

# The altered evolutionary trajectories of gene duplicates

Michael Lynch and Vaishali Katju

Department of Biology, Indiana University, Bloomington, IN 47405, USA

**Gene duplication is widely regarded as the predominant mechanism by which genes with new functions and associated phenotypic novelties arise. However, the mutational events and population-genetic mechanisms that lead to the short-term preservation of duplicate genes are not necessarily the same as those exhibited by well-established paralogs en route to the origin of new beneficial features. Thus, although recent genome-wide analyses have revealed striking patterns of protein-sequence divergence among the members of surviving paralogous gene families, the mechanisms responsible for the historical development of these patterns remain unclear.**

The idea that gene duplication has a fundamental role in the origin of phenotypic diversity has motivated numerous proposals to explain how a new gene copy can 'march to the beat of a different drummer' from its predecessor, and evolve a novel function. The alternative models differ in several fundamental respects, including: (i) the time of origin of the functional novelty (preceding or following gene duplication); (ii) the types of mutational events that drive the crucial phase of early establishment (beneficial versus degenerative); and (iii) the rates of molecular evolution in the two copies (equal versus asymmetric).

Ohno's [1] widely cited model for the origin of functional novelties following gene duplication explicitly excluded an initial role for natural selection. Instead, he postulated that gene duplication creates a redundant locus that is free to accumulate otherwise 'forbidden' mutations (i.e. changes that lead to a loss of function) as long as the original copy still performs the essential ancestral tasks. Ohno imagined that mutations arising during this period of relaxed selection direct the redundant gene copy to one of two alternative pathways: nonfunctionalization to a 'degenerate, nonsense DNA base sequence' or neofunctionalization to a 'hitherto non-existent function.' Under this model, neofunctionalization is the only mechanism by which duplicate genes can become permanently preserved. However, we now know that preservation can also be brought about by subfunctionalization, the accumulation of degenerative mutations causing complementary loss of subfunctions in the two members of the pair [2–4].

Ohno's model makes specific predictions with respect to neofunctionalized duplicate genes. First, the descendant

copy is expected to evolve its novel function after duplication. Second, neofunctionalization is expected to occur after an initial phase of nonfunctionalization under conditions of completely relaxed selective constraints [1]. Third, the two copies are expected to exhibit asymmetric rates of molecular evolution, with the ancestral copy experiencing purifying selection and the redundant copy evolving at an accelerated rate as a consequence of initially being 'ignored by natural selection' and subsequently experiencing positive selection for a new function.

An alternative model for the evolution of novel proteins, suggested by Piatigorsky and Wistow [5], and Hughes [6], invokes the presence of functional diversification before gene duplication. In its simplest form, this 'adaptive-conflict' model assumes a generalist ancestral gene capable of two or more distinct but pleiotropically constrained functions. Gene duplication then enables both copies to become specialized in distinct subsets of the ancestral functions. Similar to Ohno's model, the adaptive-conflict model invokes the accumulation of previously forbidden mutations, but in this case the partitioning of ancestral functions is brought about by positive darwinian selection for mutations that refine the restricted set of functions of each paralog. These changes are expected to arise within a relatively short time-period following duplication (before debilitating degenerative mutations have been incurred by either copy), and strong asymmetries in the rates of molecular evolution are not expected (although they are not entirely ruled out). It should also be noted that a key distinction between the adaptive-conflict and subfunctionalization models is that the subfunctionalization model requires no positive selection.

Despite their seductive simplicity, both models for the acquisition of new functions by duplicate genes ignore several key population-genetic issues. One prominent problem is the absence of a mechanism by which natural selection can sequester one member of a duplicate pair for evolutionary exploration, while retaining the other for the maintenance (and/or improvement) of the ancestral function. If both copies are identical initially and exposed to identical mutation and selection pressures, how can they embark on different evolutionary pathways? Resolving this problem is central to our understanding of the evolutionary consequences of gene duplication because the realization of any long-term advantages of duplication requires the existence of short-term mechanisms to insure the initial preservation of both members of a pair.

Corresponding author: Michael Lynch (mlynch@bio.indiana.edu).

