

**The role of population size, pleiotropy, and fitness effects of mutations  
in the evolution of overlapping gene functions.**

Andreas Wagner

The University of New Mexico, Department of Biology

and

The Santa Fe Institute

Address:

University of New Mexico, Department of Biology, 167A Castetter Hall,  
Albuquerque, NM 87131-1091, U.S.A.

Phone: +1-505-277-2021; FAX: +1-505-277-0304;

E-mail: [wagnera@unm.edu](mailto:wagnera@unm.edu)

## Abstract

Sheltered from deleterious mutations, genes with overlapping or partially redundant functions may be important sources of novel gene functions. While most partially redundant genes originated in gene duplications, it is much less clear why genes with overlapping functions have been retained, in some cases for hundreds of millions of years. A case in point are the many partially redundant genes in vertebrates, the result of ancient gene duplications in primitive chordates. Their persistence and ubiquity become surprising when considering that duplicate and original gene often diversify very rapidly, especially if the action of natural selection is involved. Are overlapping gene functions perhaps maintained because of their protective role against otherwise deleterious mutations? There are two principal objections against this hypothesis, which are the main subject of this paper. First, because overlapping gene functions are maintained in populations by a slow process of “second order” selection, population sizes need to be very high for this process to be effective. It is shown that even in small populations, pleiotropic mutations that affect more than one of a gene’s function simultaneously can slow the mutational decay of functional overlap after a gene duplication by orders of magnitude. Furthermore, brief and transient increases in population size may be sufficient to maintain functional overlap. The second objection regards the fact that most naturally occurring mutations may have much weaker fitness effects than the rather drastic “knock-out” mutations that lead to detection of partially redundant functions. Given weak fitness effects of most mutations, is selection for the buffering effect of functional overlap strong enough to compensate for the diversifying force exerted by mutations? It is shown that the extent of functional overlap maintained in a population is not only independent of the mutation rate, but also of the average fitness effects of mutation. These results are discussed with respect to experimental evidence on redundant genes in organismal development.

## Introduction

Overlapping or partially redundant gene functions are ubiquitous in eukaryotes. Their presence spans the entire taxonomic range from unicellular organisms to mammals. They are observed in proteins of any function, be it enzymatic, structural, or regulatory (Basson *et al.*, 1986; Cadigan *et al.*, 1994; Condie and Capecchi, 1995; Goldstein, 1993; González-Gaitán *et al.*, 1994; Goodson, 1995; Hanks *et al.*, 1995; Higashijima *et al.*, 1992; Hoffman, 1991; Li and Noll, 1994ab; Lundgren *et al.*, 1991). Weak or no phenotypic effects of a knock-out mutation in a gene thought to play an important biological role are often the first indicators of such redundancy. Further analysis then typically reveals one or more related genes with functions similar to that of the mutated gene. With some potential exceptions, partially overlapping gene functions can be traced to (often old) gene duplications.

While the evolutionary origins of most partially redundant gene functions are obvious, it is much less obvious why genes with overlapping functions have been retained, in some cases for hundreds of millions of years. This holds especially for many partially redundant genes in vertebrates, the result of ancient gene duplications in primitive chordates (Bailey *et al.*, 1997, Sharman and Holland, 1996). This is even more puzzling given that duplicate and original genes can diversify very rapidly after duplication (Cirera and Aguade, 1998; Tsaur *et al.*, 1998; Zhang *et al.*, 1998), especially if the action of natural selection is involved. Thus, are overlapping gene functions merely the transient remnants of past gene duplication events, a snapshot of a diversification process that will eventually be complete? Or are they perhaps maintained because of their protective role against otherwise deleterious mutations? These questions are attracting increasing interest (Cooke *et al.*, 1997; Tautz, 1992; Thomas, 1993), and for good reasons. Sheltered from some deleterious mutations, partially redundant genes may be important sources of evolutionary novelties on the biochemical level.

This paper deals with the two perhaps most serious objections to the persistence of overlapping gene functions as a buffer against mutation. The first objection concerns the population size required to stably maintain overlapping gene functions in a population. Even neutral mutations in genes with overlapping functions are likely to lead to diversification of the gene's functions. Counteracting this evolutionary force is a much slower

process of “second order” selection for maintaining overlap (Wagner, 1999), a process that acts on differences in functional overlap among genes. These differences are selectively neutral, but genes with greatly overlapping functions are less likely to undergo deleterious mutations than genes with distinct functions. Over time, genes with greater functional overlap will be preferentially retained in a population. Because of the indirect nature of this process, populations have to be large to sustain the amount of variation in overlap on which selection can act. A past contribution has shown that the required population sizes are not impossibly large (Wagner, 1999). Here, it is shown that even in small populations, pleiotropic mutations that affect more than one of a gene’s function simultaneously, can delay functional divergence by orders of magnitude. Moreover, in contrast to the maintenance of genetic variation in fluctuating populations, the evolution of functional overlap may not be predominantly influenced by occasional population bottlenecks, but by occasional population size bursts.

The second potential objection regards the fact that most naturally occurring mutations are likely to have much weaker fitness effects than the rather drastic “knock-out” mutations that lead to detection of partially redundant functions in the laboratory. Given weak fitness effects of many mutations, is the selection process outlined above strong enough to compensate for the diversifying forces exerted by mutations? It is shown that the evolution of mean fitness and functional overlap are effectively decoupled. This implies that the average effect of deleterious mutations on fitness does not greatly influence the evolution of functional overlap and vice versa. This also has implications on the genetic load associated with overlapping gene functions, as discussed below.

To address the above questions, this contribution utilizes a conceptually simple mathematical model which relies on three main assumptions. (Wagner, 1999). First, the greater the functional overlap among genes, the greater the fraction of mutations in either gene that are phenotypically neutral. Second, both neutral and non-lethal deleterious mutations cause gene functions to diverge. A third assumption, already implicit in the above discussion, is that overlap among gene functions is a quantifiable variable, at least in principle.

While the first two assumptions are unproblematic, the third requires further comment. Two examples will illustrate its validity. Overlapping gene functions are especially

prominent among regulatory developmental genes, because these genes often have multiple functions. (Most housekeeping genes, many of which catalyze specific enzymatic reactions, may be restricted in their ability to evolve functional overlap. The specificity of their function may be part of the reason why housekeeping genes are rarer among the documented cases of functional overlap.) Consider the case of a transcription factor (Fig. 1a), regulating the expression of multiple genes involved in a developmental process. Transcription factors occupy a prominent role in the many examples of genes with overlapping functions (Joyner *et al.*, 1991; Maconochie *et al.*, 1996; Weintraub, 1993). After duplication (Fig. 1b), and diversification (Fig. 1c), the functional overlap among duplicate and original can be viewed as the subset of genes regulated by both factors. Micro array technology (Schena, 1996) has made the quantification of such overlaps a realistic goal: perturb the activity of each factor in separate experiments, measure the resulting change in gene expression patterns, and assess which subset of genes is affected similarly. A second example concerns genes with identical biochemical functions but spatiotemporal expression patterns that are not completely congruent (Hanks *et al.*, 1995; Li and Noll, 1994ab). Here, the extent of spatiotemporal overlap in two expression domains can be taken as a measure of functional overlap. Recently developed automated image processing technology to quantify the concentration of multiple gene products in a developing embryo at single cell resolution enables such quantification (Kosman *et al.*, 1998).

