



On the origin and evolution of new genes—a genomic and experimental perspective

Qi Zhou, Wen Wang*

CAS-Max Planck Junior Research Group, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

Received for publication 30 August 2008; revised 18 September 2008; accepted 19 September 2008

Abstract

The inherent interest on the origin of genetic novelties can be traced back to Darwin. But it was not until recently that we were allowed to investigate the fundamental process of origin of new genes by the studies on newly evolved young genes. Two indispensable steps are involved in this process: origin of new gene copies through various mutational mechanisms and evolution of novel functions, which further more leads to fixation of the new copies within populations. The theoretical framework for the former step formed in 1970s. Ohno proposed gene duplication as the most important mechanism producing new gene copies. He also believed that the most common fate for new gene copies is to become pseudogenes. This classical view was validated and was also challenged by the characterization of the first functional young gene *jingwei* in *Drosophila*. Recent genome-wide comparison on young genes of *Drosophila* has elucidated a comprehensive picture addressing remarkable roles of various mechanisms besides gene duplication during origin of new genes. Case surveys revealed it is not rare that new genes would evolve novel structures and functions to contribute to the adaptive evolution of organisms. Here, we review recent advances in understanding how new genes originated and evolved on the basis of genome-wide results and experimental efforts on cases. We would finally discuss the future directions of this fast-growing research field in the context of functional genomics era.

Keywords: origin of new genes; gene duplication; *de novo* origination; chimeric genes

Introduction

The enormous disparity in the gene numbers of extant organisms reveals origination of new genes is a general biological phenomenon. Geneticists have started to study the patterns and underlying mechanisms of this fundamental process long before the confirmation of such disparities by recent progresses in genome sequencing. Muller and Haldane first proposed that gene duplication could produce a new gene (Haldane, 1932; Muller, 1935). This idea has been developed and the role of gene duplication has been reinforced to be most important for producing new genes in the Ohno's classical treatise *Evolution by Gene Duplication* since 1970 (Ohno, 1970). His classical model assumed that the new gene copy would be structurally and

functionally identical to its ancestor. Because of such redundancies, most new gene copies would suffer an accumulation of deleterious mutations and finally become pseudogenes (Ohno, 1973). This prediction contradicts with the later empirical results showing a large proportion of duplicated genes are functional (Force et al., 1999). It sparked intensive discussions on models addressing evolutionary fates of new gene copies, including evolution of functional redundancy, subfunctionalization, and neofunctionalization (Clark, 1994; Walsh, 1995; Force et al., 1999). All these hypotheses/models aim to answer two basic questions regarding the full origination process of a new gene, i.e., how does a new gene copy originate in the genome? And how does it evolve and become fixed within the population?

There used to be little opportunity to test the above hypotheses and explore the two questions, because most studied gene duplicates are hundreds of millions of years

* Corresponding author.

E-mail address: wwang@mail.kiz.ac.cn

old (Long et al., 2003). In 1993, Long et al. characterized the first young gene, named after a Chinese legendary goddess, *jingwei* in *Drosophila yakuba* (Long and Langley, 1993). It originated after the split of the *D. yakuba* lineage from other species like *D. melanogaster*, which indicates its extremely young age within only several million years. Such a short evolutionary time scale allows a scrutinization of the complete formation process of a new gene on the basis of sequence changes easily to be traced compared with its ancestor. It turns out that *jingwei* formed by recruiting parts from a duplicated copy and a retroposed copy of two different ancestral genes (Long and Langley, 1993; Wang et al., 2000). The case of *jingwei*, together with the following characterized new genes *sphinx*, *monkey king* and *Hun* etc., suggests the formation of a new gene usually involves complex sequence and structure changes (Wang et al., 2002, 2004; Arguello et al., 2006). Most importantly, they show that studying young genes is a direct and efficient approach to study origin of new genes (Long et al., 2003).

In this paper, we review abundant models and recent

advances in the field of new gene study. A lot of these advances are acquired from studies of young genes. We have recently screened young genes with both experimental and genomic approaches in *Drosophila* species, achieving the first comprehensive picture regarding how a new gene originated in the genome (Yang et al., 2008; Zhou et al., 2008). While the case studies on how a new gene evolves novel functions and contributes to adaptive evolution are sporadic and just rising, we will cover intriguing cases from primates and *Drosophila*, which might suggest general patterns of functional evolution of new genes to illuminate the future directions of this research field.

How does a new gene originate within the genome?

A new gene can arise through four mechanisms (Fig. 1): gene duplication, retroposition, horizontal gene transfer, and *de novo* origination from non-coding sequences. They can work individually or sometimes in combination, such as the case of *jingwei*.