Available online 11 September 2004

A second problem, specific to the Ohno model, concerns the assumption that it is always the descendent gene copy that evolves a novel function. Because a new duplicate gene is present initially as a single copy in a single individual, there are actually many more opportunities for the appearance of neofunctionalizing mutations at the ancestral locus. This issue is of relevance because the loss of an ancestral function at the original locus is equivalent to a change in chromosomal location, which has implications for the origin of species-isolating barriers [4,7]. However, discriminating parent from daughter copies is not a simple matter. For example, in relatively young duplicate pairs, the descendent copy can be reasonably inferred to be the one that is not fixed at the population level; however, with older copies, ancestral inference requires mapping and sequence data from a closely related single-copy outgroup species. To date, few empirical studies have attempted to make this distinction.

Although there are good reasons to think that key mutational events driving evolution might reside in introns, untranslated regions (UTRs) and/or nontranscribed regulatory regions, almost all molecular-evolution research has focused on the more accessible and interpretable protein-sequence data, and this also applies to the studies of duplicate-gene evolution. Several recent studies involving whole-genome sequence analysis have revealed the cumulative changes in proteins encoded by the members of surviving gene pairs, but so far these descriptive analyses have generated more questions than they have answered. In this article, we review the contributions and limitations of comparative sequence analysis to the understanding of duplicate-gene evolution, giving special attention to the relative evolutionary rates of paralogs because these data have been invoked frequently as evidence for or against the Ohno model.

### Inferences from patterns of molecular evolution

The cumulative nature of evolutionary change is a major impediment to inferring the evolutionary mechanisms leading to the divergence of extant paralogous genes. The problem is directly analogous to the bane of speciation genetics – the key preservational mutational events that occur early in the evolutionary history of a pair of duplicates become progressively obscured by subsequent changes as the pair ages. To overcome this difficulty, we introduced a method to deconvolute the historical development of selective constraints from the cumulative divergence data on the entire assemblage of duplicate-gene pairs within a species [8,9]. This method enables one to estimate the instantaneous relative rates of substitution at replacement ( $R$ ) and silent ( $S$ ) sites throughout the life-history of an average duplicate pair, with  $S$  serving as a surrogate measure of the age of a duplicate pair. During times at which  $dR/dS=1$ , replacement sites are evolving (on average) at the neutral rate, whereas  $dR/dS<1$  implies a period of average purifying selection at the amino-acid level. When applied to the genomes of a wide range of eukaryotes, this type of analysis suggests that  $dR/dS$  for newborn duplicates is generally 0.4–1.0, declining to ~0.02–0.09 for ancient duplicates.

Although these and other [10–14] observations consistently indicate that protein evolution proceeds more rapidly in young gene duplicates than in their single-copy orthologs, numerous issues remain unresolved. First, elevated levels of  $dR/dS$  in young duplicates could be a consequence of relaxed selection against degenerative mutations, as suggested by Ohno and by the subfunctionalization model, or a consequence of positive directional selection as suggested by the adaptive-conflict model, or both. Moreover, there might be an intrinsic bias to whole-genome analyses that lump all functional categories of genes together. For example, proteins that evolve slowly and are expressed broadly when in single copy appear to be more likely to become preserved as duplicates than rapidly evolving and/or narrowly expressed proteins [15], perhaps because of their greater vulnerability to subfunctionalization. In addition, highly expressed housekeeping genes tend to be smaller than genes with strong tissue-specific expression patterns [16], which increases their likelihood of being fully contained within duplication spans (discussed in the following section). The salient issue is that, to some extent, the downward trend in  $dR/dS$  with increasing  $S$  could be a consequence of the differential features of surviving versus non-surviving pairs of duplicates, rather than a simple reflection of the evolutionary trajectory of any single duplicate pair.

Second, measures of paralog sequence divergence do not provide insight into whether the evolutionary rates of one or both pair members are altered after duplication, an issue whose resolution requires information on a single-copy (pre-duplication) gene within a closely related species. If both members of a duplicate pair are evolving under the same level of constraint, they should be equally divergent from the outgroup, and this is what the first studies of this nature found [12,17,18]. However, because these analyses focused on relatively old duplicates, an early phase of asymmetry might have been obscured by subsequent nucleotide substitutions. It is therefore noteworthy that ~60% of relatively young gene duplicates ( $S<0.3$ ) from the human genome exhibit asymmetric rates of evolution [19], and that comparable studies in several other eukaryotes yield qualitatively similar results [20]. The tendency for the faster-evolving members of human gene pairs to accumulate substitution changes in a more spatially even pattern than the more slowly evolving members raises the possibility that selection is relaxed in faster-evolving pair members [19].

