*seminars in* **CELL & DEVELOPMENTAL BIOLOGY,** Vol 10, 1999: pp. 555–559
Article No. scdb.1999.0337, available online at http://www.idealibrary.com on IDEAL®

# Evolutionary preservation of redundant duplicated genes

*David C. Krakauer* and Martin A. Nowak*

*Gene duplication events produce both perfect and imperfect copies of genes. Perfect copies are said to be functionally redundant when knockout of one gene produces no 'scoreable', phenotypic effects. Preserving identical, duplicate copies of genes is problematic as all copies are prone to accumulate neutral mutations as pseudogenes, or more rarely, evolve into new genes with novel functions. We summarise theoretical treatments for the invasion and subsequent evolutionary modification of functionally redundant genes. We then consider the preservation of functionally identical copies of a gene over evolutionary time. We present several models for conserving redundancy: asymmetric mutation, asymmetric efficacy, pleiotropy, developmental buffering, allelic competition and regulatory asymmetries. In all cases, some form of symmetry breaking is required to maintain functional redundancy indefinitely.*

**Key words:** evolution of genetic redundancy / gene duplication / gene families / genome organization / mathematical model



## Introduction

'Symmetry is wearying'.[1]

THE IMMEDIATE CONSEQUENCE of the perfect duplication of an existing gene is the creation of an identical, redundant copy. Assuming that duplication mutations are rare, widespread genetic redundancy requires that genomes harbouring two or more redundant, paralogous genes, will have invaded the 'singleton', wild-type population. To become detectable, the paralogous genomes will have had to increase to non-negligible frequencies, and furthermore, remain stable across generations. The emergence and conservation of gene copies is therefore characterised by three principal difficulties. (1) *Error-free duplication of genes is rare*: probabilistically, it is more likely that copies will be imperfect, as mutations will have occurred during the duplication event. (2) *Invasion and fixation is slow*: assuming perfect replication, the copy will behave as a neutral allele and fix very slowly. (3) *Preserving redundancy from silencing or modification*: if the neutral copy is able to fix under drift and the copy remains functionally redundant, then it is likely to become transformed into a pseudogene under mutation pressure buffered by the original gene copy, or more rarely, evolve a novel function.

These represent some of the core theoretical problems associated with gene duplication and preserving redundant gene expression. Mathematical models have been used to address several of these issues, none of which are formally independent. We shall discuss the results of a small sample of models to illustrate these essential points. It is not our intention to present an exhaustive review, but a sample of theoretical efforts. First it is necessary to introduce some basic notation.

## Notation

Consider a haploid population of organisms. For our purposes, we consider only two loci and two alleles (the favoured, restricted universe of population genetics). The wild-type, functional allele is designated *A*. Upon duplication, it comes to occupy a second locus and is designated *B*. Duplication from a genome *A* to genome *AB* occurs at a rate *d* and loss of the copy at rate *e*. Deleterious mutations to either of these alleles produces the defective alleles, *a* and *b*. Mutation from *A* to *a* occurs at rate $u_a$. Mutation from *B* to *b* occurs at rate $u_b$. Rare advantageous mutations ($u_c$) to either allele produce the new functional allele *C*. Assuming a truly redundant copy, and a completely dominant functional allele, the frequency of the genotypes are $x_C$, $x_A$, $x_{AB}$, $x_{Ab}$, $x_{aB}$,

*From the* Institute for Advanced Study, Princeton, NJ 08540, USA. E-mail: krakauer@ias.edu/ nowak@ias.edu*

*1084-9521/ 99/ 050555+ 05 $30.00/ 0*

and $x_{ab}$. The fitnesses of the genotypes are ranked as: $w_{AC} > w_A = w_{AB} = w_{Ab} = w_{aB} > w_a = w_{ab} = 0$.

## Invasion and silencing of duplicate genes

Clark[2] presents models for the invasion of functionally redundant gene families ($AB$) into wild-type populations containing a single gene copy ($A$). Ignoring recombination and any effects of gene position and assuming weak selection, Clark was able to demonstrate that a duplicated genome, at a constant risk of losing the copy ($e > 0$), would invade the wild-type assuming that $u_{a,b} > 0$ and $d > 0$. Thus, recurrent duplication, when combined with the loss of gene function, is sufficient to ensure the invasion of a functionally redundant gene. In the absence of recurrent mutation for duplication and complete dominance of the functional duplicate, the gene family can still fix in the population by drift, albeit more slowly. If one allows that fitness is not completely dominant, then invasion occurs only if there is a fitness advantage for duplication $w_{AB} > w_{Ab}$. Clark also showed that in small populations the duplication is more likely to invade. This is because mutation and drift become more important and selection less important as population size declines and the effects of stochasticity increase.