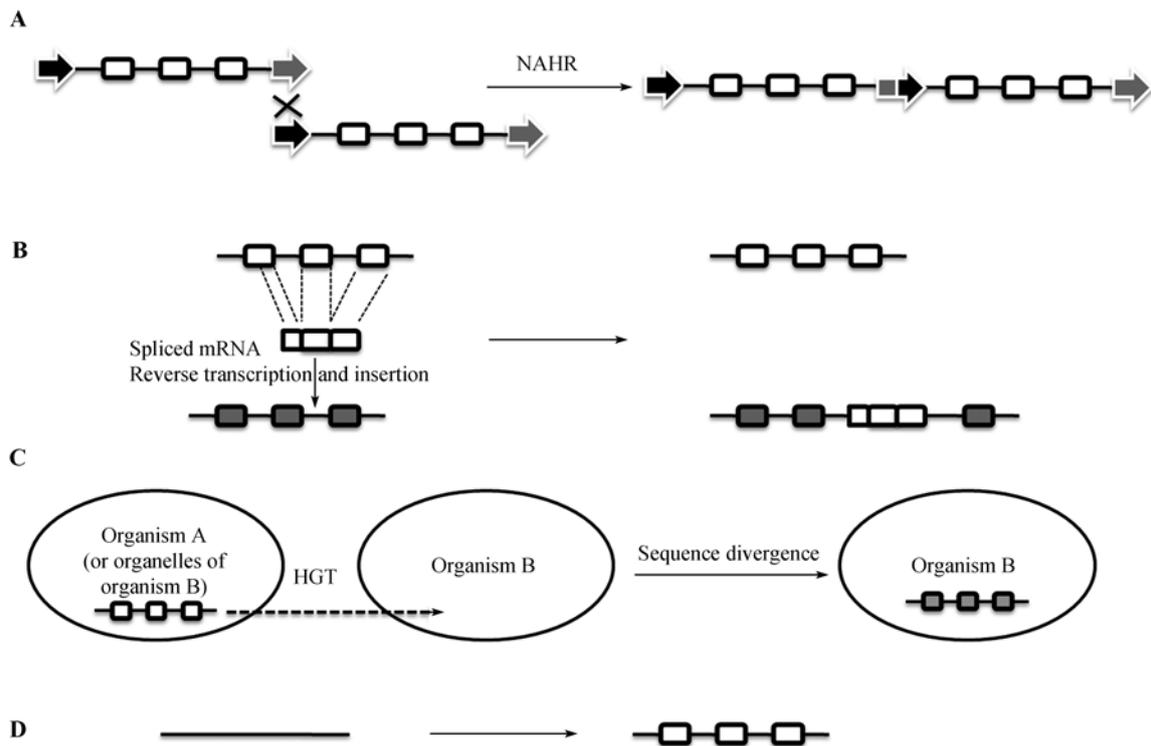


Fig. 1. Mechanisms for creating new genes. Boxes represent exons, while lines represent intronic or intergenic regions. **A:** gene duplication produced by NAHR. The black and grey arrows represent repetitive elements of the same family flanking the duplicated genes. Non-allelic homologous recombination between these elements would lead to birth of new gene copies. **B:** retroposition. When spliced mRNAs of the parental gene are fortuitously reverse transcribed and inserted into another genomic location, it will produce a new gene copy without introns. **C:** horizontal gene transfer. New genes can be introduced into organism B directly from organism A through horizontal gene transfer. **D:** *de novo* origination. New genes can be created from non-coding sequences.

Gene duplication

Gene duplication is the canonical model for origin of new genes (Ohno, 1970). Its tempo is important for understanding the significance of new gene origination. Its underlying mechanisms can provide important insights for genome structure changes besides the substitutions. Therefore, both topics have been extensively studied and corresponding results are debated in various organisms (Lynch and Conery, 2000; Gu et al., 2002; Gao and Innan, 2004; Hahn et al., 2007; Pan and Zhang, 2007; Yang et al., 2008).

There are two methods estimating the age of a gene duplicate and further estimating the duplication rate: molecular-clock based estimation from synonymous substitution rates (K_s) calculated between duplicates and non-clock based method using phylogenetic distribution of gene duplicates (Li, 1997). Lynch and Conery first studied all the gene duplicates in 8 species with whole genome sequences available. They used the former method and estimated the gene duplication rate ranges from 0.0023 to 0.021 per gene per million years (Lynch and Conery, 2000). However, subsequent analysis showed their estimates might suffer several problems. First, gene duplication is not the only mechanism producing new genes (see below), thus, gene duplication rate is not equal to new gene origination rate. Second, gene conversion homogenized sequences of gene duplicates can cause an underestimate of K_s , thus an overestimate of rate using Lynch et al.'s method (Gao and Innan, 2004). Finally, recent studies show that a large number of gene duplicates show copy number polymorphisms (CNP) within populations (Kidd et al., 2008). Such duplicates are not suitable for estimation of a general gene duplication rates. The alternative method, however, does not have the issue of gene conversion. This is because it compares the gene copy numbers between species and measures the age of new genes with species divergence time rather than sequence divergence between duplicates. A recent study used this method and new genes shared by multiple species, which are less likely to show CNPs, and obtained a much lower estimate (0.000391–0.000925 vs. 0.0023) of functional new gene origination rate compared with that of Lynch et al. (Zhou et al., 2008).

On the other hand, the mechanisms producing new duplicated gene copies have been recently investigated in *Drosophila* young genes and primate segmental duplications. New duplicates can arise through non-allelic homologous recombination (NAHR) or illegitimate recombination (IR, or non-homologous end joining, NHEJ) (Roth et al., 1985; Roth and Wilson, 1988; Bishop and Schiestl, 2000; Bailey et al., 2003). Yang et al. (2008) used more than 7,000 cDNA probes and screened 8 *Drosophila* genomes. They identified 17 lineage-specific young duplicates that had originated within 12 million years. Interestingly, they found an excess of repetitive sequences located at the breakpoints of the duplicated regions encompassing

these focal genes. This association suggests NAHR mediated by these repetitive sequences accounted for the birth of new duplicated copies (Fig. 1A). Kidd et al. (2008) studied CNP patterns for eight human individuals with a clone-based sequence strategy. They concluded 41.8% detected duplications are caused by NAHR, whereas 29.6% by IR. It seems that NAHR may play a more important role for producing dispersed duplicates, whereas IR for tandem duplicates in *Drosophila* (Zhou et al., 2008). However, our current conclusions about these two mechanisms are from limited numbers of organisms and are quite speculative. Further studies on gene duplicates from multiple closely-related species should provide more insights into their different roles underlying gene duplication.

Retroposition

Retroposition occurs when transcribed and spliced mRNAs are fortuitously reverse transcribed and inserted into a new genomic location, forming novel gene copies (Fig. 1B). It is different from regular gene duplicates with their features of lacking introns and being flanked by short direct repeats. Usually, such retrocopies are non-expressed pseudogenes, because the reverse transcription didn't include the upstream regulatory sequences with them (Brosius, 1991). However, some retrocopies can recruit nearby regulatory sequences and coding regions from unrelated genes to form chimeric structures. In this way, a functional retrogene with novel gene structure/expression pattern will arise. Such cases have been widely reported in various organisms including *Drosophila* (Long and Langley, 1993; Wang et al., 2002; Betran and Long, 2003), plants, (Drouin and Dover, 1990; Charlesworth et al., 1998) and mammals (Martignetti and Brosius, 1993; Courseaux and Nahon, 2001). Thus, retrocopies were ever called 'seeds of evolution' (Brosius, 1991).