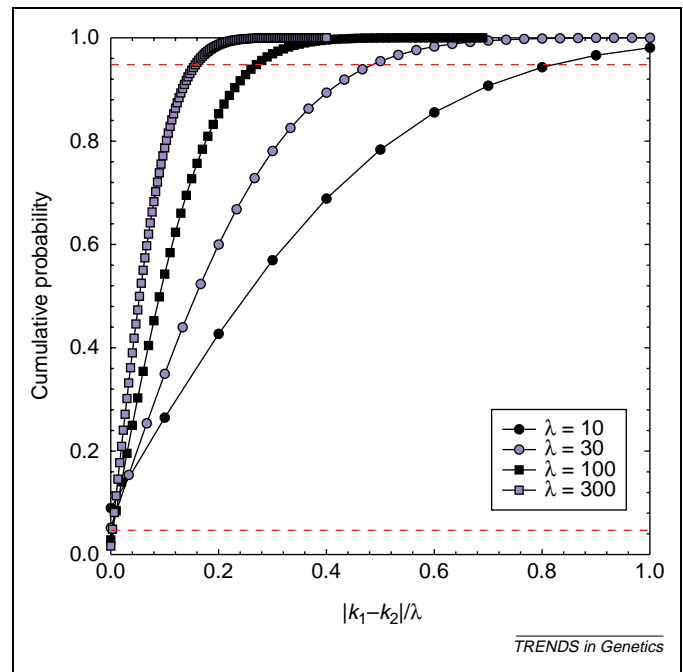
Collectively (and provisionally), these results support the idea that deleterious mutations that can never be fixed in single-copy genes might often accumulate in a nearly neutral fashion shortly after gene duplication. In principle, the chance accumulation of a sufficient number of such mutations in one copy might eventually lead to its complete nonfunctionalization, removing it from the eyes of natural selection, and setting it on the course of pseudogenization. By contrast, the small subset of duplicate genes that escape this fate [8,9] might owe their initial preservation to subfunctionalization by a more balanced distribution of degenerative mutations, which can then be followed by a phase of adaptive-conflict

resolution, neofunctionalization or further subfunctionalization. There are many ways in which subfunctionalization can proceed via coding-region evolution [2,21,22]. One intriguing observation that could help explain the initially elevated rates of protein-sequence evolution in duplicate genes is the enhanced rate of evolution of vertebrate proteins with more restricted tissue-specific patterns of expression [23–25]. A corollary prediction of this model is that pair members with the fewest subfunctions will evolve at the highest rate, and for one example in which the relevant data are available, triosephosphate isomerase in zebrafish, the more rapidly evolving copy does indeed retain the fewest ancestral gene functions [26].

### Statistical power

Before further consideration of potential biological explanations for patterns of duplicate-gene divergence, the statistical power for detecting asymmetric rates needs to be addressed. This turns out to be low for genes that have accumulated few mutations, and because the earliest mutations incurred will generally dictate the fate of a duplicate pair, this can impose a substantial barrier to deciphering the pathways of duplicate-gene evolution from comparative data. The problem can be evaluated simply by assuming the null situation in which substitutions accumulate independently in both members of a pair in an identical poisson fashion. If each copy is expected to have incurred just ten mutations since the duplication event, an absolute difference of at least nine mutations between the two copies would be required to reject the null hypothesis of equal rates, and with an expected 100 mutations per gene, an absolute difference of >27 substitutions would be required (Figure 1). Because an average eukaryotic coding region contains ~1000 bp with ~75% replacement sites, ten and 100 amino-acid changing mutations per gene translate to ~2.6% and 26% divergence at replacement sites, respectively. These levels of replacement-site divergence are beyond the point at which duplicated eukaryotic genes generally exhibit a change in substitution pattern [8,9], implying that asymmetrical patterns of evolution within duplicate-gene pairs are only likely to be detectable well after the events responsible for preservation have become obscured by secondary mutational changes. These results are conservative in that they ignore ambiguities in assigning mutations to paralogs resulting from multiple substitutions per site, parallel evolution, and so on.

