Gene Duplication

Jianzhi Zhang

Introduction

Gene duplication refers to the duplication of a segment of DNA that contains one or more genes. Gene duplication is the primary source of new genes in evolution, and duplicate genes form gene families that are abundantly found in almost all genomes. For these reasons, gene duplication has been a main focus of molecular evolutionary study for decades. Recent years have also seen examples of harmful gene duplication that causes diseases, which bring gene duplication to the attention of human and medical geneticists. Our knowledge about gene duplication has increased substantially in the last decade, due in a large part to the rapid accumulation of gene sequence and functional data at the genomic scale. We now have a basic understanding of gene duplication, including how it occurs, how often it occurs, and the common routes of duplicate gene evolution. A number of key questions remain unresolved, however, concerning issues such as the primary force behind the fixation of new gene duplicates and the relative probabilities of various modes of functional changes in duplicate genes. Mathematical modeling, large-scale genomic analysis, detailed functional study of individual genes, and laboratory experimental evolution of microbes are being used to tackle these remaining questions.

General Overviews

There are a few books on gene duplication. Ohno 1970 is a classic book that forcefully argued for the importance of gene duplication in evolution and substantially raised evolutionary biologists' interest in the topic. Ohta 1980 is a monograph that included the author's theoretical population genetic analysis of gene duplication and gene family evolution. Two edited volumes on gene and genome duplication were published in the early 21st century: Meyer and van de Peer 2003 and Dittmar and Liberles 2010. In addition, a number of reviews discuss the progress in various aspects of the study of gene duplication. For instance, Wolfe 2001 reviews the process of genome duplication and its aftermath, Conant and Wolfe 2008 reviews functional changes after gene duplication, Innan and Kondrashov 2010 discusses various evolutionary models of duplicate genes, and Zhang 2013 provides a comprehensive overview of the mechanisms of gene duplication and the evolution of duplicate genes.

Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* 9.12: 938–950.

A review on the types and mechanisms of functional changes after gene duplication.

Dittmar, K., and D. A. Liberles. 2010. *Evolution after Gene Duplication*. Hoboken, NJ: Wiley-Blackwell.

The most recent book on gene duplication, this edited volume examines post-duplication gene evolution and focuses on the mechanisms governing the retention and evolution of duplicate genes.

Innan, H., and F. Kondrashov. 2010. The evolution of gene duplications: Classifying and distinguishing between models. *Nature Reviews Genetics* 11.2: 97–108.

A review on the existing models describing the fixation and evolution of duplicate genes.

Meyer, A., and Y. van de Peer, eds. 2003. *Genome evolution: Gene and genome duplications and the origin of novel gene functions*. Dordrecht, The Netherlands, and Boston: Kluwer Academic.

The first comprehensive book on gene duplication since Ohno 1970, this edited volume examines the molecular and genomic evidence for the prevalence and evolutionary contribution of gene duplication.

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

The first and arguably the most important book on gene duplication, it proposes a central role of gene duplication in the origin of new genes and new gene functions in evolution.

Ohta, T. 1980. *Evolution and variation of multigene families*. Berlin and New York: Springer.

Collection of Ohta's early theoretical studies of gene family evolution.

Wolfe, K. H. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* 2.5: 333–341.

Discussing the occurrence and aftermath of genome duplication, this is the first major review of the topic in the genomic era.

Zhang, J. 2013. Gene duplication. In *The Princeton guide to evolution*. Edited by J. Losos. Princeton, NJ: Princeton Univ. Press.

This review article provides an overview of our current understanding of evolution by gene duplication.

Journals

Studies of gene duplication typically appear in molecular evolution journals and genetics/genomics journals, including Molecular Biology and Evolution, Journal of Molecular Evolution, Genetics, Genome Research, PLOS Genetics, Trends in Ecology and Evolution, Nature Reviews Genetics, and Trends in Genetics.

Genetics. 1916–.

Publishes original research on inheritance and is the major publication of the Genetics Society of America. Several important theoretical population genetic papers on duplicate gene evolution appeared in *Genetics*.

Genome Research. 1991–.

Publishes original research on genome structure, function, and evolution. Some large-scale analyses of gene families and gene duplication patterns appeared in *Genome Research*.

Journal of Molecular Evolution. 1971–.

Dated back to 1971, it used to be the dominant journal in molecular evolution.

Molecular Biology and Evolution. 1983–.

One of the major journals of the Society for Molecular Biology and Evolution; publishes research at the interface of molecular and evolutionary biology. The journal has published many studies on gene family evolution as well as patterns and mechanisms of duplicate gene evolution.

Nature Reviews Genetics. 2000–.

A top review journal of genetics, it has published several influential reviews on duplicate gene evolution.

PLOS Genetics. 2005–.

An open-access journal published by Public Library of Science (PLOS; formerly PLoS), it covers all areas of genetics. Some recent studies of gene duplication appeared in *PLOS Genetics*.

Trends in Ecology and Evolution.

One of the most important review journals in evolutionary biology, it publishes reviews, opinions, and letters in all areas of ecology and evolution. It has published influential reviews and opinions on gene duplication.