### Model and Results

The model is concerned with a population of haploid, randomly mating organisms with non-overlapping generations. The assumption of haploidy is chosen merely because it exposes most clearly the key evolutionary principles at work. The genic mutation rate is denoted as  $\mu$  ( $\mu \ll 1$ ). Neutral mutations that do not affect any aspect of the function of a gene product are not considered.

The central concept of the model is the notion of partially redundant or overlapping gene functions, and that the overlap in two gene's functions can be quantified. Specifically, let the variable  $r$ ,  $r \leq 0 \leq 1$ , denote a measure of the functional overlap of two genes. If  $r = 1$ , then two genes have completely identical functions; if  $r = 0$ , there is no overlap in the gene's functions. If two genes have overlapping functions, then some mutations

that eliminate one gene's function will be neutral because the function is covered by the other gene. This is conceptualized in the following way: If  $r = 1$  (identical functions), all mutations are neutral; if  $r = 0$  no mutation is neutral; and if  $0 < r < 1$  the rate of neutral mutations is some function  $f(r)$ . The simplest case is that of a linear relation, in which the rate of neutral mutations per two genes is given by  $2\mu r$  (neglecting terms of order  $\mu^2$ ). While this linear relation will be used throughout most of this section, the consequences of relaxing this assumption will be explored.

Mutations can not only affect fitness, they may also change the functional overlap among two genes, which will in general lead to a divergence in gene functions. This is modeled by a conditional probability density  $m_r(r^*|r)$ , which denotes the probability that the functional overlap  $r^*$  after mutation of two genes with overlap  $r$  before mutation lies in the interval  $(r^*, r^* + dr^*)$ . To leave the formalism as general as possible, only the minimal assumption is made that mutation reduces  $r$  on average by a factor  $\lambda_r$ , i.e.,

$$\int_0^1 r^* m(r^*|r) dr^* = \lambda_r r, \quad (1)$$

where  $0 < \lambda_r < 1$ . Because a value of  $\lambda_r = 0.5$  means that each mutation reduces  $r$  by 1/2 on average (a very rapid divergence rate),  $\lambda_r > 0.5$  will be used here. Complete loss-of-function mutations could be modeled as an important special case, but are not considered here, because their evolutionary dynamics has been extensively modeled by others (Kimura and King, 1979; Marshall *et al.*, 1994; Maruyama and Takahata, 1981; Nei and Roychoudhury, 1973; Ohno, 1970; Ohta, 1987; Watterson, 1983; Walsh, 1995). Also, empirical data suggests that the complete elimination of one gene's function may be much less frequent than commonly assumed (Allendorf *et al.*, 1975, Ferris and Whitt, 1976, 1979; Nadeau and Sankoff, 1997).

**Summary of previous results.** Because results of a previous contribution (Wagner, 1999) will be used below, they will be briefly reviewed here. In this earlier paper, it was assumed that all non-neutral mutations, occurring at a rate of  $2\mu(1-r)$  per two genes with overlap  $r$ , are effectively lethal. The evolution of the distribution of functional overlap,  $p_t(r)$ , in a population was studied under this assumption. The recurrence equation

$$p_{t+1}(r) = \frac{(1 - 2\mu)p_t(r) + 2\mu \int_0^1 zp_t(z)m_r(r|z)dz}{(1 - 2\mu) + 2\mu \int_0^1 zp_t(z)dz} \quad (2)$$

describes the distribution  $p_t(r)$  of  $r$  in generation  $t$  of a large population ( $N\mu > 50$ ; Wagner, 1999). It can be shown that the mean functional overlap among two genes  $\bar{r} = \int_0^1 rp(r)dr$  reaches a non-zero mutation-selection equilibrium independent of the initial condition and approximated by

$$\bar{r}_\infty = \sqrt{\frac{\sigma_{mr}^2 \lambda_r}{2\lambda_r^3 - 3\lambda_r^2 + 1}}. \quad (3)$$

Here,  $\sigma_{mr}^2$  is the variance of the effect of mutations on  $r$ , defined as  $\sigma_{mr}^2 = \int_0^1 (r^* - \lambda_r r)^2 m_r(r^*|r) dr^*$ . This approximation holds for values of  $\bar{r}_\infty$  in the interior of  $(0, 1)$ , requires no further assumptions about the distribution of mutational effects, and is independent of the mutation rate. It is in good agreement with numerical results (Wagner, 1999). When extending this model to  $n$  genes with overlapping functions, it was found that (i) the extent of overlap maintained in mutation-selection balance is statistically indistinguishable from that for the two-gene case, and (ii) linkage relations are not likely to favor or disfavor the evolution of overlap. These results motivate the restriction to the two-gene case with tight linkage analyzed below.

For small populations  $N\mu \ll 1$ , it was found that functional overlap diminishes (functions diverge) at a rate approximated by the equation

$$\frac{d\langle r_t \rangle}{dt} = \ln \lambda_r \frac{2\mu \langle r_t \rangle^2}{(1 - 2\mu) + 2\mu \langle r_t \rangle}, \quad (4)$$

Here,  $\langle r_t \rangle$  denotes the mean functional overlap in an ensemble of small populations. In the absence of selection, mutations (1) would lead to an exponential decline in  $\langle r_t \rangle$  from an initial value of one immediately after gene duplication. The most important feature of (4) is that the decay in  $r$  is much slower than that: The lower the functional overlap between two genes, the lower the rate at which functions continue to diverge. The reason is that most mutations that affect  $r$  are neutral for large  $r$  but deleterious for small  $r$ . This leads to a reduction in  $r$  following a polynomial rather than an exponential rate.

In sum, if all non-neutral mutations have severely deleterious effects, then (i) functional overlap and thus an increased rate of non-deleterious mutations can be maintained among genes in large populations, (ii) functions diverge, albeit very slowly, in small populations.

**Evolution of functional overlap and of mean fitness are decoupled.** In this paper, the assumption of lethality of non-neutral mutations will be relaxed. Specifically, fitness of an organism  $w$  is interpreted as a probability of survival, and non-neutral mutations in an organism with fitness  $w$  shall reduce fitness by a factor  $\lambda_w$ .

$$\int_0^1 w^* m(w^*|w) dw^* = \lambda_w w, \quad (5)$$

This formula, completely analogous to (1), allows for an increase in fitness provided that the variance of mutational effects on fitness  $\sigma_{m_w}^2 = \int_0^1 (w^* - \lambda_w w)^2 m_w(w^*|w) dw^*$  is large, but will also ensure that mutations reduce mean fitness on average by a factor  $\lambda_w$  if  $\lambda_w < 1$ . Of interest is the evolution of the joint distribution of functional overlap and fitness  $p_t(r, w)$ , as well as the moments  $\overline{r^i w^j}_t = \int_0^1 \int_0^1 r^i w^j p_t(r, w) dr dw$ . In a large population,  $p_t(r, w)$  will evolve according to

$$\begin{aligned} p_{t+1}(r, w) = & (1/D_t) [(1 - 2\mu) w p_t(r, w) + 2\mu w \int_0^1 z_r p_t(z_r, w) m_r(r|z_r) dz_r \\ & + 2\mu w \int_0^1 \int_0^1 (1 - z_r) p_t(z_r, z_w) m_w(w|z_w) m_r(r|z_r) dz_r dz_w] \end{aligned} \quad (6)$$

