

# Oxford Bibliographies

Your Best Research Starts Here



## Evolution of New Genes

Manyuan Long

### Introduction

Every gene has its first day in evolution. Thus it is no surprise that gene origination is viewed as a general evolutionary process. The outcomes of these origination events, in which new genes are created in a genome, add evolutionary novelties in genetics and functions to organisms. New genes have attracted increasing attention among evolutionary biologists and laymen, impacting studies in multiple fields of biology and public understanding of evolution. The studies of new genes have attempted to reveal the origination process, the underlying mechanisms, and the consequences on evolution of species. It has become obvious that with more known about the specifics in evolution and biological properties of new genes, more challenging and interesting scientific problems have been also raised, rendering the studies a more active field.

### General Overviews

Many authors have contributed detailed reviews of new gene evolution since the first general overview appeared in Long, et al. 2003, which was updated in Long, et al. 2013. This article attempts to provide literature for an overarching framework for understanding the general concepts and mechanisms of new gene evolution. For detailed reviews of studies of new genes generated by specific mutational processes, see Kaessmann, et al. 2009 (cited under RNA-Based Gene Duplication) and Innan and Kondrashov 2010. For overviews of the consequences of multiple mutational mechanisms of new gene formation, see Kaessmann 2010 and Ding, et al. 2012. Losos, et al. 2013 and Long, et al. 2013 provide comprehensive overviews of new gene evolution and its related issues. Chen, et al. 2013 reviews progress in understanding the role of new genes in phenotypic evolution.

**Chen, S. D., B. Krinsky, and M. Long. 2013. New genes as a driver of phenotypic evolution. *Nature Reviews Genetics* 14:645–660.**

A comprehensive review and discussion of the role of new genes in phenotypic evolution. Available online for purchase or by subscription.

**Ding, Y., Q. Zhou, and W. Wang. 2012. Origins of new genes and evolution of their novel functions. *Annual Review of Ecology, Evolution, and Systematics* 43:345–363.**

In this review, the molecular functions of new genes that originated through various mechanisms were summarized. Available online for purchase or by subscription.

**Innan, H., and F. Kondrashov. 2010. The evolution of gene duplications: Classifying and distinguishing between models. *Nature Reviews Genetics* 11:97–108.**

Among many reviews on the topic of gene duplication, this review stood out as a model-classification analysis for various consequences of gene duplication including neofunctionalization. Available online for purchase or by subscription.

**Kaessmann, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research* 20:1313–1326.**

This review called attention to the phenotypic effects of new genes that were created through several mechanisms.

**Long, M., E. Betrán, K. Thornton, and W. Wang. 2003. The origin of new genes: Glimpses from the young and old. *Nature Reviews Genetics* 4.11: 865–875.**

This early review provided the first complete picture of the active experimental research and evolutionary analysis of new genes. Available online for purchase or by subscription.

**Long, M., N. Van Kuren, S. D. Chen, and M. Vibranovski. 2013. Origination and evolution of new genes. *Annual Review of Genetics* 47:1681–1686.**

An update of the Long, et al. 2003 review of the entire field of new gene evolution. Available online for purchase or by subscription.

**Losos, J., D. Baum, D. Futuyma, et al. 2013. *Princeton guide to evolution*. Princeton, NJ: Princeton Univ. Press.**

This is a rich collection of chapters for educational purpose, including the topic of new gene evolution and related issues in molecular evolution.

---

## Journals

Cross-disciplinary approaches are often used to study new gene origination, and new gene studies are of general interest to broad audiences. A wide spectrum of journals therefore publishes papers on new gene evolution, ranging from journals of evolution and genetics to molecular biology. *Molecular Biology and Evolution*, *Genetics*, and the *Journal of Molecular Evolution* have contributed a large proportion of these papers, as have general biological journals such as *PLOS Biology*, *PLOS Genetics*, and *Proceedings of the National Academy of Sciences of the United States of America*. In addition, relevant publications can be found in many other biological journals. The general scientific magazines *Science* and *Nature* have not been stingy in providing space to the area, while *Nature Reviews Genetics* is enthusiastic in promoting exchange of the research progresses in the field in its highlights and reviews articles.

**Genetics. 1916–.**

As a peer-edited publication of the Genetics Society of America, *Genetics* mostly publishes research articles and short communications, research highlights, and other commentaries.

**Journal of Molecular Evolution. 1971–.**

This journal publishes research reports of molecular evolution including new gene evolution.

**Molecular Biology and Evolution. 1983–.**

The journal of the international Society of Molecular Biology and Evolution, publishing all types of papers, including research reports (long articles and short communications), reviews, and commentaries.

**Nature. 1869–.**

A general magazine published by the Nature Publication Group, publishing original research articles, commentaries, and the news related to scientific research.

**Nature Reviews Genetics. 2000–.**

A review journal published by the Nature Publication Group, publishing reviews, commentaries, and the news in genetic research.

**PLOS Biology.**

This peer-reviewed, open-access journal featuring research articles of exceptional significance in all areas of biological science published by the Public Library of Science. Its flagship, *PLOS Biology*, as well as *PLOS ONE*, and the more specialized *PLOS Genetics*, have published papers in the field of new gene evolution.

**Proceedings of the National Academy of Sciences of the United States of America. 1914–.**

In this weekly, the papers from all fields of scholarship are published, mostly in biological and biomedical fields. The papers containing studies of new genes are in the sections on evolution and genetics.

**Science. 1880–.**

A general magazine published by the American Association for the Advancement of Science, publishing original research articles, commentaries, and news regarding science.

---

**Defining New Genes**

A new gene is an evolutionarily new locus that harbors a distinct gene sequence and carries out a function not present before the new locus arose. Clearly the definition of “evolutionarily new” will depend on what evolutionary time scale we are concerned with. Long, et al. 2013 presents the first definition of new genes in a broad sense.

**Long, M., S. D. Chen, N. Van Kuren, and M. Vibranovski. 2013. Origination and evolution of new genes. *Annual Review of Genetics* (2013).**

This is an update on the topic and definition of new genes, which includes new protein-coding genes and new non-protein-coding genes.

---

## History

Although a number of different types of new genes have recently been discovered, one of these types of new genes—the new gene that is created by gene duplication—had already attracted the interest of the pioneers of evolutionary and genetic studies in the first half of the 20th century. For several decades, there were no other models for the origin of new genes except the duplication-based model proposed in Muller 1936 and elaborated in Ohno 1970, until Gilbert 1978 presented a recombination-based model. A systematic analysis of young genes to detect the evolutionary events in the early stage of gene origination started in 1993 when the first young chimeric gene was found in a couple of African *Drosophila* species, as reported in Long and Langley 1993. With the advent of rapid sequencing techniques that allowed detailed analysis of gene structures, more molecular mechanisms were reported, as cited under General Overviews. Unexpected from the conventional belief, a novel functional protein was found to come into being from noncoding DNAs by *de novo* origination (Levine, et al. 2006). With more new genes analyzed, it was found that there were distinct rules or patterns of new gene origination (e.g., Betrán, et al. 2002; Emerson, et al. 2004). Chen, et al. 2010 finds that new genes have rapidly evolved essential developmental functions, suggesting a genetic basis for rapid phenotypic evolution and opening another dimension of research for the field.

**Betrán, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Research* 12:1854–1859.**

This is the first report that there is a significant pattern of gene trafficking associated with the origination of new retrogenes in *Drosophila*: there is an excess of retrogenes that were copied from the X-linked parents to autosomes, and these new autosomal retrogenes evolved testis-biased expression. This pattern made a surprising prediction on the distribution of sex-related genes between the X chromosome and autosomes, which was confirmed by several studies that characterized sex-biased genes in the genomes of *Drosophila* and mouse.

**Chen, S., Y. Zhang, and M. Long. 2010. New genes in *Drosophila* quickly become essential. *Science* 330:1682–1685.**

The first large-scale experimental genetic exploration of the relationship between the age and functional importance of genes in *Drosophila*, revealing that new genes quickly evolve essential developmental functions. Available online for purchase or by subscription.

**Emerson, J. J., H. Kaessmann, E. Betran, and M. Y. Long. 2004. Extensive gene traffic on the mammalian X chromosome. *Science* 303:537–540.**

The first report of gene trafficking between the X and autosomes in mammals. This analysis revealed both similarity and difference in the gene trafficking between mammals and

*Drosophila*. In both systems, there is an excess of testis-biased retrogenes copied from the X-linked parental genes to the autosomes. Available online for purchase or by subscription.

**Gilbert, W. 1978. Why genes in pieces? *Nature* 271:501.**

A widely accepted new gene origination model based on gene recombination (domain/exon shuffling), an entirely different model from the duplication-based model. Available online for purchase or by subscription.

**Levine, M. T., C. D. Jones, A. D. Kern, H. A. Lindfors, and D. J. Begun. 2006. 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* 103:9935–9939.**

This important paper showed that several new protein-coding genes with regular structure and complexity were generated *de novo* from noncoding DNAs in *Drosophila*. Since then, *de novo* genes have been found in various organisms.