Trends in Genetics.

A major review journal of genetics, it contains reviews and opinions in all areas of genetics. The now extinct Genome Analysis session in the journal published many original discoveries on gene duplication.

History

Bridges 1936 reported the first case of gene duplication from the observation of the doubling of a small segment of the X chromosome in mutant fruit flies exhibiting extreme reduction in eye size. Gene duplication was quickly recognized by several evolutionary biologists to be a potential source of new genes in evolution, as seen in Stephens 1951 and Nei 1969. The classic book Ohno 1970 popularized the idea that gene duplication plays a major role in evolution. The study of gene duplication has thrived since the late 1990s, due in a large part to the rapid accumulation of genome sequence and functional data. Notable examples include the identification of whole-genome duplication events in fungal, plant, and animal evolution, such as Wolfe and Shields 1997 on Baker's yeast, Vision, et al. 2000 on *Arabidopsis*, and Dehal and Boore 2005 on vertebrates. The recently edited book Dittmar and Liberles 2010 covers a wide range of topics within the current study of gene duplication.

Bridges, C. B. 1936. The Bar "gene" a duplication. *Science* 83.2148: 210–211.

The first report of gene duplication.

Dehal, P., and J. L. Boore. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* 3:e314.

Provides strong genomic evidence for Ohno's hypothesis of two rounds of genome duplication in early vertebrate evolution.

Dittmar, K., and D. A. Liberles. 2010. *Evolution after gene duplication*. Hoboken, NJ: Wiley-Blackwell.

This is the most recent comprehensive book on gene duplication.

Nei, M. 1969. Gene duplication and nucleotide substitution in evolution. *Nature* 221:40–42.

An early study of the role of gene duplication in genome evolution.

Ohno, S. 1970. *Evolution by gene duplication*. Berlin and New York: Springer.

The first book on gene duplication, in which Ohno proposed the hypothesis of two rounds of genome duplication at an early stage of vertebrate evolution and the neofunctionalization model.

Stephens, S. G. 1951. Possible significances of duplication in evolution. *Advances in Genetics* 4:247–265.

An early but comprehensive evolutionary treatment of the topic of gene duplication.

Vision, T. J., D. G. Brown, and S. D. Tanksley. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290.5499: 2114–2117.

Genomic evidence for multiple rounds of genome duplication during *Arabidopsis* evolution.

Wolfe, K. H., and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387.6634: 708–713.

The first molecular evidence for genome duplication.

Mechanisms of Gene Duplication

Gene duplication typically occurs by one of three mutational mechanisms: unequal crossing-over, retroposition, and chromosomal (or genome) duplication. Zhang 2003 summarizes the main features of these mechanisms. Kaessmann, et al. 2009 provides detailed information about retroposition-mediated gene duplication. Most genetics textbooks, including Klug, et al. 2009, describe detailed mechanisms of chromosomal and genome duplication, which are also referred to as aneuploidization and polyploidization, respectively. One term often mentioned in the human and medical genetics literature is “segmental duplication,” which is defined as the duplication of a segment of DNA that has 1,000 nucleotides or more. In addition to unequal crossing-over, segmental duplication can also arise via some other mechanisms that are not well understood. See Marques-Bonet, et al. 2009 for a recent review on this topic. The relative rates of these mutations are unclear. It is likely that the relative rates vary greatly across species. In several species, the abundances of tandem duplicates (generated by unequal crossing-over) and retroduplicates (generated by retroposition) have been surveyed, for instance in Pan and Zhang 2007 and Zhou, et al. 2008. Similarly, genome duplication is much more common in plants than animals. For example, Wood, et al. 2009 infers that 15 percent of angiosperm and 31 percent of fern speciation events are accompanied by genome duplication; while only a few genome duplication events are known in animals, the most famous of which is the two rounds of genome duplication in the origin of vertebrates, which was proposed by Ohno and is now well established (e.g., see Dehal and Boore 2005). But such survey data do not necessarily tell us the relative mutational rates of different types of duplication, because the number of observed duplicates is a function of mutation rate, fixation rate, and retention rate.

Dehal, P., and J. L. Boore. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* 3:e314.

Provides strong genomic evidence for two rounds of genome duplication in early vertebrate evolution.

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

A comprehensive review on the role of retroposition in the formation of new genes, covering expressional, functional, and evolutionary patterns of retroduplicates.

Klug, W. S., M. R. Cummings, C. A. Spencer, and M. A. Palladino. 2009. *Concepts of Genetics*. 9th ed. San Francisco: Pearson Education.

A genetics textbook with detailed information on the genetic mechanisms of chromosome and genome duplication.

Marques-Bonet, T., S. Girirajan, and E. E. Eichler. 2009. The origins and impact of primate segmental duplications. *Trends in Genetics* 25.10: 443–454.

An authoritative review on segmental duplication.

Pan, D., and L. Zhang. 2007. Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. *Genome Biology* 8:R158.

The relative contributions of unequal crossing-over and retroposition to new gene originations are quantified in human and mouse.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* 106.33: 13875–13879.