The denominator

$$D_t = (1 - 2\mu) \overline{w}_t + 2\mu \overline{r w}_t + 2\mu \lambda_w (\overline{w}_t - \overline{r w}_t) \quad (7)$$

is a normalization factor representing the fraction of individuals surviving from one generation to the next. In the numerator,  $(1 - 2\mu) w p_t(r, w)$  is the fraction of individuals that do not undergo mutation in generation  $t$  and that survive into the next generation.  $2\mu w \int_0^1 z_r p_t(z_r, w) m_r(r|z_r) dz_r$  is the fraction of individuals with fitness  $w$  that undergo a neutral mutation changing  $r$  from some value  $z_r$  to  $r$  and that survive into the next generation. Finally,  $2\mu w \int_0^1 \int_0^1 (1 - z_r) p_t(z_r, z_w) m_w(w|z_w) m_r(r|z_r) dz_r dz_w$  is the fraction of individuals that undergo a mutation affecting both fitness ( $z_w \rightarrow w$ ) and functional overlap ( $z_r \rightarrow r$ ) and that survive into the next generation. In both these terms, the factor  $w$  in front of the integral represents the effect of selection, i.e., the probability that an individual of fitness  $w$  survives into the next generation. Implicit in these equations are the assumptions that functional overlap influences the probability of a deleterious mutation, but not its severity, and that mutations occur before selection.

This rather formidable equation, while perhaps too complicated to solve analytically, can be used to obtain a rough estimate for the correlation  $\rho(r, w)$  between fitness and overlap in mutation-selection balance. The estimate, derived in the Appendix, suggests that this correlation is small, i.e.,

$$\rho(r, w) = \frac{\text{Cov}(r, w)}{\sqrt{\text{Var}(w)\text{Var}(r)}} \text{ is of order } \sqrt{\mu} \quad (8)$$

This crude approximation (see Appendix) is fully supported by numerical results, exemplified in Figs. 2 and 3. Fig. 2 shows that both in cases where mutations have mildly (Fig. 2ab;  $\lambda_w = 0.9$ ) and severely deleterious effects (Fig. 2cd;  $\lambda_w = 0.3$ ), the correlation between  $r$  and  $\mu$  is of order  $\sqrt{\mu}$ . This holds even in spite of the high mutation rates of  $10^{-2}$  that were used for reasons of computational feasibility. Thus, in mutation-selection balance, there is only a weak correlation between fitness and functional overlap among genes. Moreover, the mean overlap observed both for strongly (Fig. 2c) and weakly (Fig. 2a) deleterious mutations is statistically indistinguishable from the theoretical prediction (3) for the case where all mutations are lethal. Also, mean functional overlap at equilibrium is independent of the mutation rate when this rate is varied over two orders of magnitude (Fig. 3b). Further, the difference between  $\bar{w}$  and one is of the order  $\mu$  regardless of the values of  $\lambda_r$  (Fig. 2cd) or  $\lambda_w$  (Fig. 3a). In sum, functional overlap at equilibrium is independent of the mutation rate, and of the severity of mutational effects on fitness ( $\lambda_w$ ). Conversely, mean equilibrium fitness is not sensitive to changes in the rate at which mutations lead to a decay of redundancy ( $\lambda_r$ ). In this sense, evolution of redundancy and fitness are decoupled. The reasons for this are discussed below.

**Transient population size bursts may be sufficient to maintain redundancy.**

While the maintenance of functional overlap by means of natural selection requires large populations ( $N\mu \gg 1$ ), overlap may decay relatively slowly in small populations (Wagner, 1999). This prompts a question as to what fraction of time a population of fluctuating size has to spend in a large  $N\mu$  regime for overlap to be maintained by selection. It might appear that the concept of effective population size (Hartl and Clark, 1997) from population genetic theory holds the answer to this question. For populations that spend a fraction  $f$  of their time in a regime of small  $N\mu$  by having a small population size  $N_S$ , and a fraction  $1 - f$  in a regime of large  $N\mu$  by having a large  $N_L$ , one calculates the

(inbreeding) effective population size as

$$N_e = \frac{N_S N_L}{f N_L + N_S(1 - f)}. \quad (9)$$

A population of fluctuating size with a given  $N_e$  should then behave in exactly the same manner as a population of constant size  $N = N_e$ . Numerical results from Figure 4 indicate that this is not the case for the evolutionary process studied here, because fluctuating and constant populations with the same  $N_e$  show different mean overlap.

Figure 5 shows more extensive numerical results for the evolution of overlap under fluctuating  $N\mu$ , where a population spends alternatingly  $t_S$  generations in a regime of small  $N\mu$ , in which redundancy will decay under the influence of drift, and  $t_L$  generations in a regime of large  $N\mu$ , where selection will maintain functional overlap. The figure shows mean functional overlap in a population ensemble after  $10^5$  generations, as a function of the ratio  $t_S/t_L$  of time spent in the small  $N\mu$  regime. For comparison, a dashed line shows  $N_e\mu$  as calculated via (9), where  $f = t_S/(t_L + t_S)$ . Notice that  $N_e\mu$  is smaller than one for all the values of  $t_S$  and  $t_L$ , such that drift should dominate the dynamics of overlap (Fig. 7 in Wagner, 1999). However, for many of the values of  $t_S$  and  $t_L$  in Figure 5 populations maintain levels of overlap indistinguishable from or close to those in mutation-selection balance for infinite populations, despite the fact that they spend only a small fraction of time in the large  $N\mu$  regime. Whereas the absolute amount of time  $t_L$  spent per cycle in the large  $N\mu$  regime is clearly important in determining how much overlap is maintained, it is equally clear from the figure that substantial overlap is maintained even for transient bursts in  $N\mu$ . For instance, Fig. 5a shows that functional overlap close to that observed in mutation-selection equilibrium (infinite population size) is maintained if the population spends only one-hundredth of its time, or 100 out of 10000 generations in the large  $N\mu$  regime. A caveat to these results is that computational limitations in simulating large populations necessitated modulating  $\mu$  rather than  $N$  to simulate changes in  $N\mu$ . Note also that the specific values of  $t_S$  and  $t_L$  necessary to maintain overlap will depend on multiple factors, such as the location of the mutation-selection equilibrium, the variance in overlap introduced by mutation, and probably even the selection regime (soft selection vs. hard selection; Hartl and Clark, 1997). Unfortunately, an analytical approach to this problem, which could both circumvent computational limitations and account for these

factors, is currently not at hand.

Why is  $N_e$  not a reliable indicator for the evolution of overlap (Figure 4) in fluctuating populations? Even in the absence of analytical predictions for the change of  $r$  in fluctuating populations, it is safe to say that this must have to do with the different evolutionary dynamics of genetic variation and functional overlap. Genetic variation gets lost much faster in small populations than in large populations. We surmise that this is not true in the case of functional overlap. Whereas the amount of genetic variation maintained in a fluctuating population may be most heavily influenced by the population bottlenecks, the amount of functional overlap may be most heavily influenced by the largest population sizes.