Several studies have attempted to estimate the time period required to silence copies once they have invaded. A very general result is that this time increases as a function of the effective population size ($N_e$). In effect, the silencing of the copy can be treated in the same way as the loss of a neutral allele.[3–5] Things become a little more complex when selection becomes important, such as through incomplete dominance or a cost of additional gene copies.[6] In general terms, genes are silenced more rapidly in these models.

Ohta[7,8] and Walsh[9] address the contrasting question, how often do the duplicated genes evolve new functions, rather than become silenced? Starting from a population with two identical copies, Walsh derives the probability that a new gene evolves from a duplicate, rather than becoming a pseudogene, as $u_c = [(1 - e^{-s})/\rho S + 1]^{-1}$. Where $S = 4N_e s$ (a measure of the intensity of selection corrected for population size), $\rho = u_c/u_b$ (the advantageous mutation rate to neutral mutation rate ratio) and $s$ is the selection coefficient. The take home message from this result is that the majority of genes will become silent even when $S \gg 1$. Thus, even strong selection in favour of advantageous alleles can not ensure that duplicate genes are not silenced in the long run. Only when $\rho \gg S^{-1}$ can new genes emerge. Verbally, the successful emergence of new genes requires that selection be very efficient and that advantageous alleles emerge relatively frequently (see later discussion on regulatory asymmetries for an alternative possibility).

Ohta[7] has demonstrated another interesting feature of duplication, that redundant expression can increase the rate at which compensatory mutations to a duplicate gene fix within a population. This is effective when $2N_e u < 1$. Thus, if the mutation rate is not too high, the transiently redundant gene is able to buffer mutations arising at the duplicate locus while awaiting the substitution of a compensatory allele in the mutant gene.

Gene silencing and the evolution of novel gene function have been the focus of the greater part of population genetics models of gene duplication. In the remainder of this article we will consider the null case, the preservation of identical, redundant copies of a gene. As Gould and Eldredge[10] insist, 'stasis is data'.

## Conserving redundant genes

The absence of a modified or 'scoreable' phenotype, following gene knock-out studies, has alerted biologists to the presence of genes with overlapping functions. These include, among numerous others,[11] the *Drosophila* gene pairs *gooseberry* and *sloppy paired*,[12] the mouse *tenascin*[13] and Hox genes[14] and yeast *myosin* genes.[15] Brookfield[16] cautions that experimental procedures remain insufficiently precise to demonstrate true redundancy (functional identity), while Tautz[17] proposes that we define redundancy as a measure of correlated, rather than degenerate, function. This allows genes to engage in independent activities, while remaining redundant with respect to a shared function. The problems for redundancy remain, as one needs to account for the persistence of the correlations. There are also very interesting cases of functional redundancy rarely alluded to in this context, such as the X-chromosome, which in the homogametic phase (XX), lives alongside a redundant duplicate. This seems to be made possible by random inactivation of one X chromosome in the homogametic phase and dominant expression in the heterogametic phase. There is also the issue of codon

redundancy within the genetic code, which varies across taxonomic groups, and requires preservation of an unnecessarily large class of tRNAs.

It is possible that obligate, functional redundancy is only apparent, as a result of recent duplication and reduplication events. However, the ancient origin of many redundant genes suggests otherwise.[18,19] Thus several mechanisms for preserving redundancy have been proposed. These include a cumulative benefit from gene copy number (dosage effects), increased fidelity of multiple copies, structural constraints such that genes with independent activities and overlapping function remain redundant through selection on their independent functions (pleiotropy),[20] and convergent functions emerging from common structures.[21]

## Mutational fidelity and functional efficacy

The case of redundancy serving to increase gene product, does not pose any real theoretical difficulties. One simply asserts that more gene product is better and allow that positive selection fixes copies while purifying selection preserves them. Obvious examples are the multiple copies of rRNA genes, and most dramatically, entire copies of thousands of mitochondrial and chloroplast genomes resident in numerous cells.[22]