Using cytogenetic and phylogenetic data, this paper shows that the prevalence of genome duplication in plants is four times what was previously thought. It estimates that 15 percent of angiosperm and 31 percent of fern speciation events are accompanied by ploidy increase.

Zhang, J. 2003. Evolution by gene duplication: An update. *Trends in Ecology and Evolution* 18.6: 292–298.

A comprehensive review of the status of evolution by gene duplication in the genomic era; discusses common modes of gene duplication.

Zhou, Q., G.-J. Zhang, Y. Zhang, et al. 2008. On the origin of new genes in *Drosophila*. *Genome Research* 18.9: 1446–1455.

The relative contributions of various gene duplication mechanisms to new gene originations in *Drosophila* are quantified at the genomic scale.

Fixation of Duplicate Genes

A gene duplication event first occurs in one individual. As is true for other types of mutations, most duplicate genes do not get fixed in a population. The probability that a duplicate gene spreads through a population and gets fixed is determined by the fitness effect of duplication. Gene duplication may be deleterious, beneficial, or neutral to an organism, depending on the type of gene duplication and the genes involved. Wagner 2005 proposed that the energy cost to the cell arising from the replication, transcription, and translation of the extra gene makes gene duplication deleterious. Another reason is dosage imbalance caused by gene duplication. Papp, et al. 2003 and Yang, et al. 2003 show that this possibility is quite high for genes encoding members of large protein

complexes, especially when they are haploinsufficient, as demonstrated in Qian and Zhang 2008. The third reason is that a retroduplicate may be inserted into a gene or a functional element, disrupting the normal function of the genome. As discussed in Zhang 2003 and Kondrashov and Kondrashov 2006, gene duplication could also be advantageous if extra gene product is useful to the organism. For example, Perry, et al. 2007 shows that duplication of the salivary amylase gene is beneficial in human populations that eat high starch food, because amylase helps digest starch. Gene duplication could also be neutral, or nearly so, if its fitness effect is smaller than the inverse of the effective population size. An unresolved question is: Among those duplicates that did get fixed and are observed today, were they fixed mostly by positive selection or by random genetic drift?

Kondrashov, F. A., and A. S. Kondrashov. 2006. Role of selection in fixation of gene duplications. *Journal of Theoretical Biology* 239.2: 141–151.

This paper argues that many duplicate genes are fixed by positive selection because of the benefit of the increased gene dosage.

Papp, B., C. Pal, and L. D. Hurst. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424.6945: 194–197.

This paper shows that gene duplication can be harmful when it causes dosage imbalance, and that protein complex members are particularly sensitive to dosage balance.

Perry, G. H., N. J. Dominy, K. G. Claw, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 39:1256–1260.

This population genetic study demonstrates the advantage of extra copies of the salivary amylase gene to human populations with high-starch diets.

Qian, W., and J. Zhang. 2008. Gene dosage and gene duplicability. *Genetics* 179.4: 2319–2324.

This paper clarifies the difference between gene dosage and gene dosage balance, showing that it is gene dosage balance rather than dosage itself that impacts gene duplicability.

Wagner, A. 2005. Energy constraints on the evolution of gene expression. *Molecular Biology and Evolution* 22.6: 1365–1374.

Based on theoretical calculation, this paper shows that, for many genes, duplication increases the energy cost of protein production to such an extent that its harm is detectable by natural selection, at least in species with large populations (e.g., the budding yeast *Saccharomyces cerevisiae*).

Yang, J., R. Lusk, and W.-H. Li. 2003. Organismal complexity, protein complexity, and gene duplicability. *Proceedings of the National Academy of Sciences of the United States of America* 100.26: 15661–15665.

This paper shows that genes encoding members of large protein complexes tend not to duplicate, presumably due to the sensitivity of protein complex members to dosage balance.

Zhang, J. 2003. Evolution by gene duplication: An update. *Trends in Ecology and Evolution* 18.6: 292–298.

A comprehensive review on evolution by gene duplication in the genomic era; discusses various fates of duplicate genes.

Copy Number Variation

Just as one would find many single nucleotide polymorphisms in a population, recent studies have revealed a large number of gene copy number variations (CNVs) in humans and other species (e.g., Redon, et al. 2006; She, et al. 2008). These CNVs are the results of gene duplication or gene deletion that have not been fixed in a population, as reviewed in Hastings, et al. 2009. Although many CNVs probably have negligible fitness effects, some are known to cause diseases, as shown in Craddock, et al. 2010; Merikangas, et al. 2009; and Inaki and Liu 2012. The suspected mechanism is a disturbance of the dosage of the gene involved or the dosage of the involved gene relative to those of other genes in the genome (i.e., dosage balance). This is the population-level manifestation of the evolutionary observation that genes sensitive to changes in dosage or dosage balance tend not to duplicate (see the previous section, Fixation of Duplicate Genes).

Craddock, N., M. E. Hurles, N. Cardin, et al. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464:713–720.