Recent available genome sequences of several organisms further allow the large-scale characterization of functional retrogenes. It is estimated that the emergence rate of retrogenes ranges from 0.5 in *Drosophila* to 7 in rice per million years (Marques et al., 2005; Zhang et al., 2005; Wang et al., 2006; Bai et al., 2007). Such a drastic difference usually correlates with proportions of retroposons of different organisms. Several general conclusions arise from these genome-wide analyses. First, functional retrogenes frequently form chimeric structures. For example, Wang et al. (2006) recently used stringent criteria to screen functional retrogenes in rice. They found 81% retrocopies with intact open reading frames show ratios of nonsynonymous and synonymous substitution rates (K_a/K_s) lower than 0.5, indicating signatures of purifying selection. Also, up to 55% intact retrocopies are able to transcribe with ESTs, RT-PCR, or microarray data. Among them, 42% retrogenes have recruited flanking sequences forming chimeric structures. Similar patterns have also been observed in *Droso-*

phila and human (Marques et al., 2005; Bai et al., 2007). Second, retrogenes often recruited novel regulatory regions, and a lot of them are found to be specifically expressed in testis compared with their ubiquitously-expressed ancestors (Marques et al., 2005; Bai et al., 2007).

Horizontal gene transfer

Horizontal gene transfer (HGT) describes movements of genetic elements across normal mating barriers, i.e., between different species or between organelles and nuclei (Fig. 1C). This mechanism frequently happens between prokaryotes and from organelles to nuclei (Koonin et al., 2001; Keeling and Palmer, 2008). But only a few validated cases have been reported between eukaryotes or from prokaryotes to eukaryotes. This might be because of the fact that most animals have a separated germ line which tends to be sheltered from foreign DNAs (de Koning et al., 2000; Keeling and Palmer, 2008). One intriguing study reported the intracellular endosymbiont *Wolbachia* has transferred its partial or even entire genome to its host insects. However, only 2% transferred genes are able to transcribe (Hotopp et al., 2007), and none of the genes have been proved with function. Therefore, HGTs during origin of eukaryotic new genes are rare and its role and mechanism remain largely uncharacterized so far.

de novo origination

Another formerly unappreciated mechanism, *de novo* origination (Fig. 1D), has been recently shown to be important for origin of new genes (Levine et al., 2006; Begun et al., 2007; Cai et al., 2008). It was believed that fast turnovers of non-coding sequences leading to the birth of functional new genes should be rare, if not absent (Ohno, 1970; Long et al., 2003; Long, 2007). In an influential essay, Jacob even stated “The probability that a functional protein would appear *de novo* by random association of amino acids is practically zero” (Jacob, 1977). However, a pioneer study of five genes originated from non-coding sequences in *Drosophila* suggests *de novo* origination can contribute to origin of new genes (Levine et al., 2006). Moreover, a recent scan for all sorts of new genes originated in *D. melanogaster* has identified a total of nine transcribed *de novo* genes, which comprises 11.9% new genes (Zhou et al., 2008). This proportion is even higher than that of retroposition (see below), indicating the role of *de novo* origination is not ignorable.

The identified *de novo* new genes have several interesting features and are very likely to be functional. First, all of them have intact open reading frames and most have evolved testis-specific expression patterns (Levine et al., 2006; Chen et al., 2007). Second, repetitive elements play an important role during origin of *de novo* genes. For example, a *D. melanogaster de novo* gene CG33235 emerged

through the lineage-specific expansion of short-tandem repeats, forming a gene putatively encoding a protein longer than 1,500 amino acids. Finally, *de novo* genes may first acquire the ability of transcription before becoming protein-coding genes. This can be revealed from two parallel *de novo* gene cases from yeast and *Drosophila* (Cai et al., 2008). Both of them have non-coding orthologous sequences in the closely-related species, which do not contain any intact open reading frames. But EST or RT-PCR results show these orthologous regions are able to transcribe, which suggest the ancestral states of *de novo* genes might be non-coding RNA genes.

A comprehensive picture of origination mechanisms

Several new mechanisms have been reported since Ohno proposed gene duplication as the predominant mechanism for origin of new genes in the 1970s (Ohno, 1970; Long et al., 2003). Subsequent analyses have been focused on case studies of functions of new genes or the genome-wide characterization of an individual mechanism (Long et al., 2003; Levine et al., 2006; Wang et al., 2006; Bai et al., 2007). An unaddressed fundamental question is: how are the relative roles of all these mechanisms contributing to origin of new genes? This question is important for our accurate evaluations of each mechanism and an integrative understanding of all of them. The recently released genome sequences of 12 *Drosophila* species have provided an unprecedented opportunity for studying this question (Clark et al., 2007). Because the phylogeny and divergence times of these species have been well documented (Tamura et al., 2004), a characterization of all the young genes should shed light on origin of new genes at a whole-genome level and in a chronological order.

Zhou et al. (2008) used comparative genomics approaches and identified more than 300 lineage-specific new genes in four *Drosophila* species with short divergence time (within 12.8 million years). Without finding horizontal transferred genes, they divided all the new genes into four categories in the light of their origination mechanisms: tandem duplication, dispersed duplication, retroposition, and *de novo* origination (Table 1). They found: tandem duplication predominantly (about 80%) accounts for the births of nascent copies. However, for new genes shared by multiple species, i.e., those with older ages and are more likely to be functional, are majorly dispersed duplicates. At last, they found *de novo* origination mechanism surprisingly accounted for births of more new genes (11.9%) even than retroposition (10.2%). This work highlighted the remarkable role of *de novo* origination and provided a dynamic view of different types of gene duplication. It represents the first whole-genome exploration of how new genes originated and the obtained major patterns should have general significance for understanding origin of new genes in other species.