To gain some appreciation for the significance of this problem, consider a recent study of 457 pairs of ancient gene duplicates in the yeast *Saccharomyces cerevisiae*, which exhibit an average 24% deviation in the number of amino-acid altering mutations between paralogs [27]. Although the authors concluded that this level of asymmetry provides strong support for Ohno's model, using the method outlined in Figure 1, it can be shown that there is a 5% probability that a gene pair with the average amount of divergence of the *S. cerevisiae* paralogs ( $\lambda \approx 123$ ) will deviate by >24%. Extrapolating back to an earlier phase of evolution with an expectation of 30 amino-acid altering mutations per paralog (a substantial amount of protein evolution), the results in Figure 1 show that



**Figure 1.** The cumulative probability distributions for the difference in numbers of mutations accumulated in gene duplicates assumed to be identical at time zero, under the assumption of equal rates of evolution. Results are given for four different expected numbers of mutations per copy ( $\lambda$ ). The absolute deviation in observed number of changes between the copies ( $k_1$  and  $k_2$ ) is given relative to the expected number. The probability of deviations is given by  $P(d = k_1 - k_2) = e^{-2\lambda} \cdot I_d(2\lambda)$ , where  $I_d(2\lambda)$  is a Bessel function of the  $d$ th order [44]. Data points are given for incrementing absolute differences in units of one; for example, for an expected level of ten mutations per copy, results are given for  $|k_1 - k_2| = 0, 1, 2, \dots, 10$ . The points at which the red horizontal broken lines intersect with the curves give the upper and lower 5% confidence limits for the null distributions. For a given level of  $\lambda$ , observed deviations above the upper limit would be regarded as evidence for significant differences in evolutionary rates (i.e.  $\lambda_1 \neq \lambda_2$ ).

under the assumption of equal rates there is a 32% probability of obtaining a higher deviation than the observed 24% by chance. Thus, although there is reason to believe that some gene duplicates in *S. cerevisiae* have evolved new functions [27], the comparative sequence data do not provide compelling support for Ohno's model in which neofunctionalization is the dominant mechanism of duplicate-gene preservation. Indeed, it is fair to ask whether the observed asymmetries in rates of evolution for well-established duplicate genes in yeast is any greater than what one expects for a random pair of single-copy orthologs of the same age. This being said, it is worth pointing out that it would be a surprise if subfunctionalization were a common mechanism of duplicate-gene preservation in yeast, because this is an evolutionary pathway that is highly unlikely to exist in populations with large effective sizes [22,28].

Although considerable attention has been given to asymmetric rates of evolution of duplicate genes, it is of equal interest to evaluate whether the members of a pair evolve at rates that are more similar than expected under the null model of equal and independent rates (as might be the case if two genes are under identical purifying selection pressures). A situation like this is expected if duplicate genes are maintained by selection to mask the effects of segregating null alleles maintained by recurrent mutation at the ancestral locus, which, unlike the results of subfunctionalization and neofunctionalization, would

serve to preserve all ancestral gene functions in the duplicate copy [29–31]. Simple population-genetic principles suggest the implausibility of the masking model as a mechanism for the permanent preservation of a duplicate gene [4,28]. For example, the initial selective advantage of a newly arisen duplicate under this model is equal to the ancestral-locus fitness load associated with deleterious mutations, which is on the order of  $\mu$ , the per locus mutation rate [32]. Selection of this magnitude will be ineffective unless the power of random genetic drift ( $1/2N_e$ , where  $N_e$  is the effective population size) is even smaller than  $\mu$ , whereas a substantial body of evidence indicates that  $2N_e\mu \gg 1$  for all eukaryotes [33]. The problem is exacerbated by the fact that a duplicate gene also mutates to a defective allele at rate  $\mu$  (i.e. on the order of the selective advantage). Some duplicate genes have a masking effect on a mutant ancestral gene background, as demonstrated in gene-knockout experiments [34,35], but this is a different issue than demonstrating that such effects serve to permanently preserve duplicate genes.