**Pleiotropic effects slow down divergence of gene functions.** Among the many possible relations of functional overlap  $r$  and the probability  $f(r)$  that a mutation leading to a loss of one or more functions in one gene is phenotypically neutral, only  $f(r) = r$  has been explored so far. The following scenario illustrates the role that pleiotropic mutations may have in this relation and motivates a more general mathematical form for  $f$ . Consider duplicate and original of a gene shortly after a gene duplication, where both copies carry out each of  $k$  functions. These could be  $k$  spatiotemporal expression domains, in which each of the two genes are involved in a developmental process, or  $k$  downstream genes regulated by each of two genes encoding a transcription factor (see Fig. 1). Assume that, over time, mutations randomly eliminate the capability to carry out some of these functions in each of the genes, such that only a fraction  $r = k_r/k$  of the original  $k$  functions is still performed by both genes. In this context, mutations with pleiotropic effect can be viewed as those that eliminate more than one, say  $l$ , functions. If  $k$  is sufficiently large, then the probability that such a pleiotropic mutation is neutral is approximated by  $r^l$ . As a consequence, for a given overlap  $r$ , the probability  $f(r)$  that a mutation eliminating  $l$  functions is neutral is given by  $f(r) = r^l$ . In other words, the more extensive pleiotropic effects are, the smaller the protective effect against mutations afforded by a given overlap  $r$ . Only in the absence of pleiotropic effects does the simple relation  $r = f(r)$  hold. In practice, mutations will span a range of pleiotropic effects, suggesting that  $f(r) < r$ , under the constraints that  $f(1) = 1$  and  $f(0) = 0$ . Below, the effect on the evolution of redundancy of a moderate degree of pleiotropy ( $l = 2, 3$ ) via the relation  $f(r) = r^l$  will be explored (Fig. 6a).

The evolution of  $\langle r_t \rangle$ , the mean functional overlap in a population ensemble evolving under the influence of genetic drift, can be approximated by an ordinary differential equation

$$\frac{d\langle r_t \rangle}{dt} = \ln \lambda_r \frac{2\mu \langle r_t \rangle^{1+l}}{(1-2\mu) + 2\mu \langle r_t \rangle^l}, \quad (10)$$

which is a simple extension of (4) to  $l \geq 1$ . Here, one unit of time corresponds to one (discrete) generation, and  $l \geq 1$  is the exponent in the function  $f(r) = r^l$ . A solution to this equation can be obtained as an implicit function by a separation of variables, yielding

$$\left( \frac{1}{\ln \lambda_r} \right) \left[ \left( \frac{1-2\mu}{2\mu} \right) \left( \frac{1}{l} \right) \left( 1 - \frac{1}{\langle r_t \rangle^l} \right) + \ln \langle r_t \rangle \right] - t = 0$$

for the initial condition  $\langle r_0 \rangle = 1$ . From this form of the solution, one can directly determine the time  $\tau_x$  necessary for  $\langle r \rangle$  to decrease from one shortly after a gene duplication to a value of  $x$ . Let us consider the special case of the times  $t_k = \tau_{(1/2)^k}$  necessary to reduce  $\langle r_t \rangle$  to  $(1/2)^k$ . One can show that

$$t_{k+1} - t_k = - \left( \frac{1}{\ln \lambda_r} \right) \left[ \left( \frac{1-2\mu}{2\mu} \right) \left( \frac{2^{lk}}{l} \right) (2^l - 1) + \ln 2 \right] \propto \frac{2^{lk}(2^l - 1)}{l}. \quad (11)$$

The times  $t_{k+1} - t_k$  necessary for successive halving of  $\langle r_t \rangle$  not only scale as  $2^k$ , i.e., each consecutive halving of functional overlap takes twice as long, they also scale as  $2^l$ . Thus, how fast functional overlap will decay depends critically on the extent of pleiotropy of mutations. Figure 6b shows the decrease of functional overlap from  $\langle r_0 \rangle = 1$  obtained from a numerical solution of (10) (solid lines) and from Monte Carlo simulations (dots and bars) for realistic mutation rates ( $\mu = 2.5 \times 10^{-6}$ ) and small populations ( $N = 100$ ). The decrease in mean ensemble overlap clearly slows down as the extent of pleiotropy increases. However, over longer evolutionary time scales than those depicted here, differences among the different degrees of pleiotropy are more dramatic than Fig. 6b suggests. Figure 6c shows, on a linear-log scale, the times  $\tau_{1/2}$ ,  $\tau_{1/4}$ , and  $\tau_{1/8}$  obtained from (11) as a function of  $l$ . It shows that, (i) for an increase of  $l$  from one to only three, each of these times increases by a factor 10 to 100, and (ii) for  $l = 3$ , the time to decrease  $\langle r \rangle$  by each additional factor of 1/2 increases by approximately a factor 10.

In sum, even moderate degrees of pleiotropy lead to a severe constraint on how much genetic drift can reduce functional overlap among genes. The intuitive explanation is

simple: Mutations with pleiotropic effects are more and more likely to eliminate non-redundant functions as the functional overlap between two genes decreases. Because such mutations are deleterious and are eliminated from the population, the decay of overlap via neutral mutations and drift is slowed down. If most mutations have effects on many of a gene's function one may observe significant functional overlap (congruent expression patterns etc.) maintained over long evolutionary time scales. This is because most mutations will also affect the few unique functions of a gene and thus be deleterious. In this case, extensive functional overlap may not indicate effective buffering against mutations.

### Discussion

The model explored here rests on three simple assumptions: (i) genes have functions that overlap in a quantifiable way; (ii) overlapping functions affect the probability of mutations being neutral; (iii) mutations, on average, reduce functional overlap. As in any mathematical model, simplifications and abstractions were made. A haploid system has been modeled to illustrate the issues discussed below most clearly.

**Purifying and directional selection.** The model does not incorporate advantageous mutations leading to novel functions. Instead, its restriction to purifying selection and neutral mutations reflects an implicit assumption that genes diversify mostly by (i) loss of some of their functions and, thus, functional specialization, or (ii) acquisition of new functions by neutral mutations. Whether this is the case may well depend on the type of genes considered. For instance, genes involved in self-recognition or reproductive functions are more likely to be subject to positive selection for diversification (Cirera and Aguadé, 1998; Ting *et al.*, 1998; Tsauro *et al.*, 1998). On the other hand, many of the most striking examples of partial redundancies come from genes embedded in highly conserved developmental pathways, such as that of muscle determination (Weintraub 1993). In these cases, there may be a limited potential for the evolution of novel functions, such that divergence by specialization may be prevalent (Force *et al.*, 1999). Corroborating evidence comes from experiments where a gene from an invertebrate (before duplication) can substitute for the function of one of its duplication products in very distantly related vertebrates, or vice versa. If the vertebrate gene had evolved radically new functions, such substitution might not be possible. Examples include *nautilus* and *decapentaplegic* from *Drosophila*

*melanogaster* and their respective counterparts MyoD and BMP in mammals (Padgett *et al.*, 1993, Venuti *et al.*, 1991). However, the restriction to divergence by neutral mutations should not be read as an exclusive endorsement of one evolutionary scenario. It is rather an acknowledgment that this scenario is relevant for many developmental genes.