Table 1
Contributions of different mechanisms to the origin of new genes in *Drosophila* at the whole-genome level

	Gene duplication		Retroposition	<i>de novo</i>
	Tandem	Dispersed		
<i>D. mel</i>	81.9%	8.3%	6.9%	2.8%
<i>D. mel-D. sch-D. sim</i>	33.9%	44.1%	10.2%	11.9%
<i>D. yak</i>	78.0%	22.0%	NA	NA

'*D. mel*' stands for '*Drosophila melanogaster*', '*D. sch*' for '*D. sechellia*', '*D. sim*' for '*D. simulans*'. These three species together stands for the *D. melanogaster* species complex. Zhou et al. 2008 quantitatively compared contributions of different mechanisms underlying origin of new genes specific to one species (*D. mel* or *D. yak*) or shared by multiple species (*D. mel-D. sch-D. sim*). They didn't characterize retroposed or *de novo* new genes specific to *D. yakuba*, termed 'NA' in the table because of the lack of sufficient annotation information in this species. The Table was modified from Zhou et al. 2008.

How do new genes evolve?

The evolutionary fates of new genes have been extensively studied. Although a lot of new genes are to become pseudogenes, some can be fixed through evolution of redundancy, subfunctionalization, or neofunctionalization (Walsh, 1995; Nowak et al., 1997; Force et al., 1999;

Walsh, 2003) (Fig. 2).

Evolution of redundancies

Evolution of redundancies occurs when natural selection favors the increase of the dosage of genes with same functions (Fig. 2A). Most of our knowledge about this model comes from studies of duplicated genes in yeast because it has the most abundant phenotypic data for knock-out genes to define functionalities. Gu et al. (2003) estimated that at least one quarter of gene deletions without phenotypic effects are caused by the functional compensation between redundant duplicates. And a recent study concluded that yeast can maintain substantial redundant duplicates for about 100 million years (Dean et al., 2008). However, cautions should be taken when extrapolating these conclusions to multicellular eukaryotes. First, approximately 13% yeast genes are ancient duplicates from whole-genome duplications (Wolfe and Shields, 1997). They are different from new genes discussed here by their generation mechanisms, and are less likely subjected to immediate structural changes compared to newly emerged young gene copies after their births (see below). Second, genes previously defined as 'no phenotype' are probably caused by insufficient assays for different conditions (Hillenmeyer et al., 2008).

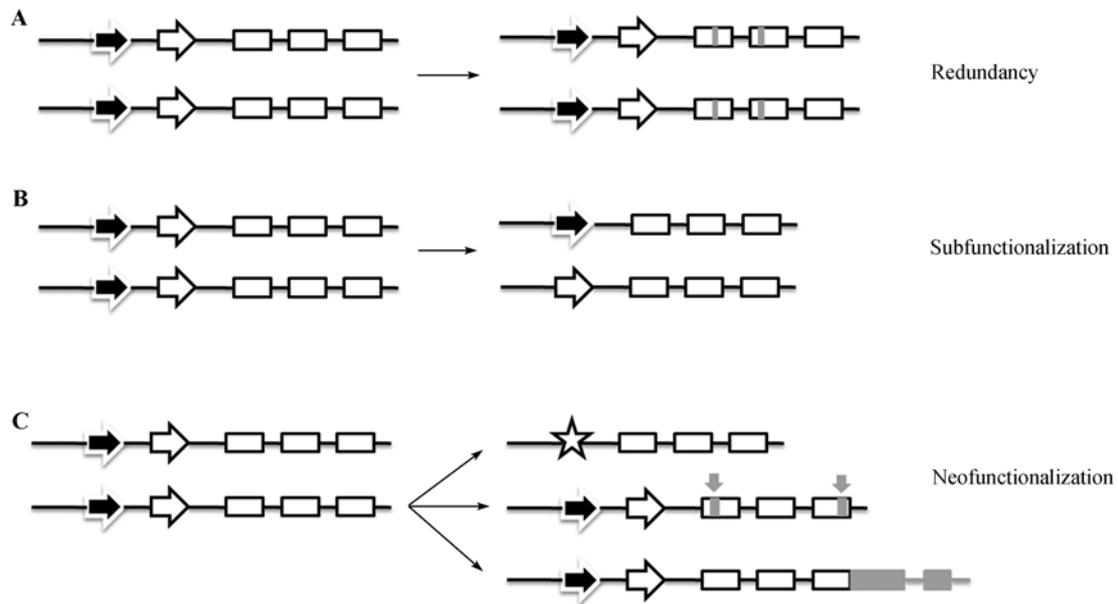


Fig. 2. New gene copies can be fixed within population through three models. Arrows or stars represent regulatory elements, while boxes represent exons. **A:** evolution of redundancies. After the births of new gene copies, substitutions (grey lines) may accumulate within the coding regions, but they retain same functions with their ancestors. **B:** subfunctionalization. Ancestral genes have multiple regulatory elements. After the births of new gene copies, these elements may degenerate in a complementary manner, leading to partitioning of ancestral functions between new and parental genes. **C:** neofunctionalization. This model can occur through evolution of novel regulatory elements which would change the expression patterns of new gene copies. Alternatively, it can happen through adaptive mutations in coding regions such as substitutions or formation of chimeric structures.

Subfunctionalization

Redundant new gene copies are most commonly followed by pseudolization according to Ohno's classical model of gene duplication (Ohno, 1970, 1973). To account for widely found functional new duplicated genes within different organisms, Lynch and Force proposed the model of subfunctionalization, also known as the duplication-degeneration-complementation (DDC) model (Force et al., 1999; Lynch and Force, 2000). It proposed, after the duplication, that the new gene copy would partition the ancestral gene functions together with its parental gene through complementary degeneration of their ancestral regulatory elements (Fig. 2B). Thus, it is expected that new genes that underwent such changes should reveal a diverse expression pattern with their parental genes.