A study by the Wellcome Trust Case Control Consortium showing associations of CNVs with Crohn's disease, rheumatoid arthritis, type 1 diabetes, type 2 diabetes, etc.

Hastings, P. J., J. R. Lupski, S. M. Rosenberg, and G. Ira. 2009. Mechanisms of change in gene copy number. *Nature Reviews Genetics* 10.8: 551–564.

A review on the genetic mechanisms of copy number variation.

Inaki, K., and E. T. Liu. 2012. Structural mutations in cancer: Mechanistic and functional insights. *Trends in Genetics* 28.11: 550–559.

A review on the role of CNVs in tumorigenesis.

Merikangas, A. K., A. P. Corvin, and L. Gallagher. 2009. Copy-number variants in neurodevelopmental disorders: Promises and challenges. *Trends in Genetics* 25.12: 536–544.

A review on the role of CNV in neurodevelopmental diseases such as autism and schizophrenia.

Redon, R., S. Ishikawa, K. R. Fitch, et al. 2006. Global variation in copy number in the human genome. *Nature* 444.7118: 444–454.

An early study of the genomic pattern of copy number variation in humans.

She, X., Z. Cheng, S. Zöllner, D. M. Church, and E. E. Eichler. 2008. Mouse segmental duplication and copy number variation. *Nature Genetics* 40.7: 909–914.

A genomic study of copy number variation in mice.

Pseudogenization after Duplication

Most duplicate genes are expected to become pseudogenes because of the lack of selective constraint due to the functional redundancy between the duplicate gene and its parent gene. This prediction is supported by the age distribution of the duplicate genes in a genome, first shown in Lynch and Conery 2000. That is, the number of (non-pseudogenized) duplicate genes decreases quickly with age. Reciprocal pseudogenizations of duplicate genes in separate populations could reduce the fecundity of their hybrids, resulting in reproductive isolation and speciation. This divergent resolution model of speciation was proposed in Werth and Windham 1991 and Lynch and Force 2000. There has not been an empirical observation that unequivocally proves the model, but

Scannell, et al. 2006 showed that patterns of yeast speciation following a whole-genome duplication event are consistent with this model. Most retroduplicates, duplicate genes generated by retroposition, are “dead on arrival” because of the lack of expression, as the promoter cannot be duplicated together with a coding region by retroposition. Pseudogenes arising from retroduplicates are referred to as processed pseudogenes. Because the number of processed pseudogenes generated from a parent gene should be proportional to the expression level of the parent gene in the germ line, Podlaha and Zhang 2009 proposed to infer the germ-line expression level of a gene at an ancient time from the number of its processed pseudogenes of the corresponding age that are found in the current genome. Furthermore, processed pseudogenes may also reveal ancient splice variants of a gene that are no longer observed today, as suggested in Shemesh, et al. 2006.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290.5494: 1151–1155.

The first genome-wide analysis of the rate of fixed gene duplication and the rate of pseudogenization after gene duplication.

Lynch, M., and A. G. Force. 2000. The origin of interspecific genomic incompatibility via gene duplication. *American Naturalist* 156.6: 590–605.

Independently proposes the divergent resolution hypothesis.

Podlaha, O., and J. Zhang. 2009. Processed pseudogenes: The “fossilized footprints” of past gene expression. *Trends in Genetics* 25.10: 429–434.

Proposes that the number of processed pseudogenes generated from a parent gene can be used to infer the expression level of the parent genes in ancient times.

Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440.7082: 341–345.

Provides some comparative genomic evidence for the divergent resolution hypothesis in yeast.

Shemesh, R., A. Novik, S. Edelheit, and R. Sorek. 2006. Genomic fossils as a snapshot of the human transcriptome. *Proceedings of the National Academy of Sciences of the United States of America* 103.5: 1364–1369.

Proposes that processed pseudogenes provide information on ancient splice variants.

Werth, C. R., and M. D. Windham. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *American Naturalist* 137.4: 515–526.

Proposes the divergent resolution hypothesis that reciprocal losses of duplicate genes in isolated populations can lead to reproductive isolation and speciation.

Stable Retention of Duplicate Genes

A duplicate gene cannot be stably retained in a genome unless it is useful to the organism such that the loss of the gene would cause an immediate decrease in fitness. Five mechanisms have been proposed to account for duplicate gene retention. First, Zhang 2003 (cited under Mechanisms of Gene Duplication) discusses the possibility that when an increased gene dose is beneficial, duplicate genes can be stably retained because gene loss would lower the dosage and fitness. Second, the neofunctionalization model presented in Ohno 1970 asserts that, after gene duplication, one gene performs the ancestral function while the other adopts