**Long, M., and C. H. Langley. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95.**

The first finding of the young gene, revealing previously expected features of new gene origination in both the underlying molecular processes and evolutionary forces. This study exemplified the advantage of examining young genes to make it possible to observe the early evolutionary processes and forces that led to the origin of new genes. Available online for purchase or by subscription.

**Muller, H. 1936. Bar duplication. *Science* 83:528–530.**

Although the Bar duplication has never evolved into a new gene with beneficial novel functions, this did not prevent Muller from imagining the duplication-based model of new gene evolution, which was the first to be described and is still the most prevalent mechanism for the evolution of new genes. Available online for purchase or by subscription.

**Ohno, S. 1970. *Evolution by gene duplication*. New York: Springer-Verlag.**

Further extension and elaboration of the role of gene duplication in evolution of genes and genomes made this an influential book.

---

## Approaches

Studies of new gene evolution have relied heavily on concepts and analyses from classic evolutionary, genetic, molecular biology, statistical, and computational approaches. Molecular evolution, population genetics, and experimental genomics have been particularly important for the development of the field. However, specific conceptual and experimental approaches have also been required to understand the evolution of new genes. These specific approaches will be emphasized in this section. For information on

general approaches, see Li 1997; Hartl and Clark 2007; Charlesworth and Charlesworth 2010; and Losos, et al. 2013.

**Charlesworth, B., and D. Charlesworth. 2010. *Elements of evolutionary genetics*. Greenwood Village, CO: Roberts.**

The most complete introduction to evolutionary genetics.

**Hartl, D. L., and A. G. Clark. 2007. *Principles of population genetics*. 4th ed. Sunderland, MA: Sinauer.**

A widely used reference book of population genetics and especially the molecular population genetics that is often used in the study of new gene evolution.

**Li, W.-H. 1997. *Molecular evolution*. Sunderland, MA: Sinauer.**

The widely used textbook provides a complete discussion of various related theories and methods of molecular evolution.

**Losos, J., D. Baum, D. Futuyma, et al. 2013. *Princeton guide to evolution*. Princeton, NJ: Princeton Univ. Press.**

Almost all major domains of evolution are discussed in the volume.

## THE VALUE OF YOUNG GENES

Many signatures of evolution are often lost over time due to mutation saturation. To understand the evolution of new genes, many authors have studied genes that recently originated. This approach of young gene study has been demonstrated to be efficient in the analyses of multiple new genes that independently originated from the parental alcohol dehydrogenase (*Adh*) genes in different *Drosophila* species. These findings of new genes from the same parental *Adh* genes were enabled by the enthusiasm of early molecular and evolutionary researchers on the *Adh* gene in fruit flies. Nozawa, et al. 2005 reports the fourth *Adh*-derived new gene in *Drosophila*. Shih and Jones 2008 and Jones and Begun 2005 identified all parallel changes in these genes. The value of young genes can be also seen from Yang, et al. 2008, which finds the role of repetitive elements in initiating gene duplication; Wang, et al. 2004, who identified the intermediate states between gene fusion and gene fission in the *D. melanogaster* subgroup species; and Wang, et al. 2000, who reported ancestral sequences that contributed to the origin of *Jingwei*, the first *Adh*-derived new gene.

**Jones, C. D., and D. J. Begun. 2005. Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences of the United States of America* 102:11373–11378.**

In Jones and Begun 2005, a clear parallel evolutionary pattern was identified from a highly resolved sequence data of all changes that have been retained in the four chimeric genes derived from independent duplication of the *Adh* gene in *Drosophila* species.

**Nozawa, M., T. Aotsuka, and K. Tamura. 2005. A novel chimeric gene, *siren*, with retroposed promoter sequence in the *Drosophila bipectinata* complex. *Genetics* 171:1719–1727.**

This paper reported a young chimeric gene, *siren*, which independently evolved from *Adh* in a *Drosophila* species group related to the model species of *Drosophila*, *D. melanogaster*.

**Shih, H. J., and C. D. Jones. 2008. Patterns of amino acid evolution in the *Drosophila ananassae* chimeric gene, *siren*, parallel those of other *Adh*-derived chimeras. *Genetics* 180.2 (October): 1261–1263.**

Shih and Jones identified the new homolog of *siren* in *D. ananassae* and analyzed the evolutionary force that shaped the gene's amino acid substitution pattern.

**Wang, W., H. Yu, and M. Long. 2004. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in *Drosophila* species. *Nature Genetics* 36:523–527.**

This work took advantage of the expression data, the difference in substitution rates between synonymous and replacement sites, the length polymorphisms outside and inside coding regions, and demonstrated that the new genes created by gene fission are functional.

**Wang, W., J. Zhang, C. Alvarez, A. Llopart, and M. Long. 2000. The origin of the *jingwei* gene and the complex modular structure of its parental gene, *yellow emperor*, in *Drosophila melanogaster*. *Molecular Biology and Evolution* 17.9 (September): 1294–1301.**

This analysis of the *jingwei* locus and its parental genes identified the 3' pseudoexons of *yande*, an ancestral gene that contributed to the N-terminal peptides of *Jingwei*. The *yande* gene was duplicated from the *yellow emperor* gene and has retained high sequence identity with *yellow emperor*, thus the duplication event that was involved in the origin of *jingwei* could be clearly identified. Available online for purchase or by subscription.

**Yang, S., R. Arguello, X. Li, et al. 2008. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLOS Genetics* 4.1: e3.**

In this study, young gene duplicates that originated within a few million years have kept the flanking transposable elements (TEs), which show continuous decay in sequence before their eventual disappearance from the genome. These extant young TE sequences provide clear evidence for their role in the duplication event.

## IDENTIFYING NEW GENES

The general approach to finding new genes is to use comparative analyses of the phylogenetic distribution of genes in a group of related species. A gene that exists only in a group of recently diverged species is a candidate new gene. Ranz and Parsch 2012; Zhang, et al. 2010; Zhou, et al. 2008; Bai, et al. 2007; and Demuth, et al. 2006 describe multiple related species-based comparative approaches to identify new genes. When the genome of only one species is available, new retrogenes can be inferred using the different exon-intron structures of the parent and child genes. The authors of Betrán, et al. 2002 and Marques, et al. 2005 developed computational methods to identify new retrogenes in the genomic sequence data of a single species and used these methods in *Drosophila* and humans, respectively.

**Bai, Y., C. Casola, C. Feschotte, and E. Betrán. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biology* 8.1: R11.**

The authors combined the phylogenetic-distribution approach and the gene structural approach to identify new retrogenes in the multiple lineages of *Drosophila* species.

**Betrán, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Research* 12:1854–1859.**

This early work identified new genes in the genome of a single species, *D. melanogaster*, based on the derived features of gene structure in the newly created copy through retroposition from a parental copy and sequence similarity among paralogs.

**Demuth, J. P., T. DeBie, J. E. Stajich, N. Cristianini, and M. W. Hahn. 2006. The evolution of mammalian gene families. *PLOS ONE* 1.1: e85.**

This study of mammalian gene families developed a computational method to identify new gene duplicates based on the phylogenetic distribution of homologous sequences.

**Marques, A. C., I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann. 2005. Emergence of young human genes after a burst of retroposition in primates. *PLOS Biology* 3.11: e357.**

This paper identified parent-daughter genes by comparative analyses of intron-presence and intron-absence in a pair of duplicates, in conjunction with the sequence divergence information among the paralogs.

**Ranz, J. M., and J. Parsch. 2012. Newly evolved genes: Moving from comparative genomics to functional studies in model systems. *BioEssays* 34:477–483.**

A summary of the general method to identify newly evolved genes based on the phylogenetic distribution of homologous genes. Available online for purchase or by subscription.

**Zhang, Y., M. D. Vibranovski, P. Landback, and M. Long. 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLOS Biology* 8.10: e1000494.**

New genes that appeared with evolution of vertebrates, especially primates, were identified using a synteny-based genome alignment from the sequenced genomes of multiple species to identify new genes.

**Zhou, Q., G. Zhang, Y. Zhang, et al. 2008. On the origin of new genes in *Drosophila*. *Genome Research* 18:1446–1455.**

This paper identified new genes based on their phylogenetic distribution in the *Drosophila* genus.



## FUNCTIONAL ANALYSES

It is often necessary to determine if a new gene is functional before further understanding of the gene's molecular role can be determined using advanced molecular and genetic analyses, which will be shown in the section on Functions and Phenotypic Effects of New Genes. Measuring the evolutionary constraint on a gene, in conjunction with expression analysis, is often a handy way to determine the functionality of new genes. This analysis is conducted by looking at the evolutionary signature of purifying selection or positive selection on the gene sequence, such as comparison of the substitution rate at synonymous and replacement sites ( $K_s$  versus  $K_a$ ), the polymorphisms at synonymous and replacement sites (e.g., single nucleotide polymorphisms), and length changes (insertion and deletion) within and outside coding regions. This approach has been used widely, for example, in Yu, et al. 2006; Bai, et al. 2007; Wang, et al. 2004 (cited under Value of Young Genes); and Rosso, et al. 2008.

