figure 6A and 6B). Furthermore, tissue-specific knockdown of these young genes can cause morphological defects in adult flies. Silencing new genes can also have a critical effect on reproduction, even when the individual can complete development. The duplicate gene *nsr* (*novel spermatogenesis regulator*) exists only in the four species of the *D. melanogaster* clade that diverged 3 MYs ago, yet it evolved an essential function required for sperm individualization (38). Similarly, silencing *Zeus*, a gene in the same group of *Drosophila*, causes sterility by disrupting testis and sperm development (27).

Brain evolution in flies and humans

Chen *et al.* (28) investigated the expression patterns of new genes in *Drosophila* and found that ~5 new genes per MYs evolved brain expression patterns, mostly in structures involved in olfaction and learning/memory. All new brain genes are expressed in the $\alpha\beta$ lobe, an evolutionarily new set of neurons, implicating new genes in the evolution of this brain structure. Some of the new brain genes have significant effects on the behavior. For example, *Xcbp1* and *Desr* influence foraging behaviors (28) and *sphinx* influences courtship behaviors (33). The incorporation of new genes into the brain is not specific to *Drosophila*. Zhang *et al.* (160) found a correlation between new genes and brain evolution in the human lineage. A high proportion of hominoid-specific and human-specific genes are expressed in the prefrontal cortex and temporal lobe, the newest brain structures, in early fetal development. Strikingly, 54 / 380 human-specific genes are expressed in these two brain regions, regions that are critical for proper cognitive functions. One of these genes, SRGAP2, is involved in neocortical development (22, 36).

Sexual dimorphism and sexual reproduction

New genes impact sexual dimorphism by participating in the genetic systems that control sexual reproduction and sex determination (85). As the aforementioned patterns of new gene origination show, the vast majority of new genes are sex-biased, especially male-biased, expressed and their origination processes show directional copying between the sex chromosomes and autosomes (*e.g.* 10, 42). A number of new genes have been identified with various phenotypic effects ranging from the testicular descent in theria (*RLN3*; 113), testis size in mouse (noncoding RNA gene, *Poldi*; 57), sperm-competition in *D. melanogaster* (*Sdic*; 156) to spermatogenesis in *Drosophila* (*nsr*; 38).

The ability of new genes to be incorporated into such “conserved” pathways, networks, and developmental programs warrants considerable further study. What specific roles can new genes play, and what characteristics of new genes enable them to become essential components of these processes so quickly? New genes now appear to be potent drivers of phenotypic evolution and the genetic control of important biological processes, and show that organismal development and organ development have evolved species-specific and lineage-specific components. Understanding the evolution and modification of these components through the incorporation of new genes is a crucial area of further research.

CHALLENGES FOR THE FUTURE

It is apparent that we have just a glimpse of the emerging world of new genes, and that they play crucial roles in rapidly the evolving genetic systems governing biological diversity. Questions about new gene evolution have opened many doors for both our understanding of existing diversity and for new research. For example, most studies examined new genes generated from a few mechanisms, e.g. duplication and *de novo* origination, leaving open a vast array of mechanisms to be investigated. In addition, even those new genes which are easiest to identify are seriously biased against, and biological studies tend to neglect them (161). Continued efforts will be invaluable for understanding the abundance of new genes, the mechanisms which have been neglected so far, and even new gene evolution in non-model organisms. An outstanding challenge is to understand the roles of new genes in the evolution and biology of phenotypes, and the studies we have highlighted have left important, unresolved questions to be answered. For example, what evolutionary forces drive gene traffic? How do new genes evolve essential developmental functions, and how quickly? How is CNV driven to fixation, and when do CNVs acquire novel functions? How are important structures, such as the human brain, able to incorporate new gene functions, and how do new genes contribute to novel cognitive function? Future studies of more, diverse phenotypes will help shed light onto the general patterns and modes of new gene evolution and the influence of new genes on evolving systems. In addition, understanding how phenotypes rapidly evolve will require a deep understanding underlying local and global gene networks. This will be a tremendous challenge, ranging from the experimental deciphering and graphic description of the gene networks to a valid comparative analysis of the ancestral and derived networks shaped by new genes, and eventually the causal relationship of the altered networks with the evolution of phenotypes.

ACKNOWLEDGMENTS

We thank all members in the M. L. lab, past or present, for their scientific contribution to the relevant topics we discussed in this review. We also thank NIH, NSF, Packard Foundation and the late Edna K. Papazian for their supports to new gene studies throughout the past fifteen years when we explored this new and exciting area. M. L. is currently supported by NIH 1R01GM100768-01A1, NSF1051826 and NSF1026200; N.W.V. by the NSF GRF and partially by the NIH genetics training grant T32 GM007197; S.C. by NSF dissertation improvement grant DEB-1110607; M.D.V. by Pew Latin American Postdoctoral Fellowship.