If one interprets increased fidelity to mean reduced mutational load, a more sophisticated reasoning is required.[23,24] Consider two identical genes with slightly different mutation rates ($u_a < u_b$), such as might arise from adopting a telomeric rather than centromeric position along a chromosome. One finds that the gene with the higher mutation rate ($B$) is eliminated by selection. If mutation rates are equal, one of two genes will eventually become silenced by drift. Thus, the simplest form of redundancy, as Fisher first noted,[25] remains unstable. By allowing a slight asymmetry in each of the gene's abilities to perform their function, redundancy can be preserved (Figure 1a). The requirement is that the less efficacious gene also possesses the lower rate of mutation ($w_A < w_B$, $u_A < u_B$). A possible scenario is when duplication renders a gene less expressible, perhaps due to local chromatin structure. The dual effect will be to reduce the selective value of the duplicate and to reduce the mutation rate (assuming the level of expression is positively correlated with mutation rate).

Some forms of pleiotropy can also ensure the conservation of redundant function.[23] Consider again two genes, $A$ and $B$ with two independent functions,

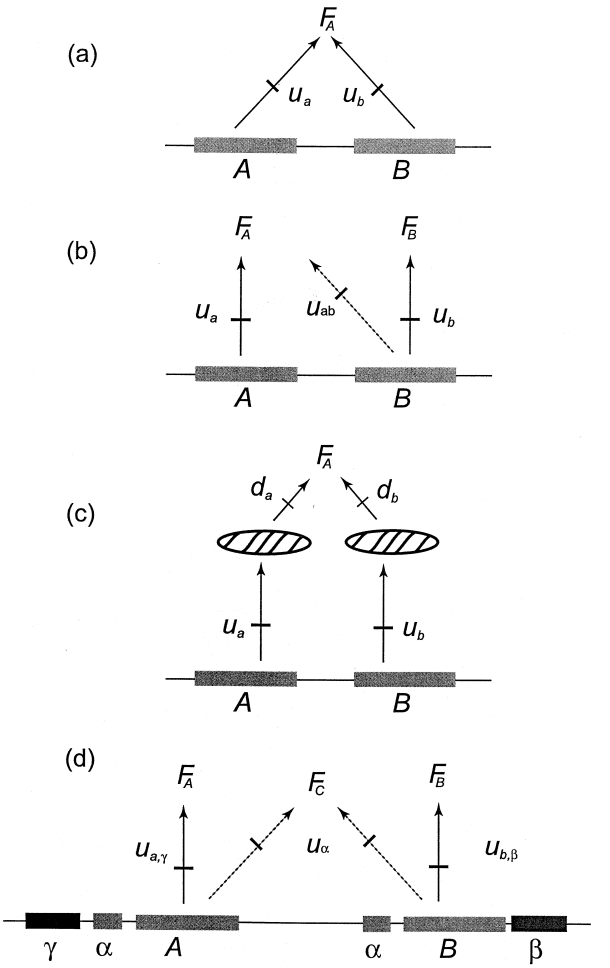


**Figure 1.** The two identical genes $A$ and $B$ have arisen through a gene duplication event. Deleterious mutations to either of these alleles produces the defective alleles, $a$ and $b$. Mutation from $A$ to $a$ occurs at rate $u_a$. Mutation from $B$ to $b$ occurs at rate $u_b$. (a) *Unstable genetic redundancy.* Both duplicates contribute to the single function $F_A$. When $u_a > u_b$, $B$ is silenced by mutations. When $u_a > u_b$ one of either $A$ or $B$ is eventually lost by drift. (b) *Pleiotropy can preserve redundant function.* $A$ and $B$ perform independent functions $F_A$ and $F_B$, while $B$ also performs $F_A$ but less effectively. Redundancy requires that $u_a > u_{ab}$. (c) *Error buffering during development and ontogeny.* Genes $A$ and $B$ encode a developmental programme related to function $F_A$. Development can produce errors in the phenotype relating to this function. Redundancy requires that developmental fidelity compensate for mutational error in the developmental programme or *vice versa*. Thus, for example, redundancy requires that $u_A < d_B$ and $u_b < d_A$. (d) *Redundancy through cistron duplication.* $\alpha$, $\beta$ and $\gamma$ are regulatory elements. Upon duplication these elements are parcelled out unevenly to $A$ and $B$. $A$ and $B$ only overlap with respect to regulation through element $\alpha$. Independent expression of $A$ and $B$ through their unique regulatory elements preserves the gene and can promote adaptive diversification rather than silencing. Redundancy is maintained for long periods as $\alpha$ is a relatively small mutational target.