**Bai, Y., C. Casola, C. Feschotte, and E. Betrán. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biology* 8.1: R11.**

This paper determined the functionality of new genes by comparing  $K_a$  and  $K_s$ , assuming that the new gene is a pseudogene and the parental gene is extremely constrained, as a conservative standard.

**Rosso, L., A. C. Marques, M. Weier, et al. 2008. Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein. *PLOS Biology* 6:e140.**

This study predicted significant functionality in new gene by identifying excess amino acid replacement changes to synonymous substitutions, suggesting positive selection on an ancestral form of a hominoid-specific retrogene (CDC14BRetro).

**Yu, H., H. Jiang, Q. Zhou, et al. 2006. Origination and evolution of a human-specific transmembrane protein gene, *c1orf37-dup*. *Human Molecular Genetics* 15:1870–1875.**

The human-specific gene *clorf37-dup* was shown to evolve rapidly, likely driven by positive selection in the coding sequences. Available online for purchase or by subscription.

## GENE ORIGINATION ON BENCHES

Taking advantage of rapid growth of bacteria, gene evolution experiments have revealed that duplicate and chimeric genes can rapidly originate and fix. These molecular novelties were shown to help microbial organisms, bacteria and yeast, to become adapted to controlled new environmental factors, as demonstrated by Blount, et al. 2012; Näsvall, et al. 2012; Riehle, et al. 2001; and Brown, et al. 1998.

**Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513–518.**

This finding reflected a long and systematic effort after more than thirty-one thousand generations of selection for the ability of *Escherichia coli* to use citrate, an ancestral trait that was estimated to exist thirteen million years ago. It was found that during experimental evolution, a citrate transporter gene was duplicated into a new chimeric gene with a new promoter, which turned the transporter from a dormant state into an active state so that the ancestral ability to eat citrate appeared. Available online for purchase or by subscription.

**Brown, C. J., K. M. Todd, and R. F. Rosenzweig. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Molecular Biology and Evolution* 15:931–942.**

This classic experiment in *Saccharomyces cerevisiae* showed that a new chimeric gene had been created in 450 generations of growth in a glucose-limited culture, with an increase in the ability to transport hexose. Available online for purchase or by subscription.

**Näsval, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338:384–387.**

A few thousand generations of continuous selection on *Salmonella enterica* led to duplicate copies accumulating mutations with diverged functions. Available online for purchase or by subscription.

**Riehle, M. M., A. F. Bennett, and A. D. Long. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 98:525–530.**

This paper showed that growth of *E. coli* for two thousand generations in a high temperature led to three duplication events that might help the strains to adapt to high temperature.

---

## Mutational Mechanisms of New Gene Formation

Mutation toward a new gene structure is the first step of new gene evolution. Various molecular processes, eleven listed below, are known to contribute to the initial structure of new genes, including the classic models of gene duplication (see Muller 1936 and Ohno 1970, cited under History) and gene recombination (in Gilbert 1978, cited under History). Although there is a rich literature about gene duplication, how gene duplication leads to neofunctionalization is rarely discussed. The following references present the known mechanistic processes leading to the evolution of new gene functions. It should be noted that in most situations, some of these mechanisms likely work together to define new genes and functions.

### DNA-BASED GENE DUPLICATION

Gene duplication is thought to contribute most to the generation of new genes. Numerous studies have detailed how new functions can evolve from duplicate genes (e.g., Li 1983; Zhen, et al. 2012; and see Conant and Wolfe 2008, cited under Selective Models and Population Genetic Predictions, and Innan and Kondrashov 2010, cited under General

Overviews, for recent reviews). Genes may be completely or partially duplicated and remain as new, intact loci or integrate into existing loci, sometimes immediately forming chimeric molecules with distinct biochemical properties, as seen in Arguello, et al. 2006 and Yang, et al. 2008. As shown in Xu, et al. 2012, even completely duplicated genes provide targets for beneficial, function-changing mutations.

**Arguello, J. R., Y. Chen, S. Yang, W. Wang, and M. Long. 2006. Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. *PLOS Genetics* 2.5: e77.**

This report provided a highly resolved observation of how a duplicate was created and further changed into a chimeric gene with new function.

**Li, W. H. 1983. Evolution of duplicate genes and pseudogenes. In *Evolution of genes and proteins*. Edited by M. Nei and R. K. Koehn, 14–37. Sunderland, MA: Sinauer.**

A comprehensive evolutionary analysis of the duplicates and pseudogenes of hemoglobins.

**Xu, G., C. Guo, H. Shan, and H. Kong. 2012. Divergence of duplicate genes in exon–intron structure. *Proceedings of the National Academy of Sciences of the United States of America* 109:1187–1192.**

The comparison of more than six hundred paralogous pairs with three hundred pairs of orthologs in plants revealed prevalent evolution of exon-intron structures by duplication.

**Yang, S., R. Arguello, X. Li, et al. 2008. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLOS Genetics* 4.1: e3.**

A correlation was detected between the flanking repetitive sequences and duplicates in analysis of a number of new genes in *Drosophila*, implicating the role of repetitive sequences in duplication.

**Zhen, Y., M. L. Aardema, E. M. Medina, M. Schumer, and P. Andolfatto. 2012. Parallel molecular evolution in an herbivore community. *Science* 337:1634–1637.**

The independent duplication of ATP $\alpha$ , along with convergent evolution of expression and parallel nucleotide substitutions, was documented in herbivorous insects. Available online for purchase or by subscription.

#### **RNA-BASED GENE DUPLICATION**

RNA-based duplication, also called “retroposition,” was predicted in the early 1990s to be able to quickly generate novel functions, as shown in Brosius 1991; Brosius 2003; and Kaessmann, et al. 2009. While DNA-based duplications with parental regulatory elements are often tandem (e.g., Thornton 2003), retroposed genes must acquire new regulatory elements or risk becoming processed pseudogenes. In addition, retroposed sequences may jump into or near existing genes and recruit existing exons, or be recruited into an existing coding sequence. In at least one case, co-retroposition of two genes immediately generated a novel chimeric gene, as was reported in Zhang, et al. 2009.

**Brosius, J. 1991. Retroposons—seeds of evolution. *Science* 251:753.**

This essay elaborated the idea that retroposition can contribute to gene evolution. Available online for purchase or by subscription.

**Brosius, J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118:99–116.**

This paper summarized the contribution of retroposition to evolutionary novelties. Available online for purchase or by subscription.

**Kaessmann, H., N. Vinckenbosch, and M. Long. 2009. RNA-based gene duplication: Mechanistic and evolutionary insights. *Nature Reviews Genetics* 10:19–31.**

This widely cited review provides an overview of the studies involving retroposition. Available online for purchase or by subscription.

**Thornton, K. 2003. Gene conversion and natural selection at duplicate loci in *Drosophila melanogaster*. PhD diss., Office of Graduate and Postdoctoral Affairs, Division of Biological Sciences, Univ. of Chicago.**

A computational analysis of the *D. melanogaster* genome characterized the frequencies of the new genes created by DNA-based duplication and RNA-based retroposition.

**Zhang, Y., S. Lu, S. Zhao, X. Zheng, M. Long, and L. Wei. 2009. Positive selection for the male functionality of a co-retroposed gene in the hominoids. *BMC Evolutionary Biology* 9:252.**

This article reported two adjacent genes were co-retroposed.

## DE NOVO GENES

After discussing the probability of gene duplication and neofunctionalization, Muller stated 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’” (Muller 1936, pp. 528–530, cited under History). However, *de novo* gene origination does occur. A number of *de novo* genes originated from noncoding regions were identified in *D. melanogaster* and its closely related species, many of which exhibit male-biased expression and very short open reading frames according to Levine, et al. 2006 (cited under History) and Begun, et al. 2007. Besides the previous references (see Wu, et al. 2011 and Knowles and McLysaght 2009, cited under Cautions for Genome Annotation), Murphy and McLysaght 2012 reports *de novo* genes in mouse; Sabath, et al. 2012 in virus; Xie, et al. 2012 in hominoid; Carvunis, et al. 2012 in yeast; Li, et al. 2010 in humans; Xiao, et al. 2009 in rice; and Cai, et al. 2008 in yeast. Other studies (see Chen, et al. 2010, cited under History, and Zhang, et al. 2011, cited under New Genes in Development, Brain, and Behaviors) have detected the important role of *de novo* genes in *Drosophila* development and human brain development.

**Begun, D. J., H. A. Lindfors, A. D. Kern, and C. D. Jones. 2007. Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta***

**clade. *Genetics* 176:1131–1137.**

A further report of *de novo* genes in species closely related to *D. melanogaster*.

**Cai, J., R. Zhao, H. Jiang, and W. Wang. 2008. *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179:487–496.**

This was the first case of *de novo* genes reported in yeast.

**Carvunis, A. R., T. Rolland, I. Wapinski, et al. 2012. Proto-genes and *de novo* gene birth. *Nature* 487:370–374.**

A mechanistic model for generating *de novo* transcripts in yeast. Available online for purchase or by subscription.