LITERATURE CITED

1. Arguello JR, Chen Y, Yang S, Wang W, Long M. Origination of an X-linked Testes Chimeric Gene by Illegitimate Recombination in *Drosophila*. *PLoS Genetics*. 2006; 2(5):0745–0754.
2. Bachtrog, D.; Toda, NRT.; Lockton, S. *Current Biology*. Vol. 20. Elsevier Ltd.; 2010. Dosage Compensation and Demasculinization of X Chromosomes in *Drosophila*; p. 1476-1481.
3. Bai Y, Casola C, Feschotte C, Betrán E. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biology*. 2007; 8(1):R11.1–R11.9. [PubMed: 17233920]
4. Baker DA, Russell S. Role of Testis-Specific Gene Expression in Sex-Chromosome Evolution of *Anopheles gambiae*. *Genetics*. 2011; 189(3):1117–1120. [PubMed: 21890740]
5. Begun DJ, Lindfors HA, Kern AD, Jones CD. Evidence for *de Novo* Evolution of Testis-Expressed Genes in the *Drosophila yakuba/Drosophila erecta* Clade. *Genetics*. 2007; 176(2):1131–1137. [PubMed: 17435230]

6. Berezikov, E. *Nature Reviews Genetics*. Vol. 12. Nature Publishing Group; 2011. Evolution of microRNA diversity and regulation in animals; p. 846-860.
7. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*. 2003; 424(6945):197–201. [PubMed: 12853958]
8. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(43):17004–17009. [PubMed: 17942681]
9. Betrán E, Long M. Dntf-2r, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics*. 2003; 164(3):977–988. [PubMed: 12871908]
10. Betrán E, Thornton K, Long M. Retroposed new genes out of the X in *Drosophila*. *Genome research*. 2002; 12:1854–1859. [PubMed: 12466289]
11. Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff J-N. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome research*. 2008; 16(1):203–215. [PubMed: 18293113]
12. Bowers JE, Chapman BA, Rong J. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*. 2003; 422:433–438. [PubMed: 12660784]
13. Brosius J. Retroposons -- seeds of evolution. *Science*. 1991; 251(4995):753. [PubMed: 1990437]
14. Brosius J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica*. 2003; 118(2–3):99–116. [PubMed: 12868601]
15. Brunet FG, Roest Crollius H, Paris M, Aury J-M, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M. Gene Loss and Evolutionary Rates Following Whole-Genome Duplication in Teleost Fishes. *Molecular Biology and Evolution*. 2006; 23(9):1808–1816. [PubMed: 16809621]
16. Cai J, Zhao R, Jiang H, Wang W. De Novo Origination of a New Protein-Coding Gene in *Saccharomyces cerevisiae*. *Genetics*. 2008; 179(1):487–496. [PubMed: 18493065]
17. Capra, JA.; Pollard, KS.; Singh, M. *Genome biology*. Vol. 11. BioMed Central Ltd.; 2010. Novel genes exhibit distinct patterns of function acquisition and network integration; p. R127
18. Cardoso-Moreira M, Emerson JJ, Clark AG, Long M. *Drosophila* Duplication Hotspots are Associated with Late-Replicating Regions of the Genome. *PLoS Genetics*. 2011; 7(11):e1002340. [PubMed: 22072977]
19. Cardoso-Moreira M, Long M. Mutational bias shaping fly copy number variation: implications for genome evolution. *Trends in Genetics*. 2010; 26(6):242–243.
20. Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charleat B, et al. Proto-genes and de novo gene birth. *Nature*. 2012; 487(7407):370–374. [PubMed: 22722833]
21. Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*. 1987; 130(1):113–146.
22. Charrier C, Joshi K, Coutinho-Budd J, Kim J-E, Lambert N, De Marchena J, Jin W-L, et al. Inhibition of SRGAP2 Function by Its Human-Specific Paralog Induces Neoteny During Spine Maturation. *Cell*. 2012; 149(4):923–935. [PubMed: 22559944]
23. Chen L, DeVries AL, Cheng CH. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(8):3811–3816. [PubMed: 9108060]
24. Chen M, Zou M, Fu B, Li X, Vibranovski MD, Gan X, Wang D, Wang W, Long M, He S. Evolutionary Patterns of RNA-based Duplication in Non-Mammalian Chordates. *PloS One*. 2011; 6(7):e21466. [PubMed: 21779328]
25. Chen S-T, Cheng H-C, Barbash DA, Yang H-P. Evolution of hydra, a Recently Evolved Testis-Expressed Gene with Nine Alternative First Exons in *Drosophila melanogaster*. *PLoS Genetics*. 2007; 3(7):e107. [PubMed: 17616977]
26. Chen S, Krinsky BH, Long M. New genes as drivers of phenotypic evolution. *Nature Reviews Genetics*. 2013 in prep.
27. Chen, S.; Ni, X.; Krinsky, BH.; Zhang, YE.; Vibranovski, MD.; White, KP.; Long, M. *The EMBO journal*. Vol. 31. Nature Publishing Group; 2012. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene; p. 2798-2809.