a new and useful function. Third, Force, et al. 1999 proposes that an ancestral gene may already possess dual functions; after duplication, each copy may adopt one of the ancestral functions but lose the other function. Such subfunctionalization dictates that neither gene can be lost without reducing fitness. Fourth, the escape from adaptive conflict (EAC) model, also known as the specialization model, is a hybrid of the neofunctionalization and subfunctionalization models and was originally proposed in Hughes 1994. In EAC, the ancestral gene possesses dual functions, but neither function can be optimized because optimizing one function compromises the other. After duplication, the ancestral functions can be subdivided for the two genes, and the removal of the conflict allows each function to be optimized. Fifth, Näsvall, et al. 2012 suggests that the progenitor gene before duplication possesses a major function and a minor function; duplication is immediately beneficial due to the increase of the dose of the minor function. Duplication further provides the opportunity for the two functions to be subdivided for the two genes and optimized. This IAD (innovation, amplification, and divergence) model is identical to the EAC model except that doubling the gene dose per se is adaptive in the former but not necessarily so in the latter. While each of these models has been supported by some real cases (e.g., Zhang, et al. 1998 for the neofunctionalization model, Force, et al. 1999 for the subfunctionalization model, and Hittinger and Carroll 2007 for the EAC model), the relative importance of these models is unclear, especially at the genomic scale, as discussed in Zhang 2013. Lynch and Force 2000 points out that, theoretically, the probability of duplicate gene retention by subfunctionalization is much greater than that by neofunctionalization, especially when the population size is not very large. When the population size gets larger, the chance that a beneficial neofunctionalizing mutation occurs and gets fixed before the occurrence of subfunctionalization increases and the relative role of neofunctionalization in duplicate gene retention expands.

Force, A., M. Lynch, F. Bryan Pickett, A. Amores, Y.-L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151.4: 1531–1545.

This seminal paper proposes the idea of subfunctionalization as a mechanism of duplicate gene retention.

Hittinger, C. T., and S. B. Carroll. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449.7163: 677–681.

This beautiful study demonstrates how gene duplication contributes to adaptation by resolving a conflict between a regulatory function and a catalytic function of a dual-function gene.

Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society B: Biological Sciences* 256.1346: 119–124.

This paper proposes the idea of functional specialization after gene duplication, a precursor of the EAC model.

Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154.1: 459–473.

This theoretical study calculates the relative probabilities of subfunctionalization, neofunctionalization, and pseudogenization under certain population genetic parameters.

Näsvall, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338.6105: 384–387.

This paper uses experimental evolution to demonstrate that a minor function of a progenitor gene may be enhanced in one of the two daughter genes, leading to the stable retention of both duplicates, thereby providing evidence for the IAD model.

Ohno, S. 1970. *Evolution by Gene Duplication*. Berlin and New York: Springer-Verlag.

It proposes the neofunctionalization model of duplicate gene evolution.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences of the United States of America* 95.7: 3708–3713.

This paper reported the origin of a novel function in one of the daughter genes generated by gene duplication and showed for the first time that positive selection was responsible for the post-duplication neofunctionalization.

Zhang, J. 2013. Gene duplication. In *The Princeton Guide to Evolution*. in press. Edited by J. Losos. Princeton, NJ: Princeton Univ. Press.

Provides a summary of the current view on various mechanisms of stable retention of duplicate genes in evolution.

Rate of Gene Duplication

The standard method to estimate the rate of gene duplication at the mutational level is to use mutation accumulation (MA) lines, which are maintained for many generations at such a low population size that natural selection is virtually absent. Lynch, et al. 2008 showed that the rate of the appearance of new duplicates is on the order of 10^{-6} per gene per generation in yeast (*Saccharomyces cerevisiae*) MA lines, and Lipinski, et al. 2011 showed that it is 10^{-7} per gene per generation in worm (*Caenorhabditis elegans*) MA lines. These rates are substantially greater than the rate of gene duplication that gets fixed in evolution, which is estimated in Lynch and Conery 2000 to be about 0.01 per gene per million years in many eukaryotes. This discrepancy suggests that the vast majority of gene duplication events are deleterious and thus do not reach fixation. This conclusion is also supported by surveys of CNVs in fruit flies in Emerson, et al. 2008.

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* 320.5883: 1629–1631.

This study estimates the distribution of the direction and strength of natural selection on *Drosophila* CNVs.

Lipinski, K. J., J. C. Farslow, K. A. Fitzpatrick, M. Lynch, V. Katju, and U. Bergthorsson. 2011. High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Current Biology* 21.4: 306–310.

This study estimates the gene duplication rate at the mutational level in *C. elegans*.

Lynch, M., W. Sung, K. Morris, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 105.27: 9272–9277.

This study offers the first genomic estimate of gene duplication rate at the mutational level in any species.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290.5494: 1151–1155.

Based on comparative genomics, estimates that the rate of gene duplication that gets fixed in evolution is ~0.01 per gene per million years in several eukaryotes.

Determinants of Gene Duplicability

Gene duplicability refers to the probability that a gene is duplicated at the mutational level, fixed, and stably retained in the genome.