**Li, C. Y., Y. Zhang, Z. Wang, et al. 2010. A human-specific *de novo* protein-coding gene associated with human brain functions. *PLOS Computational Biology* 6:e1000734.**

This study reported a human-specific *de novo* gene.

**Murphy, D. N., and A. McLysaght. 2012. *De novo* origin of protein-coding genes in murine rodents. *PLOS ONE* 7.11: e48650.**

In mouse, seventy-six *de novo* genes were identified and confirmed to have evolved from noncoding regions.

**Sabath, N., A. Wagner, and D. Karlin. 2012. Evolution of viral proteins originated *de novo* by overprinting. *Molecular Biology and Evolution* 29.12: 3767–3780.**

This paper showed the *de novo* origination of protein in viruses. Available online for purchase or by subscription.

**Xiao, W., H. Liu, Y. Li, et al. 2009. A rice gene of *de novo* origin negatively regulates pathogen-induced defense response. *PLOS ONE* 4:e4603.**

This is the first *de novo* gene known in plants.

**Xie, C., Y. E. Zhang, J.-Y. Chen, et al. 2012. Hominoid-specific *de novo* protein-coding genes originating from long non-coding RNAs. *PLOS Genetics* 8.9: e1002942.**

This report revealed a *de novo* gene that originated in the ancestor of hominoids.

#### GENE RECOMBINATION

Gilbert 1978 (cited under History) proposed that new genes can be generated by recombining existing exons and domains, called “exon” or “domain shuffling,” which was shown to be generally important in evolution of gene structure by Long, et al. 1995. Chimeric proteins, the product of exon/domain shuffling, have been found in many organisms since their discovery in the LDL receptor gene as observed by Südhof, et al. 1985: for example, Rogers, et al. 2009 and Rogers and Hartl 2012 in *Drosophila*; Katju and Lynch 2006 in *Caenorhabditis elegans*; Marques, et al. 2005 in mammals; and Wang, et al. 2006 in plants. The first evidence for how chimeric proteins are formed came from *Jingwei*, as reported by Long and Langley 1993 (cited under History), but chimeric genes can be formed in a number of different ways.

**Katju, V., and M. Lynch. 2006. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Molecular Biology and Evolution* 23:1056–1067.**

This reported several chimeric genes in the worm. Available online for purchase or by subscription.

**Long, M. Y., C. Rosenberg, and W. Gilbert. 1995. 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* 92:12495–12499.**

A computational analysis of eukaryotic intron-containing genes detected significant signals of exon shuffling.

**Marques, A. C., I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann. 2005. Emergence of young human genes after a burst of retroposition in primates. *PLOS Biology* 3:e357.**

This article identified new chimeric genes in the human genome.

**Rogers, R. L., and D. L. Hartl. 2012. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Molecular Biology and Evolution* 29:517–529.**

A genomic screening for chimeric genes in *D. melanogaster* led to the finding of rapid evolution of genome caused by chimeric genes. Available online for purchase or by subscription.

**Rogers, R. L., T. Bedford, and D. L. Hartl. 2009. Formation and longevity of chimeric and duplicate genes in *Drosophila melanogaster*. *Genetics* 181:313–322.**

An investigation of chimeric and duplicate genes in *Drosophila* and their impact on the evolution of *Drosophila*.

**Südhof, T. C., J. L. Goldstein, M. S. Brown, and D. W. Russell. 1985. The LDL receptor gene: A mosaic of exons shared with different proteins. *Science* 228:815–822.**

The first observation of chimeric structure in the LDL receptor. Available online for purchase or by subscription.

**Wang, W., H. Zheng, C. Fan, et al. 2006. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18:1791–1802.**

The highest rate of chimeric gene origination was reported in the grass genomes.

#### **GENE FISSION/FUSION**

New genes may also be formed by splitting or joining existing genes. Wang, et al. 2004, for example, finds that gene duplication is an intermediate stage in an evolutionary process leading to gene fission. A fusion protein was observed from splicing over a run-through transcript of two adjacent genes in Thomson, et al. 2000.

**Thomson, T. M., J. J. Lozano, et al. 2000. Fusion of the human gene for the polyubiquitination coeffector UEV1 with *Kua*, a newly identified gene. *Genome Research* 10:1743–1756.**

This report described run-through transcription and splicing that results in a fusion protein.

**Wang, W., H. Yu, and M. Long. 2004. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in *Drosophila* species. *Nature Genetics* 36:523–527.**

A mechanism was found for the process of gene fission that included the intermediate duplicates and subsequent complementary degeneration of the protein terminals, which made two genes that encoded the domains in N- and C-terminals of an ancestral protein, respectively. When this process of gene evolution recently occurred, such as in the reported case, the complementary degenerated portions of the duplicates are observable in the individuals of natural populations.

#### **FRAMESHIFT MUTATIONS**

Frameshift mutations are often extremely disadvantageous mutations, but sometimes generate novel proteins. For example, Okamura, et al. 2006 finds 470 human gene duplicates that had undergone frameshifts to form novel protein sequences. Xue, et al. 2003 finds that Epstein-Barr virus contains a gene that undergoes frequent frameshifts, probably to combat host immunity.

**Okamura, K., L. Feuk, T. Marquès-Bonet, A. Navarro, and S. W. Scherer. 2006. Frequent appearance of novel protein-coding sequences by frameshift translation. *Genomics* 88:690–697.**

The first report that a large number of human genes were created by frameshift translation of previously existing protein-coding genes.

**Xue, S. A., M. D. Jones, Q. L. Lu, J. M. Middeldorp, and B. E. Griffin. 2003. 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* 3 (March): 2192–2201.**

This analysis detects some transcripts from the changed reading frame in the LF3 genes that is encoded in the Epstein-Barr virus.

#### **HORIZONTAL GENE TRANSFER**

Horizontal gene transfer (HGT) is a major mechanism for the addition of new genes to prokaryotic genomes, as reviewed by Ochman 2001 and Dagan, et al. 2008. HGT has also been observed between organellar and nuclear genomes in flowering plants by Bergthorsson, et al. 2003.

**Bergthorsson, U., K. L. Adams, B. Thomason, and J. D. Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197–201.**

This test detected several HGT events in flowering plants.

**Dagan, T., Y. Artzy-Randrup, and W. Martin. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105.29: 10039–10044.**

This analysis derived vertical and HGT gene histories from 181 sequenced prokaryotic genomes and suggested that 81 percent of their genes are involved in HGTs.

**Ochman, H. 2001. Lateral and oblique gene transfer. *Current Opinion in Genetics and Development* 11:616–619.**

A short review on the prevalence of extensive lateral gene transfer in bacteria. Available online for purchase or by subscription.

#### **TRANSPOSABLE ELEMENT DOMESTICATION**

Transposable elements (TEs) can contribute to functional divergence between duplicate genes in several ways, as was shown in Böhne, et al. 2008. TEs can mediate gene recombination by carrying coding sequences from one part of the genome to another, as reported in Jiang, et al. 2004, and can even be incorporated into existing coding sequences, as found in Makalowski, et al. 1994; Nekrutenko and Li 2001; Lorenc and Makalowski 2003; Feschotte 2008; and Iwashita, et al. 2013. In addition, Piriyaopongsa and Jordan 2008 finds that TEs are a source of micro-RNAs, major components of post-transcriptional regulation.

**Böhne, A., F. Brunet, D. Galiana-Arnoux, C. Schultheis, and J. N. Volff. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* 16:203–215.**

This article provided a comprehensive review of the roles of TEs in evolution, including their involvement in protein-coding functions and regulatory systems. Available online for purchase or by subscription.

**Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9:207–215.**



Many DNA binding proteins were generated from genes within DNA transposons that encode transposases. Available online for purchase or by subscription.

**Iwashita, S., N. Osada, T. Itoh, et al. 2013. A transposable element-mediated gene divergence that directly produces a novel type bovine Bcnt protein including the endonuclease domain of RTE-1. *Molecular Biology and Evolution* 20:1556–1563.**

An in-depth experimental analysis of the functions of a mammalian TE-domesticated nuclear gene. Available online for purchase or by subscription.

**Jiang, N., Z. Bao, X. Zhang, S. R. Eddy, and W. Wessler. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431L:569–573.**

A TE element showed a great power to generate new chimeric genes in plants. Available online for purchase or by subscription.

**Lorenc, A., and W. Makalowski. 2003. Transposable elements and vertebrate protein diversity. *Genetica* 118:183–191.**

An extensive genomic analyses of TEs domesticated into host protein-coding genes in mammals. Available online for purchase or by subscription.

**Makalowski, W., G. A. Mitchell, and D. Labuda. 1994. Alu sequences in the coding regions of mRNA: A source of protein variability. *Trends in Genetics* 10:188–193.**

This was the first description that an Alu element in humans was inserted into the coding portion of the decay-accelerating factor gene.

**Nekrutenko, A., and W. H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics* 17:619–621.**

This computational analysis of the human genomic sequences detected TE-derived sequence in approximately 4 percent of human protein-coding genes. A majority of them resided in introns and were then recruited into the protein-coding exons. Available online for purchase or by subscription.