$F_A$ and $F_B$. Assume that $A$ performs $F_A$ with efficacy $E_A$, while $B$ performs $F_B$ with efficacy $E_B$ ($E_B = 1$); in addition $B$ performs function $F_A$ with efficacy $E_{AB}$ ($E_{AB} < E_B$). Redundancy is partial and measured in terms of correlated function. When mutations to gene $B$ can either eliminate its unique function ($u_b$) or both functions ($u_{ba}$), then redundancy is preserved whenever $u_{ba} < u_a$. In other words, when $B$'s pleiotropic function is more robust than $A$'s unique function, the correlated function can be preserved (Figure 1b).

## Error-buffering in ontogeny

While mutations are a very important source of error as they are heritable, developmental errors or errors transpiring during maturity can also select for redundancy. An appropriate analogy is that of the duplicate flight-systems employed in aircraft design or an external storage device used to back-up data from a personal computer hard-disk. The recurrent risk of error during the life time of a mechanism selects for noise-buffering. Considering our two genes $A$ and $B$, we now introduce the additional error probabilities, $d_a$ and $d_b$, to denote the probability that the two gene-products are rendered deficient during ontogeny. It can be shown that to preserve a sizeable frequency of both genes within the population ($0 < x_{AB} < 1$), we require that $u_A < d_B$ and $u_B < d_A$ (Figure 1c). This is an intuitive result. It states that each gene must mutate less frequently than its duplicate experiences an ontogenetic defect. If this were not so, selection would not be able to 'detect' those errors arising during development, and hence through evolution, put in place mechanisms for buffering. This insight holds true for multiple, redundant copies of a single gene. Specifically $n$ redundant genes can be conserved when $n < 1 + \ln u / \ln d$.[23]

Wagner[26] has emphasised this error-buffering function of redundancy by exploring the evolution of redundancy in a quantitative model. Unlike the models thus far, which assume a fixed degree of redundancy and explore its persistence, Wagner considers how redundancy can evolve through time by treating redundancy as a continuous variable. It is found that redundancy can increase when mutation rates are high and population sizes are sufficiently large, $N_e u > 1$ (this is in keeping with the population genetics models). Reasonable convergence rates towards redundant function requires $N_e u \gg 1$. This implies that redundancy is more likely to evolve in large populations with reasonably high mutation rates. When this is not the case ($N_e u < 1$), redundancy decays to silence under the influence of genetic drift. It is stressed that selection does not operate directly on redundant genes, which after all have no phenotype, but on the reduced number of offspring spawned by non-redundant, and hence potentially defective, parents.

## Multidomains and regulators

Throughout the preceding discussion the models assume that duplicated alleles function independently rather than interactively. There are examples where point mutations to a duplicated gene are selectively more harmful than the knock-out of the entire gene, for example key mutations to the Src-type protein kinases.[27] This result has been ascribed to negative interactions among alleles forming multi-domain proteins,[28] or to put it differently, competition among variant alleles making up a protein (allelic competition). In support of this position are the larger number of multi-gene families in vertebrates (as compared to invertebrates), thought to correlate positively with an increased abundance of multi-domain proteins. Thus, redundancy is preserved simply because $w_{AB} > \{w_{aB}, w_{Ab}, w_{ab}\}$. While this seems like a plausible hypothesis, it remains necessary to establish quantitatively the degree of irreversible association between wild-type and mutant alleles and those making up the protein domains. This is because mutant duplicates are also likely to have a lower affinity for the target domains than the wild-types.

When we have spoken of genes we have not distinguished coding regions from regulatory elements. From the perspective of genetic redundancy, however, the difference may be rather important. Thus, for example, redundancy maintained through allelic competition is not problematic, if mutations to regulatory elements completely inactivate duplicates. A compelling role for regulatory elements in promoting divergence and partial redundancy among duplicates has been proposed by Averof and Akam.[29] Averof and Akam consider not only the duplication of the coding region of a gene, to give genes $A$ and $B$, but also duplication of several regulatory elements associated with this gene ($\alpha, \beta, \gamma$). If the two duplicate genes are each accompanied by a subset of the original elements (A: $\alpha, \gamma$; B: $\alpha, \beta$), in which each set overlaps to some degree (in this case through the shared element $\alpha$), the redundant genes can be maintained by selection acting through their unique

patterns of expression and the shared regulatory element(s) (Figure 1d). If one assumes that the shared regulatory element is a smaller mutational target than the coding region ($u_\alpha < \{u_a, u_b\}$), this will prolong the half-life of the redundant function. This effect is not sufficient, however, to prevent one or more shared elements from becoming silenced in the long term. The virtue of the process is that it might help explain why it is so many Hox genes, derived from duplication events, have acquired novel functions rather than become silenced.[30] To preserve redundancy indefinitely we would require, as with the pleiotropic model, some asymmetry in mutation and/or efficacy of the regulators.