Papp, et al. 2003 and Yang, et al. 2003 show that gene dose balance is an important factor impacting gene duplicability. That is, genes sensitive to dosage balance tend not to duplicate unless all genes that need to be dosage-balanced are duplicated together via whole-genome duplication. He and Zhang 2005 shows that gene duplicability increases with the complexity of the gene involved, probably because complex genes are prone to subfunctionalization. Also, He and Zhang 2006 shows that duplicability decreases as the importance of a gene rises, probably because gene duplication tends to disturb the established cellular homeostasis, and such homeostasis is more sensitive to the dose of important genes. There are several other gene functional properties that have been observed to correlate positively with gene duplicability, although the underlying mechanisms are often unclear. These include functioning as metabolic enzymes, as shown in Marland, et al. 2004; interacting with the external environment, as shown in Prachumwat and Li 2006; interacting with fewer protein partners, as shown in Prachumwat and Li 2006; controlling physiological traits (rather than morphological traits), as shown in Liao, et al. 2010; and having more phosphorylation sites, as shown in Amoutzias, et al. 2010.

Amoutzias, G. D., Y. He, J. Gordon, et al. 2010. Posttranslational regulation impacts the fate of duplicated genes. *Proceedings of the National Academy of Sciences of the United States of America* 107.7: 2967–2971.

Genes encoding proteins with more phosphorylation sites are more likely to be retained after duplication.

He, X., and J. Zhang. 2005. Gene complexity and gene duplicability. *Current Biology* 15.11: 1016–1021.

This paper shows that gene duplicability increases with gene complexity, where gene complexity refers to the number of protein domains, protein length, and number of *cis*-regulatory motifs.

He, X., and J. Zhang. 2006. Higher duplicability of less important genes in yeast genomes. *Molecular Biology and Evolution* 23.1: 144–151.

This study demonstrates in fungi that gene duplicability decreases with the rise of gene importance, which is measured by the fitness defect of gene deletion.

Liao, B.-Y., M.-P. Weng, and J. Zhang. 2010. Contrasting genetic paths to morphological and physiological evolution. *Proceedings of the National Academy of Sciences of the United States of America* 107.16: 7353–7358.

Genes affecting physiological traits duplicate more frequently than genes affecting morphological traits in mammalian evolution.

Marland, E., A. Prachumwat, N. Maltsev, Z. Gu, and W.-H. Li. 2004. Higher gene duplicabilities for metabolic proteins than for nonmetabolic proteins in yeast and *E. coli*. *Journal of Molecular Evolution* 59.6: 806–814.

This study demonstrates that enzyme genes are more likely to duplicate than non-enzyme genes.

Papp, B., C. Pal, and L. D. Hurst. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424.6945: 194–197.

Using yeast genomic data, this paper shows that gene duplication can be harmful when it causes dosage imbalance, and that protein complex members are particularly sensitive to dosage balance.

Prachumwat, A., and W.-H. Li. 2006. Protein function, connectivity, and duplicability in yeast. *Molecular Biology and Evolution* 23.1: 30–39.

This study shows that the duplicability of a gene decreases as the number of its protein interactions increases. It further shows higher gene duplicability for proteins interacting with external environments than for proteins localized within intracellular

compartments.

Yang, J., R. Lusk, and W.-H. Li. 2003. Organismal complexity, protein complexity, and gene duplicability. *Proceedings of the National Academy of Sciences of the United States of America* 100.26: 15661–15665.

Demonstrates that duplicability of protein-coding genes depends on the involvement of the protein in protein complexes and organismal complexity.

Functional Redundancy among Duplicate Genes

Because duplicate genes share a common ancestry, it is not surprising that many duplicate genes are functionally redundant to some extent (Wagner 2000). For example, Gu, et al. 2003 showed in *S. cerevisiae* and Conant and Wagner 2004 showed in *C. elegans* that deleting a duplicate gene tends to impose a smaller fitness effect than deleting a singleton gene. Kim, et al. 2010 and Hanada, et al. 2009 report a similar pattern in fission yeast and the plant *Arabidopsis thaliana*, respectively. However, Liao and Zhang 2007 and Liang and Li 2007 did not find this pattern in mammals. He and Zhang 2006 suggests that the above estimates of functional redundancy may have been exaggerated, because of the higher duplicability of less important genes.

Conant, G. C., and A. Wagner. 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proceedings of the Royal Society B: Biological Sciences* 271.1534: 89–96.

Genomic demonstartion of functional redundancy between duplicate genes in *C. elegans*.

Gu, Z., L. M. Steinmetz, X. Gu, C. Scharfe, R. W. Davis, and W.-H. Li. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421.6918: 63–66.

Genomic demonstartion of functional redundancy between duplicate genes in budding yeast.

Hanada, K., T. Kuromori, F. Myouga, T. Toyoda, W.-H. Li, and K. Shinozaki. 2009. Evolutionary persistence of functional compensation by duplicate genes in *Arabidopsis*. *Genome Biology and Evolution* 1:409–414.

Genomic demonstartion of functional redundancy between duplicate genes in *Arabidopsis*.

He, X., and J. Zhang. 2006. Higher duplicability of less important genes in yeast genomes. *Molecular Biology and Evolution* 23.1: 144–151.

This study shows that less important genes tend to duplicate, suggesting an alternative explanation of why deleting a duplicate gene is less deleterious than deleting a singleton gene.