**Piriyapongsa, J., and I. K. Jordan. 2008. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14:814–821.**

This article demonstrated an important role of TEs in creating siRNAs and miRNAs.

#### **DIVERGENCE IN ALTERNATIVE SLICING BETWEEN DUPLICATES**

Alternative splicing (AS) of individual genes can generate distinct transcripts that can produce noncoding RNAs (ncRNAs) or polypeptides with slightly or entirely different functions, so it may be no surprise that divergence in AS patterns between duplicate genes can rapidly alter duplicate gene functions and promote divergence between copies. Keren, et al. 2008 studies the mechanistic processes involved new gene evolution through AS, and Gardiner, et al. 2008 compares the importance of AS and exon exchanges relative to gene duplication in the formation of new genes. Harr and Turner 2010 estimates the proportion of genes that evolved different AS patterns in closely related mouse species. Zhang, et al.

2010 and Zhou, et al. 2011 analyze the contribution of AS to functional evolution.

**Gardiner, A., D. Barker, R. K. Butlin, W. C. Jordan, and M. G. Ritchie. 2008. Evolution of a complex locus: Exon gain, loss and divergence at the Gr39a locus in *Drosophila*. *PLOS ONE* 3:e1513.**

This is a detailed examination of exon evolution caused by exon loss and AS in a single locus Gr39a involved in olfaction in *Drosophila*, identifying a great change in AS across widely related species. They concluded that the divergence caused by AS and other exon changes are as important as gene duplication in driving functional evolution.

**Harr, B., and L. M. Turner. 2010. Genome-wide analysis of alternative splicing evolution among *Mus* subspecies. *Molecular Ecology* 19 (Suppl. 1): 228–239.**

This report detected that 873 (6.5 percent) testis-expressed genes in the experimental mouse *Mus musculus* evolved different AS patterns in its three subspecies and close relative *M. spretus*.

**Keren, H., G. Lev-Maor, and G. Ast. 2008. Alternative splicing and evolution: Diversification, exon definition and function. *Nature Reviews Genetics* 11:345–355.**

The mechanistic process by which AS has led to the evolution of new gene functions is discussed in detail. Available online for purchase or by subscription.

**Zhang, P. G., S. Z. Huang, A.-L. Pin, and K. L. Adams. 2010. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of *Arabidopsis*. *Molecular Biology and Evolution* 27:1686–1697.**

Divergence in AS among duplicate copies in *Arabidopsis thaliana* was detected, which revealed the contribution of AS divergence to functional evolution. Available online for purchase or by subscription.

**Zhou, R. C., N. Moshgabadi, and K. L. Adams. 2011. 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* 108:16122–16127.**

This experiment revealed great divergence in AS following the allopolyploidization event that formed *Brassica napus* or in resynthesized polyploids, indicating the unusually quick rate of AS evolution.

## **NCRNAS**

Not all new genes code for proteins. A large number of functional RNAs from noncoding regions have been reported to play vital roles in a wide variety of organisms in Li, et al. 2008 and Berezikov 2011. MicroRNAs appear to turn over rapidly, but can be strongly influenced by positive selection, according to the observations of Lu, et al. 2008a; Lu, et al. 2008b; and Nozawa, et al. 2010. Dai, et al. 2008 shows that a new long ncRNA influences courtship behavior in *D. melanogaster*. Heinen, et al. 2009 identifies a new ncRNA gene that originated recently and plays an important function during spermatogenesis in mouse.

**Berezikov, E. 2011. Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics* 12:846–860.**

This extensive review of microRNAs described their phylogenetic distribution and their originations in various evolution times. The mechanisms for the origination, structural features, and targeting were also discussed. Available online for purchase or by subscription.

**Dai, H., Y. Chen, S. Chen, et al. 2008. 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* 105:7478–7483.**

This report detected a new long ncRNA gene in *D. melanogaster* that is associated with a courtship phenotype.

**Heinen, T. J., F. Staubach, D. Haming, and D. Tautz. 2009. Emergence of a new gene from an intergenic region. *Current Biology* 19:1527–1531.**

This new gene in mouse originated recently as a ncRNA gene, and a knockout experiment showed that it has a postmeiotic function during spermatogenesis in males.

**Li, Z., M. Liu, L. Zhang, et al. 2008. Detection of non-coding RNAs encoded in the intergenic regions expressed in main developmental stages in *Drosophila melanogaster*. *Nucleic Acids Research* 37:4308–4314.**

This genomic experiment detected a high number of ncRNAs expressed from the noncoding regions of the *Drosophila* genomes. Available online for purchase or by subscription.

**Lu, J., Y. Fu, S. Kumar, et al. 2008a. Adaptive evolution of newly emerged micro-RNA genes in *Drosophila*. *Molecular Biology and Evolution* 25:929–938.**

By comparing sequence divergence with polymorphism in five microRNA genes in *D. melanogaster* and *simulans*, the authors revealed that adaptive evolution may be important for the evolution of these genes. Available online for purchase or by subscription.

**Lu, J., Y. Shen, Q. Wu, et al. 2008b. The birth and death of microRNA genes in *Drosophila*. *Nature Genetics* 40.3: 351–355.**

This article reported a high rate of birth and death of microRNA genes in *Drosophila* and proposed the *de novo* origination of microRNA genes. Available online for purchase or by subscription.

**Nozawa, M., S. Miura, and M. Nei. 2010. Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biology and Evolution* 2:180–189.**

This article reported frequent gain and loss of microRNA genes and suggested that old microRNA genes had important functions and detected the origin of microRNA genes through duplication. Available online for purchase or by subscription.

## NEW REGULATORY FUNCTIONS FROM PSEUDOGENES

What if a duplicate gene accumulates nonsense mutations? Except for a few cases (e.g., Duret, et al. 2006), pseudogenes are conventionally thought of as dead genes, which play no functional roles. However, Zheng and Gerstein 2007 recently found that many pseudogenes in mammals are transcribed and may thus still function. McCarrey and Riggs 1986 predicts that pseudogenes may regulate their parental genes, similar to long ncRNAs 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 the authors of Marques, et al. 2011 and Marques, et al. 2012. Thus pseudogenes may evolve into new genes that have regulatory functions in regulating expression of related genes.

**Duret, L., C. Chureau, S. Samain, J. Weissenbach, and P. Avner. 2006. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 312:1653–1655.**

This case exemplifies how a pseudogene evolved a new silencing function for dosage compensation of X-linked genes in eutherian mammals. Available online for purchase or by subscription.

**Marques, A. C., J. Tan, and C. P. Ponting. 2011. Wrangling for microRNAs provokes much crosstalk. *Genome Biology* 12:132.**

In this commentary, the authors noticed the wide cross-talk of transcripts from distinct genes mediated by microRNAs, a critical regulatory player, and started to speculate on the related functional role of pseudogenes.

**Marques, A. C., J. Tan, S. Lee, L. Kong, A. Heger, and C. P. Ponting. 2012. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biology* 13:R102.**

The authors presented a mechanistic model for the hypothesized regulatory function of pseudogenes (as decoys to compete for the microRNA binding) and tested it with both computational and experimental evidence.

**McCarrey, J. R., and A. D. Riggs. 1986. 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* 83:679–683.**

An early interesting study to examine the function of pseudogenes.

**Zheng, D., and M. B. Gerstein. 2007. The ambiguous boundary between genes and pseudogenes: The dead rise up, or do they?. *Trends in Genetics* 23:219–224.**

Given that a big proportion of pseudogenes (5–20 percent in humans and 50 percent in mouse) were found to transcribe RNAs, the authors suspected that these pseudogenes might keep residual transcriptional activity, but might also evolve noncoding functions. Available online for purchase or by subscription.

## Evolutionary Forces Acting on New Genes

Evolutionary forces, such as natural selection and neutrality, can be operational on two levels in new gene evolution. The first is the forces that fix the new gene that first appears in a single individual genome. Meanwhile, the evolutionary forces are also working on the new sequence changes in the newly created gene, before or after its fixation in a species. In general, the role of natural selection is tested using population genetic and substitution analyses, as the publications in the first two sections, Selective Models and Population Genetic Predictions and Empirical Data, show; functional and structural analyses have also provided insight into the role of adaptive evolution, as exemplified by publications cited under Analysis of New Gene Structure and Function.

### SELECTIVE MODELS AND POPULATION GENETIC PREDICTIONS

In addition to the classic models of duplication presented by Muller 1936 and Ohno 1970 (cited under History) that selected for new mutations—the new copies and mutations in a duplicate copy with relaxed constraint may bring up new functions or be silenced to a pseudogene or subfunctionalization (as proposed by Force, et al. 1999, which was the first model that predicted various consequences of duplication)—there are a few other models for the evolution of new functions. These models, including (1) the Escape-from-Adaptive Conflict Model (EAC) by, for example, Deng, et al. 2010; (2) IAD by Näsvall, et al. 2012 and Bergthorsson, et al. 2007; (3) adaptive radiation (AR) by Francino 2005, invoking ancestral secondary functions or potentials, provided expectations to test; additionally proposed were (4) the Coalescent Model by Thornton 2007, and (5) the population genetic models by Walsh 2003 and Walsh 1995. All five types of models, mechanistically or statistically, all duplication based, are explained below. Conant and Wolfe 2008 reviews experimental evidence for the ancestral multiple functions that were amplified by duplication.