## Conclusions

A common feature of almost all models for conserving redundancy is the requirement for some degree of symmetry-breaking. The strict equality among duplicate genes must be disrupted in order that they be preserved. To maintain functional equality one must posit mutational inequality, while mutational equality necessitates divergent efficacy. Thus, in the very simplest models, identical genes must mutate at different rates and perform with different efficacies. In models including pleiotropy, overlapping functions are asymmetrical with respect to robustness; this applies to both regulatory elements and coding regions of genes. When allowing for developmental and ontogenetic errors in duplicate genes, redundancy only persists when developmental accuracy is accompanied by reduced mutational fidelity.

## References


1. Humboldt A (1850) Cosmos. Harper and Brothers, New York
2. Clark AG (1994) Invasion and maintenance of a gene duplication. Proc Natl Acad Sci USA 91:2950–2954
3. Kimura (1983) The Neutral Theory of Molecular Evolution. CUP, Cambridge
4. Watterson GA (1983) On the time for gene silencing at duplicate loci. Genetics 105:745–766
5. Li W (1980) Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics 95:237–258
6. Allendorf FW (1979) Rapid loss of duplicate gene expression by natural selection. Heredity 43:247–258
7. Ohta T (1987) Simulating evolution by gene duplication. Genetics 115:207–213
8. Ohta T (1988) Further simulation studies on the evolution by gene duplication. Evolution 42:375–386
9. Walsh JB (1995) How often do duplicated genes evolve new functions? Genetics 139:421–428
10. Gould SJ, Eldredge N (1996) Punctuated equilibrium comes of age. Nature
11. Erickson HP (1993) Gene Knockouts of c-src, transforming growth factor Beta-1 and Tenascin suggest superfluous, nonfunctional expression of proteins. J Cell Biol 120:1079–1081
12. Cadigan KM, Grossniklaus U, Gehring WJ (1994) Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. Proc Natl Acad Sci USA 91:6324–6328
13. Saga Y, Yagi T, Ikawa Y, Sakakura T, Shinichi A (1992) Mice develop normally without tenascin. Genes Dev 6:1821–1831
14. Condie BG, Capecchi MR (1995) Mice targeted disruptions in the paralogous Hox genes Hoxa-3 and Hoxd-3 reveal genetic interactions. J Cell Biochem Suppl O:A1–109
15. Goodson HV, Spudich JA (1995) Identification and molecular characterization of a yeast myosin I. Cell motility and the cytoskeleton 30:73–84
16. Brookfield J (1992) Can genes be truly redundant? Curr Biol 2:553–554
17. Tautz F (1992) Redundancies, development and the flow of information. BioEssays 14:263–266
18. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713
19. Spring J (1997) Hypothesis: vertebrate evolution by interspecific hybridisation—are we polyploid? FEBS Lett 400:2–8
20. Thomas JH (1993) Thinking about genetic redundancy. Trends Genet 9:395–399
21. Pickett FB, Meeks-Wagner DR (1995) Seeing double: appreciating genetic redundancy. Plant Cell 7:1347–1356
22. Gillham NW (1994) Organelle Genes and Genomes. OUP, Oxford
23. Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J (1997) Evolution of genetic redundancy. Nature 388:167–171
24. Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 13:360–364
25. Fisher RA (1935) The sheltering of lethals. Am Nat 69:446–455
26. Wagner A (1999) Redundant gene functions and natural selection. J Evol Biol 12:1–16
27. Lowell CA, Soriano P (1996) Knockouts of Src-family kinases: stiff bones, wimpy T cells and bad memories. Genes Dev 10:1845–1857
28. Gibson TJ, Spring J (1998) Genetic redundancy in vertebrates: polyploidy in the common ancestor and persistence of genes encoding multidomain proteins. Trends Genet 14:46–49
29. Averof M, Akam M (1995) Hox genes and the diversification of insect and crustacean body plans. Nature 376:420–423
30. Averof M, Dawes R, Ferrier D (1996) Diversification of arthropod Hox genes as a paradigm for the evolution of gene functions. Semin Cell Dev Biol 7:539–551