**Mutational load and deleterious mutations.** For the model studied here, mean fitness and mean functional overlap among genes are only weakly correlated at mutation-selection balance (Fig. 2). The reason is the following. If organisms in a population have different fitnesses, selection acts immediately on these differences, e.g., via differential survival probabilities. In contrast, if some organisms in a population have genes with lower functional overlap than other organisms, these differences are selectively neutral *until* mutations occur at these genes, which takes of the order of  $1/\mu$  generations per individual. In this case, the genes with the higher functional overlap  $r$  are less likely to undergo deleterious mutations (which are then rapidly eliminated by selection), and they therefore accumulate slowly in the population. Thus, one might think of the evolution of mean fitness and of mean overlap as occurring on different time scales, and this is what causes them to be effectively decoupled.

As a consequence, mean functional overlap is independent of the fitness effects of deleterious mutations (Fig. 2). Whether deleterious mutations have only moderate fitness effects, or whether they are invariably lethal, they are eliminated at a rate much faster than that at which the evolution of overlap occurs. Conversely, mean fitness at equilibrium differs from one only by an amount that is of the order of the mutation rate, as would be expected from a haploid two-locus model with recurrent mutations. This is because overlap influences only the rate of deleterious mutations  $(1 - r)\mu$ . Whether  $r$  is large or small, the reduction in mean fitness from 1 is thus still of the order of  $\mu$ . An important corollary is that overlapping gene functions are not maintained because they convey a higher mean fitness on a population with greater functional overlap. In this sense maintenance of functional overlap is not an adaptive phenomenon (Futuyma, 1998).

**Evolution of diploidy and redundant gene functions.** A phenomenon superficially related to the evolution of redundancy is the evolution of diploidy. Diploids are able to “mask” the fitness effects of deleterious alleles, which may have been an important reason why diploidy evolved (Crow and Kimura, 1965; Kondrashov and Crow, 1965; Maynard

Smith, 1978; Otto and Goldstein, 1992; Perrot *et al.*, 1992). On the other hand, diploids carry a higher mutational load (mean fitness reduction due to deleterious mutations) than haploids, because they carry twice as many alleles (e.g., Crow and Kimura, 1965). The higher mutational load counteracts the effects of masking, raising the question of which of these two forces wins out in the evolution of ploidy levels. The question has been studied by various authors, mostly in the framework of modifier theory (Bengtsson, 1992; Otto and Goldstein, 1992; Perrot *et al.*, 1991). The answer depends on details of parameters such as recombination rates and degree of dominance.

There are key differences in the questions studied here from those at issue in existing models of the evolution of ploidy levels. First, the increase in frequency of original gene and duplicate (corresponding to the rise of diploidy from haploid organisms) is not at issue here. Fixation of gene duplications must be a ubiquitous phenomenon, because a vast number of known proteins fall into a small number of gene families. Theoretical work suggests that such fixation is easily accomplished by neutral evolution (even without any selective advantage through “masking”) provided that duplications occur at a finite rate (Clark, 1994). At issue is rather the question if and how functional overlap is maintained, once established. A second distinction stems from the notion of deleterious mutations. It is best illustrated with the example of a set of genes regulated by one transcription factor whose gene has undergone gene duplication and subsequent loss of some functions between original and duplicate (Fig. 1). As long as all genes are properly regulated by at least one of the two factors (Fig. 1c), no fitness disadvantage will result. However, from the perspective of a classical population genetic model, all transcription factor genes that have lost the ability to regulate at least one gene in the set would harbor “deleterious” mutations, because if the allele complementing their missing functions (at a different locus in the case of duplicate genes, at the same locus on a sister chromosome in the case of diploidy) did not occur in the same organism, the organism would have reduced fitness. This creates the somewhat paradoxical situation where an entire gene pool may consist of “deleterious” alleles that nevertheless complement each other functionally, so that no organism with reduced fitness exists in the population. Thus, where multifunctional proteins may undergo specialization after gene duplications via loss of some of their functions, the notion of intrinsically deleterious mutations may not be a natural one. It may be more expedient

to only call a mutation deleterious if it actually affects fitness. However, even if the notion of “intrinsically” deleterious mutations is abandoned, the question remains whether overlapping gene functions influence mean fitness (mutational load) in mutation-selection balance. The answer is yes, but to no significant extent. In the simple case where all deleterious mutations are lethal, mean equilibrium fitness is  $1 - 2\mu$  if two genes have completely diverged in function; it is  $1 - 2\mu(1 - \bar{r})$  if mean equilibrium overlap is equal to  $\bar{r}$ . Thus, differences in mutational load are small. This result will not be greatly affected if deleterious mutations have less severe effects, because mean equilibrium fitness is still of order  $1 - \mu$ .

Decoupling of the evolution of overlap and redundancy shows that  $\lambda_r$  and its associated standard deviation of mutational effects  $\sigma_{mr}$  are the only parameters determining the evolution of functional overlap. Although functional overlap can be measured, as discussed above, such measurements are not available, and thus it is difficult to estimate these parameters from experimental data. As far as  $\sigma_{mr}$  is concerned, it is perhaps most important that *some* mutations must increase functional overlap among genes in order for selection to sustain overlap (Wagner, 1999). Ample indirect evidence that this is the case comes from the many examples of functionally convergent proteins in various organisms (Doolittle, 1994). Measurement of  $\lambda_r$  is hampered by two additional factors, the significant stochastic component in the evolution of overlap (see below), and its dependency on the nature of the genes considered.

**Population size fluctuations and redundancy.** Natural populations differ from the constructs of theoretical population genetics in many ways, e.g., mating does not occur at random, individuals show a clumped spatial distribution, and populations fluctuate in size. Such deviations can be dealt with effectively by calculating an effective population size  $N_e$  according to standard formulae (Hartl and Clark, 1997). A recent survey on the relation of  $N_e$  and  $N$  in a large number of vertebrate and invertebrate populations suggests that fluctuations in population size are by far the most important factor influencing  $N_e$  (Frankham, 1995).

Population size is a key factor in determining whether selection can maintain functional overlap among genes. Only if the influence of drift is weak, i.e., in populations with large size  $N\mu$ , can overlap be maintained indefinitely, or even be built up from disjoint

functions (Wagner, 1999). Perhaps surprisingly,  $N_e$  does not seem to be sufficient to predict the dynamics of overlap in fluctuating populations (Fig. 4). With some caveats due to computational limitations and the absence of analytical predictions, the results shown in Fig. 5 suggest that short spikes (e.g., 50-100 generations every 10,000 generations) in  $N\mu$  may sometimes be sufficient to sustain high levels of functional overlap. The reason for the inadequacy of  $N_e$  to predict the dynamics of overlap has to do with the different evolutionary dynamics of genetic variation (for which  $N_e$  was designed) and functional overlap. Whereas genetic variation gets lost much faster in small populations than in large populations, this may not be true for functional overlap, which may be lost slowly in small populations. It would clearly be desirable to have a measure of effective population size suitable to predict the decay of redundancy in fluctuating populations.