28. Chen S, Spletter M, Ni X, White KP, Luo L, Long M. Frequent Recent Origination of Brain Genes Shaped the Evolution of foraging behavior in *Drosophila*. *Cell Reports*. 2012; 1(2):118–132. The Authors. [PubMed: 22832161]
29. Chen S, Zhang YE, Long M. New Genes in *Drosophila* Quickly Become Essential. *Science*. 2010; 330(6011):1682–1685. [PubMed: 21164016]
30. Cheng C-HC, Chen L. Evolution of an antifreeze glycoprotein. *Nature*. 1999 Sep;401:443–444. [PubMed: 10519545]
31. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 2007; 450(7167):203–218. [PubMed: 17994087]
32. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*. 2008; 9(12):938–950.
33. Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W, Long M. The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(21):7478–7483. [PubMed: 18508971]
34. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The Evolution of Mammalian Gene Families. *PloS One*. 2006; 1(1):e85. [PubMed: 17183716]
35. Deng C, Cheng C-HC, Ye H, He X, Chen L. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(50):21593–21598. [PubMed: 21115821]
36. Dennis, MY.; Nuttle, X.; Sudmant, PH.; Antonacci, F.; Graves, TA.; Nefedov, M.; Rosenfeld, JA., et al. *Cell*. Vol. 149. Elsevier Inc.; 2012. Evolution of Human-Specific Neural SRGAP2 Genes by Incomplete Segmental Duplication; p. 912-922.
37. Díaz-Castillo C, Ranz JM. Nuclear Chromosome Dynamics in the *Drosophila* Male Germ Line Contribute to the Nonrandom Genomic Distribution of Retrogenes. *Molecular Biology and Evolution*. 2012; 29(9):2105–2108. [PubMed: 22427708]
38. Ding Y, Zhao L, Yang S, Jiang Y, Chen Y, Zhao R, Zhang Y, et al. A Young *Drosophila* Duplicate Gene Plays Essential Roles in Spermatogenesis by Regulating Several Y-linked Male Fertility Genes. *PLoS Genetics*. 2010; 6(12):e1001255. [PubMed: 21203494]
39. Dopman EB, Hartl DL. A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(50):19920–19925. [PubMed: 18056801]
40. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. *Science*. 2006; 312(5780):1653–1655. [PubMed: 16778056]
41. Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science (New York, N.Y.)*. 2008; 320(5883):1629–1631.
42. Emerson JJ, Kaessmann H, Betrán E, Long M. Extensive gene traffic on the mammalian X chromosome. *Science*. 2004; 303(5657):537–540. [PubMed: 14739461]
43. Fan C, Chen Y, Long M. Recurrent Tandem Gene Duplication Gave Rise to Functionally Divergent Genes in *Drosophila*. *Molecular Biology and Evolution*. 2008; 25(7):1451–1458. [PubMed: 18408233]
44. Fan C, Vrbancin MD, Chen Y, Long M. A Microarray Based Genomic Hybridization Method for Identification of New Genes in Plants : Case Analyses of *Arabidopsis* and *Oryza*. *Journal of Integrative Plant Biology*. 2007; 49(6):915–926.
45. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*. 2008; 9(5):397–405.
46. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nature Reviews Genetics*. 2006; 7(2):85–97.
47. Francino MP. An adaptive radiation model for the origin of new gene functions. *Nature Genetics*. 2005; 37(6):573–577. [PubMed: 15920518]

48. Fu, B.; Chen, M.; Zou, M.; Long, M.; He, S. BMC Genomics. Vol. 11. BioMed Central Ltd.; 2010. The rapid generation of chimerical genes expanding protein diversity in zebrafish; p. 657
49. Gallach M, Chandrasekaran C, Betrán E. Analyses of Nuclearily Encoded Mitochondrial Genes Suggest Gene Duplication as a Mechanism for Resolving Intralocus Sexually Antagonistic Conflict in *Drosophila*. *Genome Biology and Evolution*. 2010; 2:835–850. [PubMed: 21037198]
50. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. Evolution of a Complex Locus: Exon Gain, Loss and Divergence at the Gr39a Locus in *Drosophila*. *PloS One*. 2008; 3(1):e1513. [PubMed: 18231599]
51. Gilbert W. Why genes in pieces? *Nature*. 1978; 271:501. [PubMed: 622185]
52. Gillespie, J. Molecular Evolution and the Neutral Allele Theory. In: Harvey, P.; Partridge, L., editors. *Oxford Surveys in Evolutionary Biology*. Vol. 4. USA: Oxford University Press; 1987. p. 10-37.
53. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, et al. A High-Resolution Map of Segmental DNA Copy Number Variation in the Mouse Genome. *PLoS Genetics*. 2007; 3(1):e3. [PubMed: 17206864]
54. Haldane J. The Time of Action of Genes, and Its Bearing on some Evolutionary Problems. *The American naturalist*. 1932; 66(702):5–24.
55. Hall C, Brachat S, Dietrich FS. Contribution of Horizontal Gene Transfer to the Evolution of *Saccharomyces cerevisiae* Contribution of Horizontal Gene Transfer to the Evolution of *Saccharomyces cerevisiae* †. *Eukaryotic Cell*. 2005; 4(6):1102–1115. [PubMed: 15947202]
56. Harr B, Turner LM. Genome-wide analysis of alternative splicing evolution among *Mus* subspecies. *Molecular Ecology*. 2010; 19:228–239. [PubMed: 20331782]
57. Heinen TJAJ, Staubach F, Häming D, Tautz D. Emergence of a new gene from an intergenic region. *Current Biology*. 2009; 19(18):1527–1531. [PubMed: 19733073]
58. Hense W, Baines JF, Parsch J. X Chromosome Inactivation During *Drosophila* Spermatogenesis. *PLoS Biology*. 2007; 5(10):e273. [PubMed: 17927450]
59. Hudson RR, Kreitman M, Aguade M. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics*. 1987; 116:153–159. [PubMed: 3110004]
60. Innan, H.; Kondrashov, F. *Nature Reviews Genetics*. Vol. 11. Nature Publishing Group; 2010. The evolution of gene duplications: classifying and distinguishing between models; p. 97-108.
61. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004; 432(7018): 695–716. [PubMed: 15592404]
62. Jiang N, Bao Z, Zhang X, Rddy SR, Wessler SR. Pack-MULE transposable elements mediate gene evolution in plants. *Nature*. 2004 Sep.431:569–573. [PubMed: 15457261]
63. Jones CD, Begun DJ. Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(32):11373–11378. [PubMed: 16076957]
64. Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics*. 2009; 10(1):19–31.
65. Katju V, Lynch M. The Structure and Early Evolution of Recently Arisen Gene Duplicates in the *Caenorhabditis elegans* Genome. *Genetics*. 2003; 165(4):1793–1803. [PubMed: 14704166]
66. Katju V, Lynch M. On the Formation of Novel Genes by Duplication in the *Caenorhabditis elegans* Genome. *Molecular Biology and Evolution*. 2006; 23(5):1056–1067. [PubMed: 16500928]
67. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*. 2004; 428(6983):617–624. [PubMed: 15004568]
68. Keren, H.; Lev-Maor, G.; Ast, G. *Nature Reviews Genetics*. Vol. 11. Nature Publishing Group; 2010. Alternative splicing and evolution: diversification, exon definition and function; p. 345-355.
69. Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature Genetics*. 2004; 36(6):642–646. [PubMed: 15156144]
70. Knowles DG, McLysaght A. Recent de novo origin of human protein-coding genes. *Genome Research*. 2009; 19(10):1752–1759. [PubMed: 19726446]