Despite these theoretical objections, it would be useful if the viability of the masking model (or any other dosage-dependent model, [12,29]) could be formally evaluated on empirical grounds. The failure to observe significant asymmetry among paralogs [12] is certainly of interest but this is not equivalent to testing for symmetry in excess of random expectations, which could be more indicative of shared selective pressures. Unfortunately, the results illustrate the extraordinary difficulty in testing for exceptionally symmetrical rates of evolution of paralogous genes (Figure 1). Even with an average of 100–300 substitutions per gene, it is virtually impossible to detect an exceptionally symmetric rate of evolution across a pair of duplicates – a chance difference of just zero or one mutations per gene still has a cumulative probability of  $>5\%$  under a model of equal underlying rates. Thus, the only realistic way of evaluating whether excess symmetry (or asymmetry) is common during the preservational phase of gene duplications is to pool results over many gene pairs and evaluate whether the full distribution of differences among pair members is consistent with the null model.

### Are duplicate genes equal at birth?

We now move on to biological issues that bear on the early fates of duplicate genes. Although most models of gene duplication have assumed that the two members of a pair are functionally identical at birth, there are two simple ways in which this assumption can be commonly violated. First, the breakpoints associated with duplication spans appear to be effectively random, with no obvious mechanisms restricting them to intergenic regions. Thus, except in the case of whole-genome- or large-segmental duplications, the span of a duplicated gene can often fail to contain the full endowment of regulatory elements and/or coding sequence of the ancestral source. Moreover, the movement of a duplicated gene to a novel genomic location can result in a modification of transcriptional activity via position effects. Thus, from the moment of origin, many duplicate genes differ from their parental copy in ways that might restrict the subsequent potential evolutionary

pathways for the pair (e.g. the first step in the subfunctionalization process, loss of an essential feature from one copy, might be fulfilled at the time of gene birth [36]), and in some rare cases incomplete duplication might lead to a neofunctionalized allele. The features of newborn duplicate genes have been examined most closely in the nematode *Caenorhabditis elegans* [37], where the median duplication span ( $\sim 1.4$  kb) is substantially smaller than the average  $\sim 2.5$  kb gene span, implying that partial duplications are common events. Approximately 50% of newborn *C. elegans* duplicates have a unique exon in one or both pair members, and nearly half of them are in adjacent locations but in opposite orientation (i.e. transcribed from opposite strands). Thus, few newborn duplicates in this species are expected to be functionally identical to their parental copy, and there is no reason to expect that the situation will be different in other organisms.

Second, the evolutionary potential of a duplicate pair can be radically altered if the ancestral locus harbors an allelic polymorphism of functional significance. Such a situation is particularly likely for alleles that are homozygous lethal but beneficial in the heterozygous state, which can be maintained at low frequencies by balancing selection at the ancestral locus. As pointed out by Spofford [38], if a single copy of such an allele founds a duplicate locus, fixation at the new locus can be promoted by positive selection while the original locus retains the ancestral function. Alternatively, if the new locus is founded by a ‘wild-type’ allele that rises to a high enough frequency, the selective environment at the ancestral locus will be altered to one of positive selection, with fixation of the rare allele effectively resulting in neofunctionalization [28,39]. In either case, the final outcome is functionally equivalent to fixation of an overdominant gene action that would otherwise only be expressed in rare single-copy heterozygotes. Because the spectrum of mutations arising subsequent to a duplication event must be the same as that before duplication, there is no reason to expect a complete absence of neofunctionalizing mutations at the ancestral locus. Thus, allelic sampling alone, with no subsequent mutations, can sometimes set a population on a course towards neofunctionalization following a duplication event.

A striking example of this process involves the evolution of insecticide resistance in the mosquito *Culex pipiens* [40]. The acetylcholinesterase enzyme normally has an essential role in the central nervous system but a mutant allele at the locus also confers insecticide resistance. Because the mutant allele causes reduced fitness in the absence of pesticides [41], it is maintained at low frequency in insecticide-free environments. However, when challenged with organophosphates, a linked combination of a wild-type and resistant-type alleles rose rapidly to high frequency, presumably with the first member of the pair providing essential ancestral gene functions and the second ameliorating the toxicity of the environment.