Population sizes with  $N\mu \gg 1$  are realistic for most microorganisms and some invertebrates. However, even in higher vertebrates, census population sizes are sometimes well within this realm, although population sizes may undergo frequent bottlenecks. For example, a study by Avise *et al.* (1988) reports a discrepancy between current population sizes and historical population sizes in three vertebrate species including catfish (*Arius felis* and redwing blackbirds (*Agelaius phoeniceus*). Census population sizes suggest a number of  $10^7$  and  $2 \times 10^7$  breeding females, respectively. However, data on mitochondrial DNA variation suggest that past population sizes must have been substantially smaller (Avise *et al.*, 1988). However, even if large population sizes are sustained over only short periods of time, functional overlap among genes might be maintained in such populations. More data on population size history would be required to determine whether population bursts are frequent enough even in organisms normally characterized by small  $N$ .

An observation unrelated to the significance of population fluctuations can be made from the results shown here. “Error” bars shown in Figure 5 represent the standard deviation of functional overlap among populations within a population ensemble. Since these populations spend a significant amount of time in the small  $N\mu$  regime, they are monomorphic most of the time. This means that the standard deviation shown is a good measure of differences in functional overlap among populations evolving in parallel. These differences are obviously large (Fig. 5), which is due to variation in the number of mutations reducing overlap that go to fixation. In some populations, several such mutations may

have gone to fixation, in others only few. This holds not only for populations evolving in parallel, but would also apply to multiple gene pairs evolving in parallel in one lineage. The importance of stochastic effects may help explain why some developmental genes duplicated early during chordate evolution have preserved greater functional overlap than others. Examples might include members of the MyoD gene family (Weintraub, 1993), where significant overlap in functions has been maintained, versus members of the *hedgehog* gene family, which have almost completely diverged in expression pattern (Bitgood and McMahon, 1995).

**Pleiotropy slows the decay of overlap in small populations.** A large fraction of an organism's genomic gene content may be expressed during the development of each organ system (Datta *et al.*, 1993; Thaker and Kankel, 1992). This suggests that each developmental gene participates in many developmental processes, in line with evidence from many well studied genes which are expressed during more than one developmental stage. Such developmental multifunctionality is complemented by biochemical multifunctionality of proteins. Transcription factors exemplify this well (Fig.1). If each gene regulated by a transcription factor is viewed as one function of that factor, then most transcription factors are multifunctional. The abundance of multifunctionality in regulatory genes, both developmentally and biochemically, suggests that pleiotropic effects of mutations should be pervasive, as Wright (1968) already foresaw. In other words, most mutations may affect more than one function of a developmental gene. Now what does this have to do with the evolution of functional overlap?

Envision a scenario such as that shown in Figure 1, where a transcription factor regulating the expression of multiple genes undergoes gene duplication, and subsequently the original and the copy lose some of their functions. In small populations, functional overlap will decay according to (4). The key feature of this equation is that the reduction of overlap proceeds more and more slowly as more and more functions get lost. This is because as the number of functions not covered by the other genes increases, the likelihood that a loss-of-function mutation affects a unique function and is deleterious increases as well. This slowing down in evolution is quite drastic. Pleiotropic mutational effects contribute further to this phenomenon. A simple argument presented above suggests that if a mutation affects  $l$  functions at a time, then the probability of a mutation in one of two genes with overlap  $r$

being neutral is of order  $r^l$ . Thus, if pleiotropic mutations affect a sufficiently large number  $l$  of functions, then mutations will often have deleterious effects even if  $r$  is large. This should lead to a further slowing down of the decay of redundancy, an intuition confirmed by analytical and simulation results (Fig. 6). As the number of functions affected by a mutation increases linearly, the time until overlap decays from one to a given value increases exponentially (Fig. 6c).

Knock-out mutations – a severe kind of genetic perturbation – of many developmental genes in vertebrates have weak phenotypic effects. This suggests that significant functional overlap has been maintained among the genes in question, many of which are the product of ancient gene duplications having occurred sometime before the origin of tetrapods approximately 400 million years ago (Wray *et al.*, 1996). Such preservation of functional overlap is not the “default” fate of duplicated genes, as several recent studies on the rapid functional diversification of duplicated genes show (Cirera and Aguade, 1998; Tsaur *et al.*, 1998; Zhang *et al.*, 1998). Thus, it is likely that some evolutionary force is responsible for its preservation. Natural selection is likely to play an important role in this process, either actively maintaining overlap in large populations, or passively delaying its decay in small ones. Even if selection only fulfills this second role, its importance must not be underestimated given the vast amount of time that has elapsed since these duplications occurred.

## Appendix

**An order-of-magnitude estimation for the correlation of fitness and functional overlap.** One can derive from (6) a recurrence equation for  $\bar{r}_t$  and solve for  $\bar{r}$  at equilibrium by setting  $\bar{r}_t = \bar{r}_{t+1}$ . This yields

$$\bar{r} = \frac{(1 - 2\mu)\bar{r}\bar{w} + 2\mu(\lambda_r\overline{r^2w}) + 2\mu\lambda_r\lambda_w(\bar{r}\bar{w} - \overline{r^2w})}{(1 - 2\mu)\bar{w} + 2\mu\bar{r}\bar{w} + 2\mu\lambda_w(\bar{w} - \bar{r}\bar{w})}, \quad (\text{A1})$$

which can be rewritten in the following way

$$\text{Cov}(r, w) = \overline{rw} - \bar{r}\bar{w} = \frac{2\mu}{1 - 2\mu} \left[ \overline{rw\bar{r}} + \lambda_w(\bar{w} - \bar{r}\bar{w}) - \lambda_r\overline{r^2w} - \lambda_r\lambda_w(\bar{r}\bar{w} - \overline{r^2w}) \right] \quad (\text{A2})$$

Each of the terms in the square brackets on the right hand side is of order unity or less (Note that  $1 > \int_0^1 \int_0^1 r^i w^j p(r, w) \rightarrow 0$  as  $i, j \rightarrow \infty$ .) It follows that the covariance of  $r$  and  $w$  in mutation-selection balance is of order of the mutation rate, i.e., very small. By calculating the equilibrium value of  $\overline{r^k}$  in the same way, it can be shown that at mutation selection equilibrium  $\text{Cov}(r^k, w) = \overline{r^k w} - \overline{r^k} \bar{w}$  is of order  $\mu$  for all  $k > 1$ . In other words,  $\overline{r^k w} \approx \overline{r^k} \bar{w}$ . Substituting this approximation into (A1),  $\bar{w}$  and  $\mu$  cancel out, and one easily obtains

$$\bar{r} \approx \frac{\lambda_r \overline{r^2} + \lambda_r \lambda_w (\bar{r} - \overline{r^2})}{\bar{r} + \lambda_w (1 - \bar{r})} \quad (\text{A3})$$

This can be rewritten to yield  $\text{Var}(r) > \lambda_r \overline{r^2} - \bar{r}^2 = \lambda_w (1 - \lambda_r) \bar{r} / (1 - \lambda_w)$ .