71. Koonin EV, Makarova KS, Aravind L. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annual Reviews Microbiology*. 2001; 55:709–742.
72. Langley CH, Stevens K, Cardeno C, Lee Y, Schrider DR, Pool JE, Langley SA, et al. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*. 2012
73. Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(26): 9935–9939. [PubMed: 16777968]
74. Lewin, B. *Lewin's Genes X*. Krebs, J.; Goldstein, ES.; Kilpatrick, ST., editors. Sudbury, MA: Jones and Bartlett Publishers, LLC.; 2011.
75. Li, D.; Dong, Y.; Jiang, Y.; Jiang, H.; Cai, J.; Wang, W. *Cell Research*. Vol. 20. Nature Publishing Group; 2010. A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand; p. 408–420.
76. Li WH, Gojobori T. Rapid eEvolution of Goat and Sheep Globin Genes Following Gene Duplication. *Molecular Biology and Evolution*. 1983; 1(1):94–108. [PubMed: 6599963]
77. Li, W-H. *Molecular Evolution*. Sunderland, Massachusetts: Sinauer Associates; 1997. p. 432
78. Li Z, Liu M, Zhang L, Zhang W, Gao G, Zhu Z, Wei L, Fan Q, Long M. Detection of intergenic non-coding RNAs expressed in the main developmental stages in *Drosophila melanogaster*. *Nucleic Acids Research*. 2009; 37(13):4308–4314. [PubMed: 19451167]
79. Lipinski, KJ.; Farslow, JC.; Fitzpatrick, KA.; Lynch, M.; Katju, V.; Bergthorsson, U. *Current Biology*. Vol. 21. Elsevier Ltd.; 2011. High Spontaneous Rate of Gene Duplication in *Caenorhabditis elegans*; p. 306–310.
80. Llopart A, Comeron JM, Brunet G, Lachaise D, Long M. Intron presence – absence polymorphism in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(12):8121–8126. [PubMed: 12060758]
81. Long M, Langley CH. Natural Selection and the Origin of jingwei, a Chimeric Processed Functional Gene in *Drosophila*. *Science*. 1993; 260(5104):91–95. [PubMed: 7682012]
82. Long M, Rosenberg C, Gilbert W. Intron phase correlations and the evolution of the intron/exon structure of genes. *Proceedings of the National Academy of Sciences of the United States of America*. 1995; 92(26):12495–12499. [PubMed: 8618928]
83. Long, M. *The origin and evolutionary mechanisms of new genes*. UC Davis; 1992.
84. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics*. 2003; 4(11):865–875.
85. Long, M.; Vibrationovski, MD.; Zhang, YE. Evolutionary interactions between sex chromosomes and autosomes. In: Singh, R.; Xu, J.; Kulathinal, R., editors. *Rapidly Evolving Genes and Genetic Systems*. Oxford: Oxford University Press; 2012. p. 101–114.
86. Lorenc A, Makalowski W. Transposable elements and vertebrate protein diversity. *Genetica*. 2003; 118(2–3):183–191. [PubMed: 12868608]
87. Lu J, Fu Y, Kumar S, Shen Y, Zeng K, Xu A, Carthew R, Wu C-I. Adaptive Evolution of Newly Emerged Micro-RNA Genes in *Drosophila*. *Molecular Biology and Evolution*. 2008; 25(5):929–938. [PubMed: 18296702]
88. Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu C-I. The birth and death of microRNA genes in *Drosophila*. *Nature Genetics*. 2008; 40(3):351–355. [PubMed: 18278047]
89. Makalowski, W. SINES as a genomic scrap yard: an essay on genomic evolution. In: Maraia, R., editor. *The Impact of Short Interspersed Elements (SINES) on the Host Genome*. Austin, TX, USA: RG Landes Company; 1995. p. 81–104.
90. Marques, AC.; Tan, J.; Lee, S.; Kong, L.; Heger, A.; Ponting, CP. *Genome Biology*. Vol. 13. BioMed Central Ltd.; 2012. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs; p. R102
91. Marques AC, Tan J, Ponting CP. Wrangling for microRNAs provokes much crosstalk. *Genome Biology*. 2011; 12(11):132. [PubMed: 22104690]
92. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H. Emergence of Young Human Genes After a Burst of Retroposition in Primates. *PLoS Biology*. 2005; 3(11):e357. [PubMed: 16201836]