### Concluding remarks

Substantial advances in our understanding of the evolution of duplicate genes have emerged from the recent

development of theory and from the pursuit of large-scale gene-sequence analyses. However, in this article, we have shown how far we are from deciphering the mechanistic determinants of the evolutionary fates of duplicate genes. A central issue, the subtleties of which are generally missed in empirical studies, concerns the mutational events that are responsible for the initial preservation of a duplicate gene as opposed to its evolutionary modification following permanent establishment. A duplicate gene need not have a short-term advantage to arise towards fixation because purely neutral processes such as subfunctionalization in sufficiently small populations can permit such change. However, if a duplicate gene is eventually to acquire a new beneficial function, some mutational mechanism is required to insure that it remains permanently preserved and exposed to natural selection. It is becoming increasingly clear that the evolutionary fates of duplicate genes are often dictated, at least in part, by structural features that are inherited at the time of origin rather than by subsequent changes. That is, duplicate genes often evolve in different ways from their ancestral copies not because redundancy frees one copy for evolutionary exploration, as postulated by Ohno, but because the duplication event itself promotes a novel gene *de novo*, thereby creating a novel population-genetic environment. The fact that the preservation of duplicate genes can often be driven by changes in non-coding DNA [2–4,35,42], combined with the limited power of sequence analysis of weakly divergent proteins, suggests that direct functional analysis will ultimately be essential to elucidate how and how often gene duplication can promote the origin of evolutionary novelties [43].

### Acknowledgements

This work has been supported grants from the NIH and NSF to M.L. and by an NSF IGERT fellowship to V.K. We thank F. Kondrashov and A. Wagner for helpful comments.

### References

- Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag
- Force, A. *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
- Prince, V.E. and Pickett, F.B. (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3, 827–837
- Lynch, M. (2004) Gene duplication and evolution. In *Evolution: From Molecules to Ecosystems* (Moya, A. and Font, E., eds), pp. 33–47, Oxford University Press
- Piatgorsky, J. and Wistow, G. (1991) The recruitment of crystallins: new functions precede gene duplication. *Science* 252, 1078–1079
- Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B. Biol. Sci.* 256, 119–124
- Lynch, M. and Force, A. (2000) Gene duplication and the origin of interspecific genomic incompatibility. *Am. Nat.* 156, 590–605
- Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1154
- Lynch, M. and Conery, J.S. (2003) The evolutionary demography of duplicate genes. In *Genome Evolution* (Meyer, A. and Van de Peer, Y., eds), pp. 35–44, Kluwer Academic Publishers
- Li, W.-H. (1985) Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy. In *Population Genetics and Molecular Evolution* (Ohta, T. and Aoki, K., eds), pp. 333–352, Springer-Verlag
- Van de Peer, Y. *et al.* (2001) The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J. Mol. Evol.* 53, 436–446
- Kondrashov, F.A. *et al.* (2002) Selection in the evolution of gene duplications. *Genome Biol.* 3, RESEARCH0008
- Nembaware, V. *et al.* (2002) Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. *Genome Res.* 12, 1370–1376
- Seoighe, C. *et al.* (2003) Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. *Mol. Biol. Evol.* 20, 484–490
- Davis, J.C. and Petrov, D.A. (2004) Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol.* 2, E55
- Urrutia, A.O. and Hurst, L.D. (2003) The signature of selection mediated by expression on human genes. *Genome Res.* 13, 2260–2264
- Hughes, M.K. and Hughes, A.L. (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol. Biol. Evol.* 10, 1360–1369
- Zhang, L. *et al.* (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19, 1464–1473
- Zhang, P. *et al.* (2003) Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.* 4, R56
- Conant, G.C. and Wagner, A. (2003) Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13, 2052–2058
- Dermitzakis, E.T. and Clark, A.G. (2001) Differential selection after duplication in mammalian developmental genes. *Mol. Biol. Evol.* 18, 557–562
- Lynch, M. and Force, A. (2000) The probability of duplicate-gene preservation by subfunctionalization. *Genetics* 154, 459–473
- Hastings, K.E.M. (1996) Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. *J. Mol. Evol.* 42, 631–640
- Duret, L. and Mouchiroud, D. (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* 17, 68–74
- Zhang, L. and Li, W.-H. (2004) Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol. Biol. Evol.* 21, 236–239
- Merritt, T.J. and Quattro, J.M. (2001) Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics* 159, 689–697
- Kellis, M. *et al.* (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624
- Lynch, M. *et al.* (2001) The probability of fixation of a newly arisen gene duplicate. *Genetics* 159, 1789–1804
- Nowak, M.A. *et al.* (1997) Evolution of genetic redundancy. *Nature* 388, 167–171
- Wagner, A. (2000) The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* 154, 1389–1401
- Kondrashov, F.A. and Koonin, E.V. (2004) A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* 20, 287–290
- Haldane, J.B.S. (1937) The effect of variation on fitness. *Am. Nat.* 71, 337–349
- Lynch, M. and Conery, J.S. (2003) The origins of genome complexity. *Science* 302, 1401–1404
- Wagner, A. (2002) Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* 19, 1760–1768
- Papp, B. *et al.* (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429, 661–664
- Averof, M. *et al.* (1996) Diversification of arthropod Hox genes as a paradigm for the evolution of gene functions. *Semin Cell Dev. Biol.* 7, 539–551