Using (6) to establish an equation for the mean fitness  $\bar{w}$  at equilibrium, one can show that the variance of  $w$  is given by

$$\text{Var}(w) = \overline{w^2} - \bar{w}^2 = \frac{2\mu}{1 - \mu} [(1 - \lambda_w) \overline{rw} - \lambda_w^2 \overline{r^2 w^2} + \lambda_w \bar{w} - \lambda_w^2 \bar{w}^2 + \sigma_{mw}^2 (\bar{r} - 1)] \quad (\text{A4})$$

Summarizing, both  $\text{Cov}(r, w)$  and  $\text{Var}(w)$  are of order  $\mu$ , whereas the above form for  $\text{Var}(r)$  suggests that  $\text{Var}(r)$  is of order unity. Taken together, this indicates that

$$\rho(r, w) = \frac{\text{Cov}(r, w)}{\sqrt{\text{Var}(w)\text{Var}(r)}} \quad \text{is of order} \quad \sqrt{\mu} \quad (\text{A5})$$

## Literature Cited

- Allendorf, F.W., Utter, F.M. and May, B.P. (1975) *Gene duplication within the family Salmonidae. II. Detection and determination of the genetic control of duplicate loci through inheritance studies and the examination of populations.* p 415-432. In: Markert, C.L. (ed.) *Isozymes IV: Genetics and Evolution.* Academic Press, New York.
- Avise, J.C., Ball, R.M., Arnold, J. (1988) *Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations.* *Mol. Biol. Evol.* **5**, 331-344
- Bailey, W.J., Kim, J., Wagner, G.P., Ruddle, F.H. (1997) *Phylogenetic reconstruction of vertebrate Hox cluster duplications.* *Mol. Biol. Evol.* **14**, 843-853
- Basson, M.E., Thorsness, M., Rine, J. (1986) *Saccharomyces cerevisiae contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase.* *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5563-5567
- Bengtsson, B.O. (1992) *Deleterious mutations and the origin of the meiotic ploidy cycle.* *Genetics* **131**, 741-744
- Bitgood, M.J., McMahon, A.P. (1995) *Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo.* *Dev. Biol.* **172**, 126-138
- Cadigan, K.M., Grossniklaus, U., Gehring, W.J. (1994) *Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation.* *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6324-6328
- Cirera, S., Aguade, M. (1998) *Molecular evolution of a duplication: the sex-peptide (Acp70A) gene region of Drosophila subobscura and Drosophila madeirensis.* *Mol. Biol. Evol.* **15**, 988-996
- Clark, A.G. (1994) *Invasion and maintenance of a gene duplication.* *Proc. Natl. Acad. Sci.* **91**, 2950-2954
- Condie, B.G., Capecchi, M.R. (1995) *Mice with targeted disruptions in the paralogous Hox genes Hoxa-3 and Hoxd-3 reveal genetic interactions.* *J. of Cellular Biochemistry Supplement O*, A7-109

- Cooke, J., Nowak, M.A., Boerlijst, M., Maynard-Smith, J. (1997) *Evolutionary origins and maintenance of redundant gene expression during metazoan development*. Trends. Genet. **13**,360-364
- Crow, J.F., Kimura, M. (1965) *Evolution in sexual and asexual populations*. Am. Nat. **99**, 439-450
- Datta, S., Stark, K., Kankel, D.R. (1993) *Enhancer detector analysis of the extent of genomic involvement in nervous system development in Drosophila melanogaster*. J. Neurobiol **24**, 824-841
- Doolittle, R.F. (1994) *Convergent evolution: the need to be explicit*. Trends in Biochemical Sciences **19**, 15-18
- Emery, P., Strubin, M., Hofmann, K., Bucher, P., Mach, B., Reith, W. (1996) *A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity*. Mol. Cell. Biol. **16**, 4486-4494
- Ferris, S.D. and Whitt, G.S. (1976) *Loss of duplicate gene expression after polyploidization*. Nature **265**, 258-260
- Ferris, S.D. and Whitt, G.S. (1979) *Evolution of the differential regulation of duplicate genes after polyploidization*. Journal of Molecular Evolution **12**, 267-317
- Firulli, A.B., Olson, E.N. (1997) *Modular regulation of muscle gene transcription: a mechanism for muscle cell diversity*. Trends. Genet. **13**, 364-36
- Force A., Lynch, M., Pickett, F.B, Amores, A., Yan, Y.L., Postlethwait, J. (1999) *Preservation of duplicate genes by complementary, degenerative mutations*. Genetics **151**, 1531-1545
- Frankham, R. (1995) *Effective population size/adult population size ratios in wildlife: a review*. Genet. Res. Camb. **66**, 95-107
- Futuyma, D. (1998) *Evolutionary biology*. 3rd ed. Sinauer, Massachusetts.
- Goldstein, L.S.B. (1993) *Functional redundancy in mitotic force generation*. J. Cell. Biol. **120**, 1-3
- González-Gaitán E., M., Rothe, M., Wimmer, E.A., Taubert, H., Jäckle, H. *Redundant*

- functions of the genes knirps and knirps-related for the establishment of anterior Drosophila head structures.* Proc. natl. Acad. Sci. U.S.A. **91**, 8567-8571
- Goodson, H.V., Spudich, J.A. (1995) *Identification and molecular characterization of a yeast myosin I.* Cell Motility and the Cytoskeleton **30**, 73-84
- Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A., Joyner, A.L. (1995) *Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2.* Science **269**, 679-682
- Hartl, D.L., Clark, A.G. (1997) *Principles of population genetics.* (3rd ed.) Sinauer; Sunderland, Massachusetts
- Higashijima, S., Michiue, T., Emori, Y., Saigo, K. (1992) *Subtype determination of Drosophila embryonic external sensory organs by redundant homeo box genes BarH1 and BarH2.* Genes & development **6**, 1005-1018
- Hoffmann, F.M. (1991) *Drosophila abl and genetic redundancy in signal transduction.* Trends in genetics **7**, 351-355
- Joyner, A.L., Herrup, K., Auerbach, B.A., Davis, C.A., Rossant, J. (1991) *Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the en-2 homeobox.* Science **251**, 1239-1243
- Kimura, M., King, J.L. (1979) *Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift.* Proc. Natl. Acad. Sci. U.S.A. **76**, 2858-2861
- Kosman, D., Reinitz, J., Sharp, D.H. (1998) *Automated assay of gene expression at cellular resolution.* In: Altman, R.B., Dunker, K., Hunter, L., Klein, T.E. *Pacific Symposium on Biocomputing 1998.* World Scientific, Singapore
- Kirchhamer, K., Yuh, C.-H., Davidson, E.H. (1996) *Modular cis-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples.* Proc. Natl. Acad. Sci **93**, 9322-9328
- Li, W.-H. (1997) *Molecular evolution.* Sinauer, Sunderland, MA.
- Li, X., Noll, M. (1994a) *Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo.* EMBO J. **13**, 400-406