**Bergthorsson, U., D. I. Andersson, and J. R. Roth. 2007. Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America* 104:17004–17009.**

The authors proposed a selective amplification model that incorporated (1) the innovation before duplication that created minor functions, (2) the amplification of the minor functions by duplication, and (3) the divergence of the parental and new gene copies by mutation and recombination.

**Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* 9:938–950.**

This review highlighted many cases in which the secondary functions present in the ancestral gene were amplified by duplication. Available online for purchase or by subscription.

**Deng, C., C.-H. C. Cheng, H. Yea, X. He, and L. Chen. 2010. 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* 107:21593–21598.**

The authors tested the EAC, which assumes that there is adaptive conflict between the old and an emerging new function in a gene. They found that the deletion of the domain responsible for the old function was deleted.

**Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics***

**151.4: 1531–1545.**

This paper proposed a widely cited model for gene duplication, the duplication-degeneration-complementation model for subfunctionalization. It also briefly discussed the other two possible consequences of duplication: pseudogenization and neofunctionalization.

**Francino, M. P. 2005. An adaptive radiation model for the origin of new gene functions. *Nature Genetics* 37:573–577.**

This paper proposed an AR model for the origin of new gene functions: competition among duplicate copies that are amplified from a single gene leads to an adaptive radiation of new functions, predicting many pseudogenes. Available online for purchase or by subscription.

**Nåsvall, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338:384–387.**

This paper reported a few thousand generations of continuous selection on *Salmonella enterica* leading to duplicate copies accumulating mutations with diverged catalytic functions, providing evidence for the IAD model. Available online for purchase or by subscription.

**Thornton, K. R. 2007. The neutral coalescent process for recent gene duplications and copy-number variants. *Genetics* 177:987–1000.**

The theoretical predictions for the diversity in the new gene that were recently fixed or are still segregating were made by examining the coalescent process under neutrality for the distribution of the gene duplicates in a population.

**Walsh, B. 2003. Population-genetic models of the fates of duplicate genes. *Genetica* 118:279–294.**

An extended analysis of the author's 1995 model to predict the fates of duplicate genes. Available online for purchase or by subscription.

**Walsh, J. B. 1995. How often do duplicated genes evolve new functions?. *Genetics* 139:421–428.**

The first population genetic model to examine the relative probabilities of silencing versus neofunctionalization, under the classic model of duplication (e.g., Muller 1936 and Ohno 1970, cited under History).

**EMPIRICAL DATA**

Molecular evolution and population genetic theory provide predictions for testing natural selection acting on new genes. The following sections are the articles that reported statistical tests of the relevant hypotheses and predictions using empirical data.

## Fixation of New Genes within Populations

The first stage of new gene origination is the formation of a protogene in an individual, which then becomes polymorphic within the population before its fixation in the species. The following papers provided evidence that natural selection may have played an important role in the fixation of new gene duplicates within species. Using neutrality as a null hypothesis, Schrider, et al. 2013 and Schrider, et al. 2011 detect the signatures of natural selection on retroposed genes using population genomic data in humans and *Drosophila*. Llopart, et al. 2002 and Emerson, et al. 2008 directly investigate the role of natural selection using polymorphism data for new exon-intron structure and copy number variation.

**Emerson, J. J., M. Cardoso-Moreira, J. O. Borevitz, and M. Long. 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* 320:1629–1632.**

This genomic study to investigate the main evolutionary forces acting on polymorphic gene duplicates and deletions detected a role for natural selection in maintaining standing copy number variation. Available online for purchase or by subscription.

**Llopart, A., J. M. Comeron, F. G. Brunet, D. Lachaise, and M. Long. 2002. Intron presence/absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proceedings of the National Academy of Sciences of the United States of America* 99:8121–8126.**

The first evidence that polymorphism in the new exon-intron structure created by the loss of an intron is driven by positive selection in *Drosophila*.

**Schrider, D. R., F. C. P. Navarro, P. A. F. Galante, et al. 2013. Gene copy-number polymorphism caused by retrotransposition in humans. *PLOS Genetics* 9.1: e1003242.**

The application of the McDonald-Kreitman test in molecular population genetics detected an excess fixation of retrogenes, indicating a significant role of positive selection in the origination of retrogenes in the human genome.

**Schrider, D. R., K. Stevens, C. M. Cardeño, C. H. Langley, and M. W. Hahn. 2011. Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*. *Genome Research* 21:2087–2095.**

The population genetic analysis based on the McDonald-Kreitman test revealed that positive selection is responsible for the excess fixation of retrogenes in the *Drosophila* genome.

## Selection on Sequence Changes in New Genes

New genes accumulate mutations before and after the fixation of the new gene. There are numerous publications that report evidence for positive selection acting on the sequence variation of new genes. Several papers below exemplify these studies and are chosen for their generality, originality, and valuable methodology, including the methods of molecular population genetics, as were used by Long and Langley 1993; Nurminsky, et al. 1998; Jones and Begun 2005; Shih and Jones 2008 (cited under Value of Young Genes); and Chen, et al. 2010.

**Chen, S., Y. Zhang, and M. Long. 2010. New genes in *Drosophila* quickly become essential. *Science* 330:1682–1685.**

This paper analyzed many new genes of various ages in *D. melanogaster* and observed that younger genes were driven by stronger selective forces. Available online for purchase or by subscription.

**Jones, C. D., and D. J. Begun. 2005. Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences of the United States of America* 102:11373–11378.**

This paper, along with Long and Langley 1993, analyzed the substitution patterns in four chimeric genes that were independently derived from a duplicate (either RNA based or DNA based) of alcohol dehydrogenase and found that these genes share common parallel evolution in their amino acid sequence changes under positive selection.

**Long, M., and C. H. Langley. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95.**

This early paper on a new gene tested the null hypothesis of neutrality by comparing within-species variation and between-species divergence using the McDonald-Kreitman method, detecting significant positive selection acting on the *Jingwei* gene. Available online for purchase or by subscription.

**Nurminsky, D. I., M. V. Nurminskaya, D. De Aguiar, and D. L. Hartl. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* 396:572–575.**

This paper reported a species-specific chimeric gene, *Sdic*, in the *D. melanogaster* lineage that was the target of a selective sweep. Available online for purchase or by subscription.

### Analysis of New Gene Structure and Function

In addition to the population genetic analyses of new genes, the complementary structure-function approach has also been used to study new genes. The subjects of the analysis were often functions that are apparently adapted to particular environmental conditions or fit to the mechanistic process under positive selection. One widely cited example is the analysis of the protein structure and antifreeze function in Chen, et al. 1997 and Cheng and Chen 1999. Other studies have revealed that selection is sometimes jointly acting with the mechanistic processes that lead to new genes. For example, Díaz-Castillo and Ranz 2012 demonstrates that the position of chromosomal domains in the nuclear structure has an effect on the position of retrogenes in the genome. Similarly Hense, et al. 2007 and Vibranovski, et al. 2009 suggests that male meiotic sex chromosome inactivation might play a role in the trafficking of new genes from sex chromosomes to autosome (see Patterns of New Gene Movement).

**Chen, L., A. L. De Vries, and C.-H. C. Cheng. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proceedings of the National Academy of Sciences of the United States of America* 94:3817–3822.**

This paper, with Cheng and Chen 1999, is an example of how a new protein can be independently formed in response to similar selective pressures to help some Antarctic and Arctic fishes to survive in freezing seawater.

**Cheng, C.-H., and L. Chen. 1999. Evolution of an antifreeze glycoprotein. *Nature* 401:443–444.**



This is an analysis of the structural evolution of antifreeze proteins in polar fishes. Available online for purchase or by subscription.

**Díaz-Castillo, C., and J. M. Ranz. 2012. Nuclear chromosome dynamics in the *Drosophila* male germ line contribute to the nonrandom genomic distribution of retrogenes. *Molecular Biology and Evolution* 29:2105–2108.**

This analysis revealed the impact of the mechanistic process of new gene formation on retrogenes, whose locations were shown to be related to the position of chromosome domains with the nuclear periphery. Available online for purchase or by subscription.

**Hense, W., J. F. Baines, and J. Parsch. 2007. X chromosome inactivation during *Drosophila* spermatogenesis. *PLOS Biology* 5:e273.**

This paper provided the first experimental evidence that the X-linked genes in *Drosophila* are likely inactivated in male germlines. They further suggested that this meiotic sex chromosome inactivation might be one of the selective forces for the trafficking of new genes to autosomes.

**Vibrantovski, M. D., H. F. Lopes, T. L. Karr, and M. Long. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLOS Genetics* 5.11: e1000731.**

This paper presented the first genome-wide analysis in the *Drosophila* gene expression in the spermatogenesis process of male germline cells, and revealed several independent evidences consistent with the prediction of meiotic sex chromosome inactivation.