93. Matsuno M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard J-E, Pollet B, et al. Evolution of a Novel Phenolic Pathway for Pollen Development. *Science*. 2009; 325(5948):1688–1692. [PubMed: 19779199]
94. McCarrey JR, Riggs AD. Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; 83(3):679–683. [PubMed: 2418440]
95. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*. 1991; 351:652–654. [PubMed: 1904993]
96. McLysaght A, Hokamp K, Wolfe KH. Extensive genomic duplication during early chordate evolution. *Nature Genetics*. 2002; 31(2):200–204. [PubMed: 12032567]
97. Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. Sex Chromosome-Specific Regulation in the *Drosophila* Male Germline but Little Evidence for Chromosomal Dosage Compensation or Meiotic Inactivation. *PLoS Biology*. 2011; 9(8):e1001126. [PubMed: 21857805]
98. Meisel RP, Han MV, Hahn MW. A Complex Suite of Forces Drives Gene Traffic from *Drosophila* X Chromosomes. *Genome Biology and Evolution*. 2009; 1:176–188. [PubMed: 20333188]
99. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011; 470(7332):59–65. [PubMed: 21293372]
100. Moran NA, Jarvik T. Lateral Transfer of Genes From Fungi Underlies Carotenoid Production in Aphids. *Science*. 2010; 328(5978):624–627. [PubMed: 20431015]
101. Muller HJ. Bar Duplication. 1936; 83(2161):528–530.
102. Murphy DN, McLysaght A. De Novo Origin of Protein-Coding Genes in Murine Rodents. *PloS One*. 2012; 7(11):e48650. [PubMed: 23185269]
103. Näsval J, Sun L, Roth JR, Andersson DI. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science*. 2012; 338(6105):384–387. [PubMed: 23087246]
104. Nekrutenko A, Li WH. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics*. 2001; 17(11):619–621. [PubMed: 11672845]
105. Ni X, Zhang YE, Nègre N, Chen S, Long M, White KP. Adaptive Evolution and the Birth of CTCF Binding Sites in the *Drosophila* Genome. *PLoS Biology*. 2012; 10(11):e1001420. [PubMed: 23139640]
106. Nozawa M, Aotsuka T, Tamura K. A Novel Chimeric Gene, *siren*, with Retroposed Promoter Sequence in the *Drosophila bipectinata* Complex. *Genetics*. 2005; 171(4):1719–1727. [PubMed: 16143626]
107. Nozawa M, Miura S, Nei M. Origins and Evolution of MicroRNA Genes in *Drosophila* Species. *Genome Biology and Evolution*. 2010; 2:180–189. [PubMed: 20624724]
108. Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature*. 1998 Jun.396:572–575. 1994. [PubMed: 9859991]
109. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000; 405(6784):299–304. [PubMed: 10830951]
110. Ohno, S. Evolution by gene duplication. Vol. 160. New York: Springer-Verlag; 1970.
111. Okamura K, Feuk L, Marquès-Bonet T, Navarro A, Scherer SW. Frequent appearance of novel protein-coding sequences by frameshift translation. *Genomics*. 2006; 88(6):690–697. [PubMed: 16890400]
112. Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. Paucity of Genes on the *Drosophila* X Chromosome Showing Male-Biased Expression. *Science*. 2003; 299(5607):697–700. [PubMed: 12511656]
113. Park J, Semyonov J, Chang CL, Yi W, Warren W, Yu S, Hsu T. Origin of INSL3-mediated testicular descent in therian mammals. *Genome Research*. 2008; 18:974–985. [PubMed: 18463305]
114. Piriyaopongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. *Bioinformatics*. 2008; 14:814–821.

115. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-Dependent Gene Expression and Evolution of the *Drosophila* Transcriptome. *Science*. 2003; 300(5626):1742–1745. [PubMed: 12805547]
116. Rogers RL, Bedford T, Hartl DL. Formation and Longevity of Chimeric and Duplicate Genes in *Drosophila melanogaster*. *Genetics*. 2009; 181(1):313–322. [PubMed: 19015547]
117. Rogers RL, Hartl DL. Chimeric Genes as a Source of Rapid Evolution in *Drosophila melanogaster*. *Molecular Biology and Evolution*. 2012; 29(2):517–529. [PubMed: 21771717]
118. Sabath N, Wagner A, Karlin D. Evolution of Viral Proteins Originated De Novo by Overprinting. *Molecular Biology and Evolution*. 2012; 29(12):3767–3780. [PubMed: 22821011]
119. Schrider DR, Navarro FCP, Galante PAF, Parmigiani RB, Camargo AA, Hahn MW, De Souza SJ. Gene Copy-Number Polymorphism Caused by Retrotransposition in Humans. *PLoS Genetics*. 2013; 9(1):e1003242. [PubMed: 23359205]
120. Schrider DR, Stevens K, Cardeno CM, Langley CH, Hahn MW. Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*. *Genome Research*. 2011; 21(12):2087–2095. [PubMed: 22135405]
121. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Månér S, Massa H, et al. Large-Scale Copy Number Polymorphism in the Human Genome. *Science*. 2004; 305(5683):525–528. [PubMed: 15273396]
122. Sémon M, Wolfe KH. Consequences of genome duplication. *Current Opinion in Genetics & Development*. 2007; 17(6):505–512. [PubMed: 18006297]
123. Shih H-J, Jones CD. Patterns of Amino Acid Evolution in the *Drosophila ananassae* Chimeric Gene, *siren*, Parallel Those of Other *Adh*-derived Chimeras. *Genetics*. 2008; 180(2):1261–1263. [PubMed: 18780749]
124. Sturgill D, Zhang Y, Parisi M, Oliver B. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature*. 2007; 450(7167):238–241. [PubMed: 17994090]
125. Sturtevant AH. The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics*. 1925; 10:117–147. [PubMed: 17246266]
126. Sudhof TC, Goldstein JL, Brown MS, Russell DW. The LDL Receptor Gene : A Mosaic of Exons Shared with Different Proteins. *Science*. 1985; 228(4701):815–822. [PubMed: 2988123]
127. Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, Hartl DL. A sex-ratio Meiotic Drive System in *Drosophila simulans*. II: An X-linked Distorter. *PLoS Biology*. 2007; 5(11):e293. [PubMed: 17988173]
128. Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL. A sex-ratio Meiotic Drive System in *Drosophila simulans*. I: An Autosomal Suppressor. *PLoS Biology*. 2007; 5(11):e292. [PubMed: 17988172]
129. Thomson TM, Lozano JJ, Loukili N, Carrio R, Serra F, Cormand B, Valeri M, et al. Fusion of the Human Gene for the Polyubiquitination Coeffector UBEV1 with Kua, a Newly Identified Gene. *Genome Research*. 2000; 10(11):1743–1756. [PubMed: 11076860]
130. Thornton KR. Gene Conversion and Natural Selection at Duplicate Loci in *Drosophila melanogaster*. University of Chicago; 2003.
131. Thornton KR. The Neutral Coalescent Process for Recent Gene Duplications and Copy-Number Variants. *Genetics*. 2007; 177(2):987–1000. [PubMed: 17720930]
132. Tiedge H, Chen W, Brosius J. Primary Structure, Neural-Specific Expression, and Dendritic Location of Human BC200 RNA. *The Journal of Neuroscience*. 1993 Jun.13:2382–2390. [PubMed: 7684772]
133. Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, Albà MM. Origin of Primate Orphan Genes: A Comparative Genomics Approach. *Molecular Biology and Evolution*. 2009; 26(3):603–612. [PubMed: 19064677]
134. Toups MA, Hahn MW. Retrogenes Reveal the Direction of Sex-Chromosome Evolution in Mosquitoes. *Genetics*. 2010; 186(2):763–766. [PubMed: 20660646]
135. Vibranovski MD, Lopes HF, Karr TL, Long M. Stage-Specific Expression Profiling of *Drosophila* Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. *PLoS Genetics*. 2009; 5(11):e1000731. [PubMed: 19936020]