There are only sporadic genome-wide evaluations of roles of subfunctionalization, and a few cases are demonstrated by experiments (Altschmied et al., 2002; Aury et al., 2006; Tumpel et al., 2006; Hittinger and Carroll, 2007; Semon and Wolfe, 2008). One of the most thoroughly studied examples is *GAL1/GAL3* gene duplicates of *Saccharomyces cerevisiae* (Hittinger and Carroll, 2007). This pair of parental and new genes evolved from whole genome duplication (WGD) and are both required for the efficient growth of *S. cerevisiae* when galactose is the only carbon source. Species diverged before WGD, such as *Kluyveromyces lactis*, only have one copy of *GAL1*. Hittinger and Carroll separately swapped promoters and coding regions of *GAL1* and *GAL3*, and developed a sensitive flow-cytometry method to assay the functions and fitness effects of the artificial proteins. They found that new genes' promoter (*GAL3*) can drive parental genes' coding regions (*GAL1*) to partially perform new genes' function. But similar sequence mimicking of new genes in parental gene's coding region didn't make them perform like new genes. This suggests the functional divergence is mostly from regulatory changes. Because *GAL1* in outgroup species *K. lactis* is bifunctional, it is not as specialized as *GAL1* or *GAL3* of *S. cerevisiae* under certain conditions. They suggest this gene pairs evolved through subfunctionalization and further optimized each partitioned function.

Neofunctionalization

Only neofunctionalization involves innovations of gene functions among the proposed models for new genes' evolution. Therefore, they draw more attentions from evolutionary biologists given these cases are more likely to be subjected to adaptive evolution. In addition, most reported cases for the two models mentioned above are from ancient duplication events and insights into important questions, such as how new genes evolved and contributed to human evolution after the short divergence with chimpanzee are more likely to be

gained from studies of recently emerged young genes.

A new gene can alter its ancestral function through evolution of novel expression patterns and substitutions in its coding regions or forming novel chimeric gene structures (Fig. 2C). These processes are not mutually exclusive. Zhang et al. (2007b) recently compared sex-biased expression patterns among 7 *Drosophila* species using microarray. They found genes shared by multiple species (ancestral genes) tend to show female-biased expressions, whereas new genes specific to certain species tend to show male-biased expressions. This result is consistent with previously reported expression patterns for new genes including *Hun*, *hydra*, *Sdic*, *Dntf-2r*, and most of the young retrogenes in human and *Drosophila* (Nurminsky et al., 1998; Betran and Long, 2003; Marques et al., 2005; Arguello et al., 2006; Bai et al., 2007; Chen et al., 2007). They together suggest new genes frequently evolve male-related functions and their evolutions are likely to be driven by sexual selection. The extremely young gene *sphinx* of *D. melanogaster* might be the best illustration and also, the most thoroughly studied new gene reported to date. *sphinx* evolved from retroposition of ATP synthase F-chain (*ATPS-F*) gene followed by recruiting nearby sequences forming a new chimeric structure. Previous studies revealed that it contains no intact open reading frames, but it is able to transcribe and has evolved male and developmental specific splicing forms, suggesting that it also recruited novel regulatory sequences. Sequences analysis found *sphinx* contains indel polymorphism only within its non-exonic regions and reveals a substitution rate significantly higher than neutral expectations. These results suggest *sphinx* is a functional RNA gene driven by positive selection (Wang et al., 2002). Recent functional study revealed *sphinx* specifically express in the male accessory gland, an organ known as regulating reproductive behaviors (Dai et al., 2008). Most intriguingly, homozygous mutant males for *sphinx* would pursue each other for a significantly longer time than wild-types. These males would even form a courtship chain when multiple males are present. As male-male courtship behavior seems to have evolved in the ancestors of *Drosophila* species, *sphinx* is likely recruited and evolved to inhibit it specifically in *D. melanogaster* (Dai et al., 2008). It's noteworthy that such a new gene with novel functions regulating behaviors compared with its parental gene *ATPS-F* evolved within only 2–3 million years, representing quite a classic case for neofunctionalization of new genes (Dai et al., 2008).

In contrast, understanding the genetic basis underlying unique features of human being and primates has been of inherent interest to biologists. Therefore, explorations of new gene functions in human or primates have been rapidly growing by the identification of more and more human or primate specific new genes (Fortna et al., 2004; Marques et al., 2005). Classical cases, such as *FOXP2* and *RNASE1B*, have been extensively studied and reviewed

elsewhere (Enard et al., 2002; Zhang et al., 2002; Long et al., 2003; Sikela, 2006). We intend to cover here cases reported more recently. For example, *clorf37-dup* originated after the split of human with chimpanzee through retroposition. Yu et al. (2006) used *in vitro* cell experiments and found that *clorf37-dup* and its parental gene both localize at plasma membranes. Interestingly, this new gene shows a *Ka/Ks* ratio significantly higher than neutral expectations, indicating its rapid evolution driven by positive Darwinian selection. The substitutions, together with the cell experiments, show the selected sites preferentially located at N-terminal region and extra-cellular loop of the protein, suggesting novel functions have arisen in these regions. Moreover, extensive expression profiles showed *clorf37-dup* evolved novel expression patterns restricted to brain, lung, pancreas, thymus, intestine, and blood, compared with its ubiquitously-expressed ancestral gene. Other recently characterized hominoid new genes with cellular experiments include *GLUD1* and *CDC14Bretro*, both of which evolved diverged subcellular locations compared to their ancestors through substitutions in protein-coding regions (Rosso et al., 2008a, 2008b).

Finally, novel functions can also evolve from innovation of gene structures. Such innovations can be achieved through ‘exon shuffling’ or recruitments of noncoding sequences into new genes’ coding regions (Gilbert, 1978; Long and Langley, 1993; Patthy, 1996, 1999; Wang et al., 2002; Arguello et al., 2006). Such cases include *jingwei* and *sphinx*. Genome-wide characterizations in worm, fly, and rice consistently show that young genes frequently form chimeric structures through recruiting various genomic sources (Katju and Lynch, 2006; Wang et al., 2006), indicating it is a general feature of new gene. Such structures are different from Ohno’s classical model about nascent new genes, which are hypothesized to be functionally and structurally identical to their ancestors. Take *Drosophila* for example, up to 30% new genes are found to have formed chimeric structures probably upon their births, which might immediately confer new copies with novel structures and thus functions and make them to be exposed to the positive natural selection (Zhou et al., 2008). A recently reported case is *Hun*, which evolved within the last 2–3 million years. Its parental gene, *Ballchen* locates on chromosome 3R and it produced a partial duplication onto chromosome X possibly through illegitimate recombination. Subsequent incorporations of nearby intergenic sequences create the novel chimeric protein with about 33 new amino acids of *Hun*. Interestingly, *Hun* also evolves testis-specific expression compared to ubiquitously expressed *Ballchen* (Arguello et al., 2006).