---

## Rates of New Gene Origination

Advances in genome sequencing and experimental genomics techniques have made finding new genes feasible. However, accurate estimation of the rate of new gene origination depends upon the genome's annotation, which is unfortunately biased against new genes (Zhang, et al. 2012, cited under Cautions for Genome Annotation). Hence, researchers should be alert to the quality of the annotation of various databases. Nevertheless, careful examination of the genome sequences from more closely related species in mammals, *Drosophila*, and grasses have identified recently evolved genes and provide insightful estimates of the origination rates of new genes.

### CAUTIONS FOR GENOME ANNOTATION

It has been found that many annotation methods, as outlined in Flicek, et al. 2012, are good at annotating old genes but neglect young genes. Zhang, et al. 2012 finds and discusses the systematic bias in genome annotation and functional studies and proposes possible approaches to alleviate this problem. Knowles and McLysaght 2009 and Wu, et al. 2011 report *de novo* genes that are missing in the current Ensembl annotation but exist in its earlier versions.

**Flicek, P., M. R. Amode, D. Barrell, et al. 2012. Ensembl 2012. *Nucleic Acids Research* 40:D84–D90.**

An introduction of the widely used database of annotated genes. Available online for purchase or by subscription.

**Knowles, D. G., and A. McLysaght. 2009. Recent *de novo* origin of human protein-coding genes. *Genome Research* 19:1752–1759.**

This important discovery of *de novo* genes in the human genome that were missing from the annotation in Ensembl.

**Wu, D. D., D. M. Irwin, and Y. P. Zhang. 2011. De novo origin of human protein-coding genes. *PLOS Genetics* 7:e1002379.**

This work carefully checked several versions of Ensembl and identified many *de novo* genes. Many of these *de novo* genes are not present in the current annotation of Ensembl.

**Zhang, Y. E., P. Landback, M. Vibranovski, and M. Long. 2012. New genes expressed in human brains: Implications for annotating evolving genomes. *BioEssays* 34:982–991.**

This review reported systematic biases against new genes in annotation programs, for example, those embedded in Ensembl, and a correcting approach was proposed and used. Available online for purchase or by subscription.

## **DROSOPHILA**

In this group of insects, several systematic analyses have been conducted from the early efforts of Bai, et al. 2007; Zhou, et al. 2008; and Yang, et al. 2008 to a recent description of changed rates in different stages of evolution. Zhang, et al. 2010 reports a large number of new genes that originated in the lineage toward *D. melanogaster*.

**Bai, Y., C. Casola, C. Feschotte, and E. Betrán. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biology* 8.1: R11.**

This analysis was interested in the new genes created by retroposition.

**Yang, S., R. Arguello, X. Li, et al. 2008. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLOS Genetics* 4.1: e3.**

This experimental analysis identified new genes by hybridizing cDNA against the polytene chromosomes of various species in the *melanogaster* species group.

**Zhang, Y., M. D. Vibranovski, and M. Long. 2010. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Research* 20:1526–1533.**

Using the genomic sequences of twelve species of *Drosophila*, the authors provided a conservative estimate that fifteen new genes have arisen per million years by the duplication, retroposition, and *de novo* mechanisms.

**Zhou, Q., G. Zhang, Y. Zhang, et al. 2008. On the origin of new genes in *Drosophila*. *Genome Research* 18:1446–1455.**

This analysis computationally identified new genes that have arisen by DNA-based duplication, retroposition, and *de novo* origination, using genomic sequences from the *D. melanogaster* subgroup. Available online for purchase or by subscription.

## MAMMALS

The early efforts to identify new genes in mammals began by investigating the rate of the origination of retrogenes. More recently, DNA-based duplications, retroposition, and *de novo* origination have been considered in Zhang, et al. 2010; Vinckenbosch, et al. 2006; and Marques, et al. 2005.

**Marques, A. C., I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann. 2005. Emergence of young human genes after a burst of retroposition in primates. *PLOS Biology* 3.11: e357.**

This study, with the one above from the same research group, characterized the distribution of retrogenes in various evolutionary times in primates.

**Vinckenbosch, N., I. Dupanloup, and H. Kaessmann. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 103:3220–3225.**

A careful analysis of the new genes created by retroposition.

**Zhang, Y., M. D. Vibranovski, and M. Long. 2010. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Research* 20:1526–1533.**

Comparison of sequenced genomes of mammalian species led to estimation of new genes that originated in different stages of evolution toward humans. In general, this is an accelerating origination process, for example, fifty genes per million years in primates, eighty genes per million years in hominoids, and approximately one hundred genes per million years in human-specific lineage.

## PLANTS

The computational analysis of rice genomes from revealed unexpectedly high rates of origination of retroposition-derived chimeric genes, fifty chimeric genes per million years, the highest record in all known organisms, reported in Wang, et al. 2006 and confirmed in Fan, et al. 2008.

**Fan, C., Y. Zhang, Y. S. Yu, S. Rounsley, M. Long, and R.A. Wing. 2008. The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. *Molecular Plant* 1:839–850.**

This paper revealed a high origination rate of new genes that were created through DNA-based duplication, in addition to what was found in the new gene origination by the retroposition mechanism. Available online for purchase or by subscription.

**Wang, W., H. Zheng, C. Fan, et al. 2006. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18:1791–1802.**

This article revealed the high rate of origination rate of retroposition-derived chimeric genes.

---

## Evolutionary Patterns Associated with New Genes

New gene origination is governed by various evolutionary forces, as evidenced by the finding that there are patterns associated with the evolution of new genes. A few patterns have been detected, including gene trafficking in mammals and insects, distinct expression patterns between the sexes, and sequence substitution patterns. These patterns, which are described below, have provided new insights into the mechanisms that lead to the origin of new genes and the underlying forces that lead to their fixation.

### THE PATTERNS OF NEW GENE MOVEMENT

It was found that in evolution, new genes, for example retrogenes, have been directionally moved between sex chromosomes (X and Z) and autosomes, which is called “gene trafficking.” This finding eventually led to a new view that there has been a significant interaction between the sex chromosomes and autosomes in the distribution of gene contents in the genome during evolution. Since the directional movement of retrogenes from the X chromosome to autosomes was reported in *Drosophila* by Betrán, et al. 2002 (cited under History), there has been extensive interest in investigating related patterns, and the underlying reasons and consequences. Gene trafficking on the Z chromosome of silkworm was analyzed by Wang, et al. 2012; gene trafficking similar to that in *Drosophila* was found on the X chromosome of humans and mouse by Emerson, et al. 2004. Vibranovski, et al. 2009 and Meisel, et al. 2009 extend these studies of retrogenes and investigated the gene trafficking of DNA-based duplicates. Finally, the consequence of gene trafficking on the distribution of sex-biased genes in the genome was found in Ranz, et al. 2003 and Parisi, et al. 2003 in *Drosophila* and in Khil, et al. 2004 in mouse. Khil, et al. 2005 reviews these data.

**Emerson, J. J., H. Kaessmann, E. Betrán, and M. Y. Long. 2004. Extensive gene traffic on the mammalian X chromosome. *Science* 303:37–540.**

This article revealed the impact of natural selection on the retroposition events that shaped gene trafficking between the X and autosomes in mammals, showing significant difference from the chromosomal distribution of the neutrally evolving processed pseudogenes. Available online for purchase or by subscription.

**Khil, P. P., B. Oliver, and R. D. Camerini-Otero. 2005. X for intersection: Retrotransposition both on and off the X chromosome is more frequent. *Trends in Genetics* 21:3–7.**

This article, with Khil, et al. 2004; Ranz, et al. 2003; and Parisi, et al. 2003, revealed that the gene traffic might be a mechanism involved in the enrichment of male-biased genes on autosomes in mammals and *Drosophila*.

**Khil, P. P., N. A. Smirnova, P. J. Romanienko, and R. D. Camerini-Otero. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature Genetics* 36:642–646.**

This paper revealed that male-biased genes that were expressed during the meiotic sex chromosome inactivation were not X linked in mouse, suggesting that meiotic sex chromosome inactivation might play a role in the trafficking of new genes from the X chromosome to the autosomes.

**Meisel, R. P., M. V. Han, and M. W. Hahn. 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biology and Evolution* 1:176–188.**

This analysis provided an independent set of data in *Drosophila* in support of gene trafficking out of the X, for new genes created by both RNA-based duplication and DNA-based duplication. Available online for purchase or by subscription.

**Parisi, M., R. Nuttall, D. Naiman, et al. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299:697–700.**

The excess number of genome-wide identified male-biased genes in *D. melanogaster* was found on autosomes, indicating that the X is a undesirable chromosome for the male-biased genes. Available online for purchase or by subscription.

**Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300:1742–1745.**

The comparative analyses of transcriptomes in *D. melanogaster* and *D. simulans* showed that the genes with diverged sex-biased expression are nonrandomly distributed between the X and autosomes, revealing the role of sex-related selection in rapidly evolved genes. Available online for purchase or by subscription.

**Vibrantovski, M. D., Y. Zhang, and M. Long. 2009. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Research* 19:897–903.**

This analysis showed that new genes in *Drosophila* created by both DNA-based and RNA-based duplication tended to move out of the X chromosomes.