- 37 Katju, V. and Lynch, M. (2003) The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* 165, 1793–1803
- 38 Spofford, J.B. (1969) Heterosis and the evolution of duplications. *Am. Nat.* 103, 407–432
- 39 Walsh, B. (2003) Population-genetic models of the fates of duplicate genes. *Genetica* 118, 279–294
- 40 Lenormand, T. *et al.* (1998) Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. *Evolution* 52, 1705–1712
- 41 Berticat, C. *et al.* (2002) Insecticide resistance genes induce a mating competition cost in *Culex pipiens* mosquitoes. *Genet. Res.* 79, 41–47
- 42 Gu, Z. *et al.* (2004) Duplicate genes increase gene expression diversity within and between species. *Nat. Genet.* 36, 577–579
- 43 Zhang, J. (2003) Parallel functional changes in the digestive RNases of ruminants and colobines by divergent amino acid substitutions. *Mol. Biol. Evol.* 20, 1310–1317
- 44 Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons

### Articles of interest in *Trends* and *Current Opinion* journals

#### Common logic of transcription factor and microRNA action

Oliver Hobert

*Trends in Biochemical Science*, 29, 462–468

#### Gene flow from GM plants – towards a more quantitative risk assessment

Guy M. Poppy

*Trends in Biotechnology*, 22, 436–438

#### Picking a winner: new mechanistic insights into the design of effective siRNAs

Delquin Gong and James E. Ferrell, Jr

*Trends in Biotechnology*, 22, 451–454

#### The alluring simplicity and complex reality of genetic rescue

David A. Tallmon, Gordon Luikart and Robin S. Waples

*Trends in Ecology and Evolution*, 19, 489–496

#### Human intestinal bacteria as reservoirs for antibiotic resistance genes

Abigail A. Salyers, Anamika Gupta and Yanping Wang

*Trends in Microbiology*, 12, 412–416

#### Getting closer to affective disorders: the role of CRH receptor systems

Marianne B. Müller and Wolfgang Wurst

*Trends in Molecular Medicine*, 10, 409–415

#### Id genes and proteins as promising targets in cancer therapy

Sylvia Fong, Robert J. Debs and Pierre-Yves Desprez

*Trends in Molecular Medicine*, 10, 387–392

#### Schizophrenia genetics: dysbindin under the microscope

Matthew A. Benson, Roy V. Sillitoe and Derek J. Blake

*Trends in Neurosciences*, 20, 355–358

#### Pharmacogenomics in cancer therapy: is host genome variability important?

William P. Petros and William E. Evans

*Trends in Pharmacological Sciences*, 25, 457–464

#### Development through the eyes of functional genomics

Andrew G. Fraser and Edward M. Marcotte

*Current Opinion in Genetics and Development*, 14, 336–342