Overall, these cases or genome-wide patterns illustrate that new genes usually show accelerated sequence/structure changes compared with their ancestors. Together with altered expression patterns, these general features probably expose the new genes to positive selections and further more make them fix within the populations.

Future directions

Our knowledge about how new gene originated and evolved have greatly been expanding since the characterization of the first new gene *jingwei* in 1993 (Long et al., 2003). Genome-wide patterns, as well as rising numbers of case studies have provided new insights into this classical but fast-growing research field (Long et al., 2003; Zhang et al., 2007b). We try to pose and discuss several largely uncharacterized questions here to inspire the future directions, which may be fundamental and of general interests. First, *de novo* genes have been recently proved to be an important source for new genes. Such drastic turnovers of non-coding sequences leading to the birth of a new gene within short evolutionary time contrast with the classical models. Is this a general phenomenon widely distributed in other organisms? If so, what are their functions and how were they integrated into the pre-existing biological pathways to contribute to the adaptive evolution of organisms? The second question is for all types of new genes rather than being restricted for *de novo* genes. With the application of novel transcriptomic and proteomic methods, we are able to go beyond sequence or structure analysis of new genes. We are allowed to systematically compare hundreds of new genes’ expression profiles with their ancestors, or strains with new genes deleted. Together with subtle phenotypic assays, we can gain deeper insights about new genes’ functions and their interactions with other genes. Third, *sphinx* represents a good case for novel RNA genes affecting organisms’ reproductive behaviors (Dai et al., 2008). Other cases of novel RNA/micro-RNA genes have been recently characterized in human and results showed they are specifically expressed in developing human brains or correlated with sexual maturation (Pollard et al., 2006; Zhang et al., 2007a). However, because of the limited annotations for transcriptomes of multiple closely-related species, previous studies mainly focused on the identification and characterization of novel protein-coding genes. We currently know little about origin and evolutionary patterns of novel RNA genes. Fortunately, the revolutionary ‘next-generation sequencing’ technologies such as Solexa and 454 would allow us to quickly compare transcriptomes of closely related species with much lower costs (Marioni et al., 2008). A recent such exploration compared microRNA contents of three *Drosophila* species and estimated 0.3 novel microRNA genes would arise per million years (Lu et al., 2008). The upcoming genome-wide characterizations of novel microRNA or RNA genes in other species, especially in human, and functional studies of identified cases may represent the future directions of new gene studies.

Acknowledgements

This work was supported by a CAS-Max Planck Soci-

ety Fellowship, an award (No. 30325016) of the National Science Foundation of China (NSFC), two NSFC key grants (No. 30430400 and 30623007), a grant from the National Basic Research Program of China (973 Program) (No. 2007CB815703-5) to W.W., and a NSFC grant (No. 30500283) for junior researchers to S.Y.