**Wang, J., M. Long, and M. D. Vibrantovski. 2012. Retrogenes moved out of the Z chromosome in the silkworm. *Journal of Molecular Evolution* 74:113–126.**

This paper reported a retrogene movement out of the Z in the silkworm. Available online for purchase or by subscription.

#### THE PATTERNS OF NEW GENE EXPRESSION

Retrogenes were found to be more likely expressed in testis in humans, mouse, and *Drosophila*; also see Betrán, et al. 2002 (cited under History) for the first observation that the dominant majority of retrogenes were expressed in testis. The expression pattern of retrogenes was found to be complementary with the parental genes in *Drosophila* by Vibranovski, et al. 2009 and in mammals by Potrzebowski, et al. 2008. Ding, et al. 2010 reports the expression of a new gene in spermatogenesis, which resulted in a striking male phenotype. Based on these data, Vinckenbosch, et al. 2006 proposes the “out of testis” hypothesis.

**Ding, Y., L. Zhao, S. Yang, et al. 2010. A young *Drosophila* duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. *PLoS Genetics* 6.12: e1001255.**

This article reported a new gene that evolved a critical function in spermatogenesis with a striking male phenotype.

**Potrzebowski, L., N. Vinckenbosch, A. C. Marques, F. Chalmel, B. Jegou, and H. Kaessmann. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biology* 6:e80.**

This paper reported the complementary expression patterns of retrogenes and parental genes in the somatic and germline tissues of mammals.

**Vibranovski, M. D., H. F. Lopes, T. L. Karr, and M. Long. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genetics* 5.11: e1000731.**

This study detected the complementary expression patterns between the parental genes and retrogenes across different spermatogenesis stages in *Drosophila*.

**Vinckenbosch, N., I. Dupanloup, and H. Kaessmann. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 103:3220–3225.**

This study detected that the retrogenes initially evolved testis-expression patterns and later became more diverse and stronger in expression, coined the “out of the testis” hypothesis.

#### THE PATTERNS OF SEQUENCE EVOLUTION IN NEW GENES

After new genes are fixed in a species, they can continuously evolve toward further improved function. Chen, et al. 2010 analyzes the intensity of natural selection across new genes with different. Jones and Begun 2005 finds that there was rapid evolution driven by selection of the alcohol dehydrogenase (Adh)-derived new genes in the early stage of their evolution.

**Chen, S., Y. Zhang, and M. Long. 2010. New genes in *Drosophila* quickly become essential. *Science* 330:1682–1685.**

The population genetic analysis of new genes revealed the early strong positive selection with the decreasing selective intensity until the strong purifying selection was established

with the increase of the ages of genes. Available online for purchase or by subscription.

**Jones, C. D., and D. J. Begun. 2005. Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences of the United States of America* 102:11373–11378.**

This paper dissected the evolutionary dynamics in a few chimeric genes that were independently derived from DNA-based and RNA-based duplication of Adh genes, detecting early adaptive evolution and pinpointing parallel amino acid substitutions.

---

## Functions and Phenotypic Effects of New Genes

There have been extensive investigations into the functions of new genes and their roles in phenotypic evolution using molecular biological, biochemical, and genetic means. It has been found that the genetic machinery in control of a few phenotypes and molecular functions have evolved rapidly by recruiting new genetic components. This demands further understanding of the underlying genetic and evolutionary mechanisms. Chen, et al. 2013 reviews the progress in this area, and the following sections outline a few areas of major progress that were chosen from this general review. New genes have also found to be critical in the evolution of sexual reproduction and underlying genetic systems. This topic has been extensively discussed in the literature cited under Patterns of New Gene Movement and Patterns of New Gene Expression.

**Chen, S. D., B. H. Krinsky, and M. Long. 2013. New genes as drivers of phenotypic evolution. *Nature Reviews Genetics* 14:645–660.**

In this review, recent progress in the study of the functions and phenotypic effects of new genes are discussed. It was proposed that, as the title indicates, new genes are drivers of phenotypic evolution. Available online for purchase or by subscription.

## NEW GENES IN DEVELOPMENT, BRAIN, AND BEHAVIORS

Genetic control of development has been observed to be conservative. However, new genes are observed to evolve important functions in development, suggesting that development genetics, although its conserved components have been identified, were as a whole evolved so quickly that species and closely related lineages have specific genetic components to define their development. Schmidt-Ott, et al. 2010 reviews the new duplicates in the Hox gene family that were integrated into development in *Drosophila*; Chen, et al. 2010 (cited under History) provides the first evidence that a high proportion of young genes evolved essential developmental functions. New genes also have a role in the evolution of brains and behavior. Using the reverse genetic approach and computational analysis of transcriptome data, the authors of Chen, et al. 2012 and Dai, et al. 2008 investigate the brain and behavioral effects of new genes in *Drosophila*; Zhang, et al. 2011 finds the young genes were expressed in the human brain in the early development.

**Chen, S. D., M. Spletter, X. C. Ni, K. White, L. Q. Luo, and M. Long. 2012. Frequent recent origination of brain genes shaped the evolution of foraging behavior in *Drosophila*. *Cell Reports* 1:118–132.**

Around 160 new genes were found to be recruited into brain functions, as evidenced by expression in specific brain regions in *Drosophila*. Foraging behaviors were found to be controlled by two young genes, likely through the control of brain neurons in which the new gene *Xcbp1* is expressed.

**Dai, H., Y. Chen, S. Chen, et al. 2008. 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* 105:7478–7483.**

A gene mutagenesis analysis showed that the new gene, *sphinx*, might be involved in control of courtship behavior.

**Schmidt-Ott, U., A. M. Rafiqi, and S. Lemke. 2010. *Hox3/zen* and the evolution of extraembryonic epithelia in insects. In *Hox genes: Studies from the 20th to the 21st century*. Edited by Jean S. Deutsch, 133–144. New York: Springer.**

This review discussed the role of the multiple duplicates of the *Hox3/zen* gene family in the evolutionary changes in extraembryonic development in insects.

**Zhang, Y. E., P. Landback, M. D. Vibranovski, and M. Long. 2011. Accelerated recruitment of new brain development genes into the human genome. *PLOS Biology* 9:e1001179.**

This computational analysis of human genomic sequences and tissue-expression profiles detected excess hominoid-specific and human-specific genes that are upregulated in the developing brain of humans, especially in the prefrontal cortex regions and other neocortex regions.

## **NEW GENES RESHAPE GENE NETWORKS**

Gene networks are viewed as conserved in evolution as a consequence of robustness. However, Chen, et al. 2012 and Ross, et al. 2013 show that new genes are able to integrate into an ancestral gene network and reshape it extensively. In plants, Weng, et al. 2012 and Matsuno, et al. 2009 show that new genes could assemble new biochemical pathways. Li, et al. 2010 shows a new gene in yeast suppressed a mating pathway. Recently, Capra, et al. 2010 analyzes the integration into gene networks of new genes generated through different mechanisms in *Saccharomyces cerevisiae*.

**Capra, J. A., K. S. Pollard, and M. Singh. 2010. Novel genes exhibit distinct patterns of function acquisition and network integration. *Genome Biology* 11:R127.**

This work showed that new genes in yeast that were integrated into the networks of protein-protein interaction have different positions in networks based on their mechanism of origination.



**Chen, S. D., X. C. Ni, B. H. Krinsky, et al. 2012. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *EMBO Journal* 31.12: 2798–2809.**

This paper revealed that the new gene called “Zous” has rewired the gene networks involving hundreds of genes in male reproduction under positive selection.

**Li, D., Y. Dong, Y. Jiang, H. F. Jiang, J. Cai, and W. Wang. 2010. A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. *Cell Research* 20:408–420.**

This pathway analysis revealed a *de novo* gene can suppress a mating pathway and promote vegetative growth in yeast.

**Matsuno, M., V. Compagnon, G. A. Schoch, et al. 2009. Evolution of a novel phenolic pathway for pollen development. *Science* 325:1688–1692.**

This article reported that the phenolic pathway for pollen development appears as a consequence of the origin of a new gene. Available online for purchase or by subscription.

**Ross, B. D., L. Rosin, A. W. Thoma, et al. 2013. Stepwise evolution of essential centromere function in a *Drosophila* neogene. *Science* 340:1211–1214.**

This report revealed a molecular evolutionary process that drove evolution of species-centromere functions, by changing the expression network associated with a young gene in *Drosophila*. Available online for purchase or by subscription.

**Weng, J.-K., Y. Li, H. P. Mo, and C. Chapple. 2012. Assembly of an evolutionarily new pathway for  $\alpha$ -pyrone biosynthesis in *Arabidopsis*. *Science* 337:960–964.**

This report showed that origin of a single duplicate gene had triggered the evolution of a new biochemical pathway. Available online for purchase or by subscription.

LAST MODIFIED: 01/13/2014

DOI: 10.1093/OBO/9780199941728-0012

[BACK TO TOP](#)

You are browsing courtesy of: **Oxford Online OUP–USA**

Copyright © 2015. All rights reserved.