Kim, D. U., J. Hayles, D. Kim, et al. 2010. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nature Biotechnology* 28:617–623.

Genomic demonstartion of functional redundancy between duplicate genes in fission yeast.

Liang, H., and W.-H. Li. 2007. Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends in Genetics* 23.8: 375–378.

It shows that there is no evidence for functional redundancy between duplicate genes in mammals. This paper and Liao and Zhang 2007 stimulated wide discussion of why mammals behave differently from other species.

Liao, B.-Y., and J. Zhang. 2007. Mouse duplicate genes are as essential as singletons. *Trends in Genetics* 23.8: 378–381.

It demonstrates that there is no functional redundancy between duplicate genes in mice. This paper and Liang and Li 2007 stimulated wide discussion of why mammals behave differently from other species.

Wagner, A. 2000. Robustness against mutations in genetic networks of yeast. *Nature Genetics* 24.4: 355–361.

This first genomic test of the hypothesis of functional redundancy between duplicate genes. Probably because of the small sample size, the result was not positive.

Mechanisms of Long-Term Maintenance of Functional Redundancy

The strongest evidence for functional redundancy between duplicate genes is that deleting a pair of duplicate genes in *S. cerevisiae* has a greater fitness effect than expected from deleting either gene alone. This was independently discovered in DeLuna, et al. 2008, Musso, et al. 2008, and Dean, et al. 2008. Although the degree of functional redundancy between a pair of duplicates should gradually decline with the time since duplication, Vavouri, et al. 2008 reports that some duplicate genes are redundant even hundreds of millions of years after duplication. In the case of ribosomal proteins, ribosomal RNAs, histones, transfer RNAs, and other molecules that are in high demand in the cell, functional similarity among duplicates is selectively favored and thus requires no other explanation. For other duplicate genes, the unexpectedly long retention of functional redundancy is puzzling. Kafri, et al. 2005 suggests that redundancy is beneficial in itself because it protects the organism from the potential harm of deleterious mutations, much like the backup role of a spare tire for a car. Although this backup hypothesis appears consistent with observations made in a small number of duplicate genes (in DeLuna, et al. 2010), it cannot on its own be correct in cellular organisms, because the benefit of backup is likely to be too small to detect by natural selection, as argued in Zhang 2012. Another thought, known as the piggyback hypothesis (Vavouri, et al. 2008), is that paralogous genes have some nonoverlapping functions as well as some overlapping functions, and the existence of the latter is a byproduct of the former owing to strong protein structural constraints. More recently, Qian, et al. 2010 reported that reduction of gene expression after duplication, a special form of subfunctionalization, is quite common in yeast and mammals. This expression reduction prevents the loss of either duplicate gene because such loss would render the total expression level after duplication lower than that before duplication, which would be deleterious. The expression reduction, when it is sufficiently large, would require both daughter genes to retain all ancestral functions, preventing the occurrence of functional divergence.

Dean, E. J., J. C. Davis, R. W. Davis, and D. A. Petrov. 2008. Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genetics* 4:e1000113.

Using double deletion of duplicate genes in budding yeast, this study provides strong evidence for functional redundancy between duplicate genes.

DeLuna, A., M. Springer, M. W. Kirschner, and R. Kishony. 2010. Need-based up-regulation of protein levels in response to deletion of their duplicate genes. *PLoS Biology* 8:e1000347.

Provides real examples that are consistent with the backup hypothesis.

DeLuna, A., K. Vetsigian, N. Shoresh, et al. 2008. Exposing the fitness contribution of duplicated genes. *Nature Genetics* 40.5: 676–681.

The authors showed that deleting a pair of duplicate genes simutaneously in budding yeast causes a much greater fitness drop than one would expect from the individual deletions of the two genes, demonstrating that duplicate genes are highly redundant in function.

Kafri, R., A. Bar-Even, and Y. Pilpel. 2005. Transcription control reprogramming in genetic backup circuits. *Nature Genetics* 37.3: 295–299.

Proposes the backup hypothesis of the retention of functionally redundant duplicate genes.

Musso, G., M. Costanzo, A. M. Smith, et al. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Research* 18.7:1092–1099.

By deleting from the yeast genome paralogous genes that originated from a whole-genome duplication about 100 million years ago, the authors found substantial functional redundancy between relatively ancient duplicate genes.

Qian, W., B. Y. Liao, A. Y. Chang, and J. Zhang. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* 26.10: 425–430.

Duplicate genes tend to have reduced expression, compared with their progenitors, which may explain the stable retention of duplicates with functional redundancy.

Vavouri, T., J. I. Semple, and B. Lehner. 2008. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends in Genetics* 24.10: 485–488.

This study offers examples of old duplicate genes that are functionally redundant, which it explains by the piggyback hypothesis.

Zhang, J. 2012. Genetic redundancies and their evolutionary maintenance. In *Evolutionary Systems Biology*. Edited by O. S. Soyer, 279–300. New York: Springer.

A comprehensive review on the evidence for functional redundancy of duplicate genes and the potential evolutionary genetic mechanisms explaining the long retention of the redundancy.