References

- Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Voff, J.N., and Scharl, M. (2002). Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics* **161**: 259–267.
- Arguello, J.R., Chen, Y., Yang, S., Wang, W., and Long, M. (2006). Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. *PLoS Genet.* **2**: e77.
- Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aïach, N., Arnai, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Durharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J., and Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171–178.
- Bai, Y., Casola, C., Feschotte, C., and Betran, E. (2007). Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biol.* **8**: R11.
- Bailey, J.A., Liu, G., and Eichler, E.E. (2003). An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**: 823–834.
- Begun, D.J., Lindfors, H.A., Kern, A.D., and Jones, C.D. (2007). Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta* clade. *Genetics* **176**: 1131–1137.
- Betran, E., and Long, M. (2003). *Dntf-2r*, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics* **164**: 977–988.
- Bishop, A.J.R., and Schiestl, R.H. (2000). Homologous recombination as a mechanism for genome rearrangements: Environmental and genetic effects. *Hum. Mol. Genet.* **9**: 2427–2334.
- Brosius, J. (1991). Retroposons—seeds of evolution. *Science* **251**: 753.
- Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). *de novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* **179**: 487–496.
- Charlesworth, D., Liu, F.L., and Zhang, L. (1998). The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (*Brassicaceae*). *Mol. Biol. Evol.* **15**: 552–559.
- Chen, S.T., Cheng, H.C., Barbash, D.A., and Yang, H.P. (2007). Evolution of *hydra*, a recently evolved testis-expressed gene with nine alternative first exons in *Drosophila melanogaster*. *PLoS Genet.* **3**: e107.
- Clark, A.G. (1994). Invasion and maintenance of a gene duplication. *Proc. Natl. Acad. Sci. USA* **91**: 2950–2954.
- Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., Pollard, D.A., Sackton, T.B., Larracuente, A.M., Singh, N.D., Abad, J.P., Abt, D.N., Adryan, B., Aguade, M., Akashi, H., Anderson, W.W., Aquadro, C.F., Ardell, D.H., Arguello, R., Artieri, C.G., Barbash, D.A., Barker, D., Barsanti, P., Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S.A., Bradley, R.K., Brand, A.D., Brent, M.R., Brooks, A.N., Brown, R.H., Butlin, R.K., Caggese, C., Calvi, B.R., Bernardo de Carvalho, A., Caspi, A., Castrezana, S., Celniker, S.E., Chang, J.L., Chapple, C., Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S.W., Comeran, J.M., Costello, J.C., Coyne, J.A., Daub, J., David, R.G., Delcher, A.L., Delehaanty, K., Do, C.B., Ebling, H., Edwards, K., Eickbush, T., Evans, J.D., Filipowski, A., Findeiss, S., Freyhult, E., Fulton, L., Fulton, R., Garcia, A.C., Gardiner, A., Garfield, D.A., Garvin, B.E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J., Good, R., Gotea, V., Gravely, B., Greenberg, A.J., Griffiths-Jones, S., Gross, S., Guigo, R., Gustafson, E.A., Haerty, W., Hahn, M.W., Halligan, D.L., Halpern, A.L., Halter, G.M., Han, M.V., Heger, A., Hillier, L., Hinrichs, A.S., Holmes, I., Hoskins, R.A., Hubisz, M.J., Hultmark, D., Huntley, M.A., Jaffe, D.B., Jagadeeshan, S., Jeck, W.R., Johnson, J., Jones, C.D., Jordan, W.C., Karpen, G.H., Kataoka, E., Keightley, P.D., Kheradpour, P., Kirkness, E.F., Koerich, L.B., Kristiansen, K., Kudrna, D., Kulathinal, R.J., Kumar, S., Kwok, R., Lander, E., Langley, C.H., Lapoint, R., Lazzaro, B.P., Lee, S.J., Levesque, L., Li, R., Lin, C.F., Lin, M.F., Lindblad-Toh, K., Llopart, A., Long, M., Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C.A., Makalowski, W., Marzo, M., Matsuda, M., Matzkin, L., McAllister, B., McBride, C.S., McKernan, B., McKernan, K., Mendez-Lago, M., Minx, P., Mollenhauer, M.U., Montooth, K., Mount, S.M., Mu, X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M.A., O'Grady, P., Pachter, L., Papaceit, M., Parisi, M.J., Parisi, M., Parts, L., Pedersen, J.S., Pesole, G., Phillippy, A.M., Ponting, C.P., Pop, M., Porcelli, D., Powell, J.R., Prohaska, S., Pruitt, K., Puig, M., Quesneville, H., Ram, K.R., Rand, D., Rasmussen, M.D., Reed, L.K., Reenan, R., Reily, A., Remington, K.A., Rieger, T.T., Ritchie, M.G., Robin, C., Rogers, Y.H., Rohde, C., Rozas, J., Rubenfield, M.J., Ruiz, A., Russo, S., Salzberg, S.L., Sanchez-Gracia, A., Saranga, D.J., Sato, H., Schaeffer, S.W., Schatz, M.C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R.S., Sirot, L., Sirot, M., Sisneros, N.B., Smith, C.D., Smith, T.F., Spieth, J., Stage, D.E., Stark, A., Stephan, W., Strausberg, R.L., Strempel, S., Sturgill, D., Sutton, G., Sutton, G.G., Tao, W., Teichmann, S., Tobari, Y.N., Tomimura, Y., Tsolas, J.M., Valente, V.L., Venter, E., Venter, J.C., Vicario, S., Vieira, F.G., Vilella, A.J., Villasante, A., Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R.K., Wing, R.A., Wolfner, M.F., Wong, A., Wong, G.K., Wu, C.I., Wu, G., Yamamoto, D., Yang, H.P., Yang, S.P., Yorke, J.A., Yoshida, K., Zdobnov, E., Zhang, P., Zhang, Y., Zimin, A.V., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Acer, S.C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I., Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M., Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray, S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P., Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher, S., Foley, C.D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C.R., Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K., Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M.D., Hughes, L., Hurhula, B., Husby, M.E., Kamat, A., Kanga, B., Kashin, S., Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F., LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y., Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C., McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O., Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L., Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C., Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O.L., Patti, C., Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise, C., Rogov, P., Ross, K., Ryan, E., Settippalli, S., Shea, T., Sherpa, N., Shi, L., Shih, D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann,