136. Vibranovski, MD.; Zhang, YE.; Kemkemer, C.; Lopes, HF.; Karr, TL.; Long, M. BMC Biology. Vol. 10. BioMed Central Ltd.; 2012. Re-analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the drosophila X chromosome; p. 49author reply 50.
137. Vibranovski MD, Zhang YE, Kemkemer C, VanKuren NW, Lopes HF, Karr TL, Long M. Segmental dataset and whole body expression data do not support the hypothesis that non-random movement is an intrinsic property of Drosophila retrogenes. BMC Evolutionary Biology. 2012; 12:169. [PubMed: 22950647]
138. Vibranovski MD, Zhang Y, Long M. General gene movement off the X chromosome in the Drosophila genus. Genome Research. 2009; 19(5):897–903. [PubMed: 19251740]
139. Vicoso B, Charlesworth B. The Deficit of Male-Biased genes on the D. melanogaster X Chromosome is Expression-Dependent: A Consequence of Dosage Compensation? Journal of Molecular Evolution. 2009; 68(5):576–583. [PubMed: 19407921]
140. Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(9):3220–3225. [PubMed: 16492757]
141. Walsh B. Population-genetic models of the fates of duplicate genes. Genetica. 2003; 118(2–3): 279–294. [PubMed: 12868616]
142. Walsh JB. How Often Do Duplicated Genes Evolve New Functions? Genetics. 1995; 139(1):421–428. [PubMed: 7705642]
143. Wang J, Mager J, Chen Y, Schneider E, Cross JC, Nagy A, Magnuson T. Imprinted X inactivation maintained by a mouse Polycomb group gene. Nature Genetics. 2001; 28(4):371–375. [PubMed: 11479595]
144. Wang J, Long M, Vibranovski MD. Retrogenes Moved Out of the Z Chromosome in the Silkworm. Journal of Molecular Evolution. 2012; 74(3–4):113–126. [PubMed: 22535494]
145. Wang W, Zhang J, Alvarez C, Llopart A, Long M. The Origin of the Jingwei Gene and the Complex Modular Structure of its Parental gene, Yellow Emperor, in Drosophila melanogaster. Molecular Biology and Evolution. 2000; 17(9):1294–1301. [PubMed: 10958846]
146. Wang W, Yu H, Long M. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. Nature Genetics. 2004; 36(5):523–527. [PubMed: 15064762]
147. Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, et al. High Rate of Chimeric Gene Origination by Retroposition in Plant Genomes. The Plant Cell. 2006 Aug.18:1791–1802. [PubMed: 16829590]
148. Weng J-K, Li Y, Mo H, Chapple C. Assembly of an Evolutionarily New Pathway for α -Pyruvate Biosynthesis in Arabidopsis. Science. 2012; 337(6097):960–964. [PubMed: 22923580]
149. Wu D-D, Irwin DM, Zhang Y-P. De Novo Origin of Human Protein-Coding Genes. PLoS Genetics. 2011; 7(11):e1002379. [PubMed: 22102831]
150. Xiao W, Liu H, Li Y, Li X, Xu C, Long M, Wang S. A Rice Gene of De Novo Origin Negatively Regulates Pathogen-Induced Defense Response. PloS One. 2009; 4(2):e4603. [PubMed: 19240804]
151. Xie C, Zhang YE, Chen J-Y, Liu C-J, Zhou W-Z, Li Y, Zhang M, Zhang R, Wei L, Li C-Y. Hominoid-Specific De Novo Protein-Coding Genes Originating From Long Non-Coding RNAs. PLoS Genetics. 2012; 8(9):e1002942. [PubMed: 23028352]
152. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(4): 1187–1192. [PubMed: 22232673]
153. Xue S, Jones MD, Lu Q, Middeldorp JM, Griffin BE. Genetic Diversity : Frameshift Mechanisms Alter Coding of a Gene (Epstein-Barr Virus LF3 Gene) That Contains Multiple 102-Base-Pair Direct Sequence Repeats. Molecular and Cellular Biology. 2003; 23(6):2192–2201. [PubMed: 12612089]
154. Yang S, Arguello JR, Li X, Ding Y, Zhou Q, Chen Y, Zhang Y, et al. Repetitive Element-Mediated Recombination as a Mechanism for New Gene Origination in Drosophila. PLoS Genetics. 2008; 4(1):e3. [PubMed: 18208328]

155. Yang, Z.; Huang, J. FEBS Letters. Vol. 585. Federation of European Biochemical Societies; 2011. De novo origin of new genes with introns in *Plasmodium vivax*; p. 641-644.
156. Yeh S-D, Do T, Chan C, Cordova A, Carranza F, Yamamoto Ea, Abbassi M, et al. Functional evidence that a recently evolved *Drosophila* sperm-specific gene boosts sperm competition. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(6): 2043–2048. [PubMed: 22308475]
157. Yoshida S, Maruyama S, Nozaki H, Shirasu K. Horizontal Gene Transfer by the Parasitic Plant *Striga hermonthica*. *Science*. 2010; 328:1128. [PubMed: 20508124]
158. Zhang J, Dean AM, Brunet F, Long M. Evolving protein functional diversity in new genes of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(46):16246–16250. [PubMed: 15534206]
159. Zhang PG, Huang SZ, Pin A-L, Adams KL. Extensive Divergence in Alternative Splicing Patterns After Gene and Genome Duplication During the Evolutionary History of *Arabidopsis*. *Molecular Biology and Evolution*. 2010; 27(7):1686–1697. [PubMed: 20185454]
160. Zhang YE, Landback P, Vibranovski MD, Long M. Accelerated Recruitment of New Brain Development Genes into the Human Genome. *PLoS Biology*. 2011; 9(10):e1001179. [PubMed: 22028629]
161. Zhang YE, Landback P, Vibranovski M, Long M. New genes expressed in human brains: Implications for annotating evolving genomes. *BioEssays*. 2012; 34(11):982–991. [PubMed: 23001763]
162. Zhang YE, Vibranovski MD, Krinsky BH, Long M. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Research*. 2010; 20(11):1526–1533. [PubMed: 20798392]
163. Zhang YE, Vibranovski MD, Landback P, Marais GaB, Long M. Chromosomal Redistribution of Male-Biased Genes in Mammalian Evolution with Two Bursts of Gene Gain on the X Chromosome. *PLoS Biology*. 2010; 8(10)
164. Zhang Y, Lu S, Zhao S, Zheng X, Long M, Wei L. Positive selection for the male functionality of a co-retroposed gene in the hominoids. *BMC Evolutionary Biology*. 2009; 9:252. [PubMed: 19832993]
165. Zhang Y, Wu Y, Liu Y, Han B. Computational Identification of 69 Retroposons in *Arabidopsis*. *Plant Physiology*. 2005 Jun.138:935–948. [PubMed: 15923328]
166. Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. Parallel Molecular Evolution in an Herbivore Community. *Science*. 2012; 337(6102):1634–1637. [PubMed: 23019645]
167. Zheng D, Gerstein MB. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends in Genetics*. 2007; 23(5):219–224. [PubMed: 17382428]
168. Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. On the origin of new genes in *Drosophila*. *Genome Research*. 2008; 18(9):1446–1455. [PubMed: 18550802]
169. Zhou R, Moshgabadi N, Adams KL. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(38):16122–16127. [PubMed: 21900601]
170. Zhu Z, Zhang Y, Long M. Extensive Structural Renovation of Retrogenes in the Evolution of the *Populus* Genome. *Plant Physiology*. 2009; 151(4):1943–1951. [PubMed: 19789289]

Figure 1.