Functional Diversification of Duplicate Genes

Although functional redundancy among duplicate genes is not uncommon, the most common observation among stably retained duplicates is their functional divergence, which occurs at the level of gene expression and/or protein function. For example, Gu, et al. 2002 and Makova and Li 2003, two genome-wide studies, show that gene expression patterns diverge quickly after gene duplication. Chen and Zhang 2012 shows that duplicate genes tend to be less similar in their expression levels than orthologous genes of comparable divergence times. In terms of protein functional divergence, the degree of divergence varies greatly among different genes. In many cases, duplicate genes perform similar types of function, but with different activities or specificities, such as the paralogous pancreatic ribonucleases of leaf-eating monkeys studied in Zhang, et al. 2002. Other well-known examples include odorant receptor (OR) genes, which form the largest gene family in the vertebrate genome. Although each OR is able to recognize only a limited number of odorants, vertebrates are believed to be able to detect 10,000 or more odorants because of the possession of hundreds of functional OR genes that recognize different odorants. Nei, et al. 2008 provides an excellent evolutionary overview of this gene family. Xu, et al. 2012 reports rapid divergence of exon-intron structures between duplicate genes in plants, which may underlie their functional divergence. Occasionally, retroduplicates may be expressed when they are fortuitously inserted into a genomic region that harbors a promoter. There is accumulating evidence, reviewed in Kaessmann, et al. 2009, that retroposition gives rise to new genes whose functions are drastically different from their parent genes, probably because of completely different expression patterns and/or the involvement of gene fusion.

Chen, X., and J. Zhang. 2012. The ortholog conjecture is untestable by the current gene ontology but is supported by RNA sequencing data. *PLoS Computational Biology* 8:e1002784.

A genomic analysis of gene expression evolution with and without gene duplication.

Gu, Z., D. Nicolae, H. H.-S. Lu, and W.-H. Li. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends in Genetics* 18.12: 609–613.

Rapid yeast gene expression pattern changes after gene duplication.

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

A review on the role of retroposition in gene duplication, focusing on the expressional, functional, and evolutionary patterns of retroduplicates.

Makova, K. D., and W.-H. Li. 2003. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Research* 13.7: 1638–1645.

Rapid human gene expression pattern changes after gene duplication.

Nei, M., Y. Niimura, and M. Nozawa. 2008. The evolution of animal chemosensory receptor gene repertoires: Roles of chance and necessity. *Nature Reviews Genetics* 9.12: 951–963.

A review of the evolution of animal chemosensory receptors, including odorant receptors, vomeronasal receptors, and taste receptors.

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.

A surprising finding of rapid exon-intron structure divergence between duplicate genes in plants.

Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics* 30.4: 411–415.

A case study of functional modification after gene duplication.

Outstanding Questions

In spite of numerous empirical studies of functional differences between duplicate genes, it is unclear whether these differences are primarily attributable to subfunctionalization or neofunctionalization. There is also the view, from He and Zhang 2005, that both happen frequently but subfunctionalization occurs immediately after gene duplication whereas neofunctionalization is a slower process. A related question is the role of positive selection in the functional divergence of duplicate genes. Positive selection is absent in the pure subfunctionalization model, but it must be involved in the EAC and IAD models. It is commonly thought, and has been demonstrated in case studies such as Zhang, et al. 1998, that positive selection is involved in neofunctionalization. But, in theory, neofunctionalization can also occur by random fixation of neutral mutations; the utility of the new function may be realized after its fixation, upon an alteration of the genetic background or environment. While gene duplication is undoubtedly the primary source of

new genes in evolution, its contribution to adaptation requires more scrutiny. The contribution of gene duplication to speciation, reviewed in Nei and Nozawa 2011, is another area lacking critical empirical data. The same can be said to whole-genome duplication. Although many authors believe that whole-genome duplication promotes adaptation, speciation, and/or biodiversity (e.g., see Crow and Wagner 2006), critical evidence is still lacking.

Crow, K. D., and G. P. Wagner. 2006. What is the role of genome duplication in the evolution of complexity and diversity? ***Molecular Biology and Evolution*** **23.5: 887–892.**

In this perspective article, the authors proposed that genome duplication promotes biodiversity, biocomplexity, and adaptation.

He, X., and J. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. ***Genetics*** **169.2: 1157–1164.**

Based on yeast protein-protein interaction data, this study reveals substantial subfunctionalization and neofunctionalization. Subfunctionalization occurs quickly after gene duplication, whereas neofunctionalization occurs more slowly.

Nei, M., and M. Nozawa. 2011. Roles of mutation and selection in speciation: From Hugo de Vries to the modern genomic era. ***Genome Biology and Evolution*** **3:812–829.**

Gene duplication may have contributed greatly to speciation.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. ***Proceedings of the National Academy of Sciences of the United States of America*** **95.7: 3708–3713.**

This paper showed for the first time that positive selection was responsible for the neofunctionalization after gene duplication.

LAST MODIFIED: 01/13/2014

DOI: 10.1093/OBO/9780199941728-0002

BACK TO TOP