- N., Stavropoulos, S., Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiland, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Alvarez, P., Brockman, W., Butler, J., Chin, C., Grabherr, M., Kleber, M., Mauceli, E., and MacCallum, I. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.
- Courseaux, A., and Nahon, J.L. (2001). Birth of two chimeric genes in the Hominidae lineage. *Science* **291**: 1293–1297.
- Dai, H., Chen, Y., Chen, S., Mao, Q., Kennedy, D., Landback, P., Eyre-Walker, A., Du, W., and Long, M. (2008). The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **105**: 7478–7483.
- de Koning, A.P., Brinkman, F.S., Jones, S.J., and Keeling, P.J. (2000). Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**: 1769–1773.
- Dean, E.J., Davis, J.C., Davis, R.W., and Petrov, D.A. (2008). Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet.* **4**: e1000113.
- Drouin, G., and Dover, G.A. (1990). Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. *J. Mol. Evol.* **31**: 132–150.
- Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., and Paabo, S. (2002). Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**: 869–872.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., Brenton, M., Hink, R., Burgers, S., Hernandez-Boussard, T., Karimpour-Fard, A., Glueck, D., McGavran, L., Berry, R., Pollack, J., and Sikela, J.M. (2004). Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**: e207.
- Gao, L., and Innan, H. (2004). Very low gene duplication rate in the yeast genome. *Science* **306**: 1367–1370.
- Gilbert, W. (1978). Why genes in pieces? *Nature* **271**: 501.
- Gu, Z., Cavalcanti, A., Chen, F.-C., Bouman, P., and Li, W.-H. (2002). Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.* **19**: 256–262.
- Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., and Li, W.H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**: 63–66.
- Hahn, M.W., Han, M.V., and Han, S.-G. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* **3**: e197.
- Haldane, J.B.S. (1932). *The causes of evolution*. (London: Longmans and Green).
- Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., Altman, R.B., Davis, R.W., Nislow, C., and Giaever, G. (2008). The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* **320**: 362–365.
- Hittinger, C.T., and Carroll, S.B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**: 677–681.
- Hotopp, J.C.D., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Torres, M.C.M., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., and Werren, J.H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**: 1753–1756.
- Jacob, F. (1977). Evolution and tinkering. *Science* **196**: 1161–1166.
- Katju, V., and Lynch, M. (2006). On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* **23**: 1056–1067.
- Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**: 605–618.
- Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tuzun, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., Gillett, W., Phelps, K.A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J.D., Korn, J.M., McCarroll, S.A., Altshuler, D.A., Peiffer, D.A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D.A., Mullikin, J.C., Wilson, R.K., Bruhn, L., Olson, M.V., Kaul, R., Smith, D.R., and Eichler, E.E. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**: 56–64.
- Koonin, E.V., Makarova, K.S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* **55**: 709–742.
- Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A., and Begun, D.J. (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. USA* **103**: 9935–9939.
- Li, W.H. (1997). *Molecular Evolution*. (Sunderland).
- Long, M. (2007). *Journal club*. *Nature* **449**: 511–511.
- Long, M., and Langley, C.H. (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* **260**: 91–95.
- Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**: 865–875.
- Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., and Wu, C.-I. (2008). The birth and death of microRNA genes in *Drosophila*. *Nat. Genet.* **40**: 351–355.
- Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
- Marioni, J., Mason, C., Mane, S., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**: 1509–1517.
- Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., and Kaessmann, H. (2005). Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **3**: e357.
- Martignetti, J.A., and Brosius, J. (1993). *BC200* RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl. Acad. Sci. USA* **90**: 11563–11567.
- Muller, H.J. (1935). The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetica* **17**: 237–252.
- Nowak, M.A., Boerlijst, M.C., Cooke, J., and Smith, J.M. (1997). Evolution of genetic redundancy. *Nature* **388**: 167–171.
- Nurminsky, D.I., Nurminskaya, M.V., De Aguiar, D., and Hartl, D.L. (1998). Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**: 572–575.
- Ohno, S. (1970). *Evolution by gene duplication*. (New York: Springer-Verlag).
- Ohno, S. (1973). Ancient linkage groups and frozen accidents. *Nature* **244**: 259–262.
- Pan, D., and Zhang, L. (2007). Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: A novel strategy to estimate gene duplication rates. *Genome Biol.* **8**: R158.
- Patthy, L. (1996). Exon shuffling and other ways of module exchange. *Matrix Biol.* **15**: 301–310; discussion 311–312.
- Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling—a review. *Gene* **238**: 103–114.
- Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A.D., Dehay, C., Igel, H., Ares, M., Jr., Vanderhaeghen, P., and Haussler, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**: 167–172.
- Rosso, L., Marques, A.C., Reichert, A.S., and Kaessmann, H. (2008a).

- Mitochondrial targeting adaptation of the hominoid-specific glutamate dehydrogenase driven by positive Darwinian selection. *PLoS Genet.* **4**: e1000150.
- Rosso, L., Marques, A.C., Weier, M., Lambert, N., Lambot, M.A., Vanderhaeghen, P., and Kaessmann, H.** (2008b). Birth and rapid subcellular adaptation of a hominoid-specific *CDC14* protein. *PLoS Biol.* **6**: e140.
- Roth, D., and Wilson, J.** (1988). Illegitimate recombination in mammalian cells. Genetic recombination. American Society for Microbiology, Washington, DC, 621–653.
- Roth, D.B., Porter, T.N., and Wilson, J.H.** (1985). Mechanisms of non-homologous recombination in mammalian cells. *Mol. Cell. Biol.* **5**: 2599–2607.
- Semon, M., and Wolfe, K.H.** (2008). Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* **105**: 8333–8338.
- Sikela, J.M.** (2006). The jewels of our genome: The search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genet.* **2**: e80.
- Tamura, K., Subramanian, S., and Kumar, S.** (2004). Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**: 36–44.
- Tumpel, S., Cambroner, F., Wiedemann, L.M., and Krumlauf, R.** (2006). Evolution of *cis* elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc. Natl. Acad. Sci. USA* **103**: 5419–5424.
- Walsh, B.** (2003). Population-genetic models of the fates of duplicate genes. *Genetica* **118**: 279–294.
- Walsh, J.B.** (1995). How often do duplicated genes evolve new functions? *Genetics* **139**: 421–428.
- Wang, W., Brunet, F.G., Nevo, E., and Long, M.** (2002). Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **99**: 4448–4453.
- Wang, W., Yu, H., and Long, M.** (2004). Duplication-degeneration as a mechanism of gene fission and the origin of new genes in *Drosophila* species. *Nat. Genet.* **36**: 523–527.
- Wang, W., Zhang, J., Alvarez, C., Llopart, A., and Long, M.** (2000). The origin of the *Jingwei* gene and the complex modular structure of its parental gene, yellow emperor, in *Drosophila melanogaster*. *Mol. Biol. Evol.* **17**: 1294–1301.
- Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., Zhang, G., Liu, D., Zhang, J., Vang, S., Lu, Z., Wong, G.K., Long, M., and Wang, J.** (2006). High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**: 1791–1802.
- Wolfe, K.H., and Shields, D.C.** (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
- Yang, S., Arguello, J.R., Li, X., Ding, Y., Zhou, Q., Chen, Y., Zhang, Y., Zhao, R., Brunet, F., Peng, L., Wang, W., and Long, M.** (2008). Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genet.* **4**: e3.
- Yu, H., Jiang, H., Zhou, Q., Yang, J., Cun, Y., Su, B., Xiao, C., and Wang, W.** (2006). Origination and evolution of a human-specific transmembrane protein gene, *clorf37*-dup. *Hum. Mol. Genet.* **15**: 1870–1875.
- Zhang, J., Zhang, Y., and Rosenberg, H.F.** (2002). Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* **30**: 411–415.
- Zhang, R., Peng, Y., Wang, W., and Su, B.** (2007a). Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res.* **17**: 612–617.
- Zhang, Y., Sturgill, D., Parisi, M., Kumar, S., and Oliver, B.** (2007b). Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* **450**: 233–237.
- Zhang, Y., Wu, Y., Liu, Y., and Han, B.** (2005). Computational identification of 69 retroposons in *Arabidopsis*. *Plant Physiol.* **138**: 935–948.
- Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W.** (2008). On the origin of new genes in *Drosophila*. *Genome Res.* **18**: 1446–1455.