New genes are defined using syntenic and sequence comparisons between the genomes of a group of related species. A) The general procedure to identify new genes. The relationship of species S1–S4 is shown by the blue tree. The relationships between the genes G1 (yellow), G2 (red), and G3 (green) are shown within the species tree. Aligning the genomes of species S1–S4 shows that the new gene G2 is present in S1–S3 but absent in S4, indicating that G2 arose in the common ancestor of S1–S3. G2 was thus generated in the genome between old genes G1 and G3 in the common ancestor of S1, S2, and S3 (red star). B) An example of using syntenic alignments to identify new genes. *Sdic* exists only in *Drosophila melanogaster* (99). In this case, *Sdic* originated as a chimeric gene through recombination of duplicates of the two flanking genes, a 5' piece of *Cdic* encoding a cytoplasmic dynein intermediate chain and a 3' piece of *AnnX* (see text for further details).

Figure 2.

Representative new genes exhibiting various new gene origination mechanisms. A) *Jingwei*, a new gene found only in *D. teissieri* and *D. yakuba*, was generated by a combination of retroposition, DNA-based duplication and gene recombination which formed a chimeric gene consisting of *Adh*-derived enzymatic domain and a hydrophobic domain from *Ymp* (74,133). B) *PIPSL* in humans is a consequence of gene fusion between two adjacent ancestral genes by read-through transcription and subsequent co-retroposition (152). C) Gene fission split the ancestral gene *monkeying* into two distinct genes in *D. mauritiana*, revealing an intermediate process of gene fission aided by gene duplication and complementary degeneration (134). D) The gene ENSMUSG00000078384 in mouse revealed the evolutionary process of *de novo* gene origination (93). E) Two new genes in humans, *DAF* and *mNSCI*, were generated by domesticating transposable elements, Alu and short interspersed elements (B1-B4) (89, 104). *DAF* and Alu elements is a neat case where alternative splicing generated a new isoform in the mammalian genome. F) Horizontal gene transfer (HGT) is prevalent in bacteria with mechanisms including homologous recombination (101). Antibiotic resistance genes can be acquired by host genomes containing the *intl* gene, which encodes integrase, a recombination site (*att*), and a promoter to express the captured gene, as depicted by the process on the left. See the text for more details.

Figure 3.

The phylogenetic distribution of new gene origination events in *Drosophila* and vertebrates. These genes are only those that were generated by DNA-based duplication, retroposition and *de novo* origination (162, 163). The number of new genes that originated in each time period is shown above the branch. For example, branch 1 in A) shows that 220 genes originated between 36 and 41 MYA in *Drosophila*. In B) red numbers are new genes that originated in the hominoid branches or specifically in humans.

Figure 4.

Retrogene traffic in *Drosophila* (A; 10, 138) and humans (B; 42). Each arrow indicates the movement of retrogenes from the parental gene chromosomal location to the retrogene's location. The size of arrow indicates the intensity of gene movement between chromosomes, and the percentages show quantitatively the excess of movement over the null expectation (random origination and insertion). The functions of the retrogenes are indicated.

Figure 5.

Positive Darwinian selection acting on new genes in *Drosophila*. A) Positive selection for the fixation of new retrogenes in *Drosophila* (120) and humans (119). The numerator and denominator show the numbers of retrogenes that originate on the autosomes and the X, respectively. Tests based on the M-K framework indicate an excess of fixed X→A retrogenes in both species, and strong positive selection for X→A retrogene movement. B) The *jingwei* gene (81). The ratios over the branches are the numbers of nonsynonymous changes over the numbers of synonymous changes, while the ratios in the triangles are the ratios of divergence between the species to polymorphisms. M-K tests and Ka/Ks ratios indicate strong positive selection acted on *jpgw* shortly after it originated. C) Selection acted on all *Adh*-derived chimeric genes in *Drosophila* (63), as indicated by elevated Ka/Ks ratios.

Figure 6. The essential effects of new genes on development. A) Development was terminated at the final stage of development when three different genes were knocked down using RNA interference (RNAi). B) *YLL1* originated in the common ancestor of the *D. melanogaster* subgroup species 6~10 MYA, yet showed lethal effects in the pupal stage when silenced by RNAi, mutated by EMS or disrupted by P-element (29).

Figure 7.

New genes integrated into and reshaped gene networks. A) Yeast new genes that originated through duplication-based (blue) and non-duplication-based (red) mechanisms since the recent whole genome duplication (<100 MYA) were integrated into the physical interaction network (17). The orange box highlights a module composed of two new genes involved in the pathway to form and process actin. *DID4* (green box) interacts with 13 new genes within a few steps. B) New genes form hubs in protein-protein interaction networks (29). C) The *D. melanogaster-D. simulans*-specific gene *Zeus* quickly accumulated more than 100 amino acid substitutions in its nucleotide binding domains under positive selection. Consequently, it evolved into a new DNA binding motif that evolved hundreds of new gene links to rewire the gene networks that control reproduction (27).