- Li, X., Noll, M. (1994b) *Evolution of distinct developmental functions of three Drosophila genes by acquisition of different cis-regulatory regions*. Nature **367**, 83-87
- Lundgren, K., Walworth, N., Booher, R., Dembski, M., Kirschner, M., Beach, D. (1991) *mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2*. Cell **64**, 1111-1112
- Maconochie, M., Nonchev, S., Morrison, A., Krumlauf, R. (1996) *Paralogous Hox genes: function and regulation*. Annu. Rev. Genet. **30**, 529-556
- Marshall, C.R., Raff, E.C., Raff, R.A. (1994) *Dollo's law and the death and resurrection of genes*. Proc. Natl. Acad. Sci. U.S.A. **91**, 12283-12287
- Maruyama, T., Takahata, N. (1981) *Numerical studies of the frequency trajectories in the process of fixation of null genes at duplicated loci*. Heredity **46**, 49-57
- Maynard Smith, J. (1978) *The evolution of sex*. Cambridge University Press, New York.
- Nadeau, J.H. and Sankoff, D. (1997) *Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution*. Genetics **147**, 1259-1266
- Nei, M., Roychoudhury, A.K. (1973) *Probability of fixation of nonfunctional genes at duplicate loci*. Am. Nat. **107**, 362-372
- Ohno, S. (1970) *Evolution by gene duplication*. Springer, New York
- Ohta, T. (1987) *Simulating evolution by gene duplication*. Genetics **115**, 207-213
- Otto, S.P., Goldstein, D.B. (1992) *Recombination and the evolution of diploidy*. Genetics **131**, 745-751
- Padgett, R.W., Wozney, J.M., Gelbart, W.M. (1993) *Human BMP sequences can confer normal dorsal-ventral patterning in the Drosophila embryo*. Proc. Natl. Acad. Sci. U. S. A. **90**, 2905-2909
- Perrot, V., Richerd, S., Valero, M. (1991) *Transition from haploidy to diploidy*. Nature **351**, 315-317
- Press, W.H., Teukolsky, S.A., Vetterling, W.A., Flannery, B.P. (1992) *Numerical Recipes in C*. Cambridge University Press, New York.
- Ptashne, M., Gann, A. (1997) *Transcriptional activation by recruitment*. Nature **386**,

569-577

- Schena, M. (1996) *Genome analysis with gene expression microarrays*. *Bioessays* **18**, 427-431
- Sharman, A.C., Holland, P.W.H. (1997) *Conservation, duplication, and divergence of developmental genes during chordate evolution*. *Netherlands J. Zool.* **46**, 47-67
- Suckow, M., Madan, A., Kisters-Woike, B., von Wilcken-Bergmann, B., Muller-Hill, B. (1994) *Creating new DNA binding specificities in the yeast transcriptional activator GCN4 by combining selected amino acid substitutions*. *Nucleic Acids Res.* **22**, 2198-2208
- Sun, Y., Jan, L.Y., Jan, Y.N. (1998) *Transcriptional regulation of atonal during development of the Drosophila peripheral nervous system*. *Development* **18**, 3731-3740
- Tautz, D. (1992) *Redundancies, development and the flow of information*. *BioEssays* **14**, 263-266
- Thaker, H.M., Kankel, D.R. (1992) *Mosaic analysis gives an estimate of the extent of genomic involvement in the development of the visual system in Drosophila melanogaster*. *Genetics* **131**, 883-894
- Thomas, J.H. (1993) *Thinking about genetic redundancy*. *Trends in Genetics* **9**, 395-398
- Ting, C.T., Tsaur, S.C., Wu, M.L., Wu, C.I. (1998) *A rapidly evolving homeobox at the site of a hybrid sterility gene*. *Science* **282**, 1501-1504
- Tsaur, S.C., Ting, C.T., Wu, C.I. (1998) *Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of Drosophila: II. Divergence versus polymorphism*. *Mol. Biol. Evol.* **15**, 1040-1046
- van Leeuwen, H.C., Strating, M.J., Cox, M., Kaptein, R., van der Vliet, P.C. (1996) *Mutation of the Oct-1 POU-specific recognition helix leads to altered DNA binding and influences enhancement of adenovirus DNA replication*. *Nucleic Acids Res.* **23**, 3189-3197
- Venuti, J. M., Goldberg, L., Chakraborty, T., Olson, E.N., Klein, W.H. (1991) *A myogenic factor from sea urchin embryos capable of programming muscle differentiation in mammalian cells*. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6219-6223
- Wagner, A. (1999) *Redundant gene functions and natural selection*. *J. Evol. Biol.* **12**,

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

Watterson, G.A. (1983) *On the time for gene silencing at duplicate loci.* Genetics **105**, 745-766

Weintraub, H. (1993) *The MyoD family and myogenesis: redundancy, networks, and thresholds.* Cell **75**, 1241-1244

Wray, G.A., Levinton, J.S., Shapiro, L.H. (1996) *Molecular evidence for deep precambrian divergences among metazoan phyla.* Science **274**, 568-573

Wright, S. (1968). *Evolution and the genetics of populations.* Chicago, University of Chicago Press.

Zhang, J., Rosenberg, H.F., Nei, M. (1998) *Positive Darwinian selection after gene duplication in primate ribonuclease genes.* Proc. Natl. Acad. Sci. U. S. A. **95**, 3708-3713

## Figure captions

**Fig. 1.** (a) A group of genes regulated by a transcription factor (TF). (b) After duplication of the transcription factor gene, the same genes are now regulated by both original and duplicate. (c) Over time, each factor may lose the ability to regulate some of the genes, such that only a subset of genes (cross-hatched) are regulated by both factors.

**Fig. 2. Mean fitness evolution and mean functional overlap in mutation-selection balance.** (a) Mean functional overlap at mutation-selection equilibrium for  $\lambda_w = 0.9$ , i.e., non-neutral mutations reduce fitness on average by 10 percent. Dots and bars represent mean fitness and one standard deviation as obtained from Monte-Carlo simulations. Mean equilibrium overlap scales linearly with  $\sigma_{mr}/(1 - \lambda_r)$ , where  $\lambda_r$  is the mean reduction in functional overlap caused by mutation, and  $\sigma_{mr}$  is the standard deviation of this reduction. The solid line represents the analytical prediction (3) for  $\lambda_w = 0$ , i.e., where all non-neutral mutations are lethal. Notice that equilibrium overlap maintained is statistically indistinguishable from that in the lethal case (solid line). (b) One minus mean fitness (dots and bars) and correlation between fitness and functional overlap (solid line) in mutation-selection equilibrium. The dashed line represents  $\sqrt{\mu}$ . Notice that despite the high mutation rate  $\mu = 10^{-2}$  used here for reasons of computational feasibility, both correlation and deviation of mean fitness from one are still of order  $\sqrt{\mu}$ , as estimated analytically. (c) and (d), identical to (a) and (b), respectively, but for more severe effects of mutations on fitness ( $\lambda_w = 0.3$ ).  $N = 5 \times 10^4$ ,  $\sigma_{mw} = 0.05$ ,  $10^4$  generations.

**Fig. 3.** (a) Dependency of mean fitness at mutation-selection equilibrium, plotted as  $1 - \bar{w}$ , on  $\lambda_w$ , the average factor by which mutations reduce fitness.  $N = 5 \times 10^4$ ,  $\sigma_{mw} = 0.05$ ,  $10^4$  generations simulated. Dots represent population means, bars one standard deviation. (b) Dependency of mean functional overlap at mutation-selection equilibrium on the mutation rate. Note the logarithmic scale for the mutation rate. Computational feasibility alone motivated the choice of range for  $\mu$ . Note that even for high mutation rates, equilibrium overlap is statistically indistinguishable from the analytical prediction (3), represented by the solid horizontal line.  $\lambda_w = \lambda_r = 0.9$ ,  $\sigma_{mr} = 0.05$ ,  $\sigma_{mw} = 0.01$ ,  $N = 500/\mu$ , generations simulated.

**Fig. 4. Population size fluctuations and evolution of overlap.** Shown are mean and standard error of overlap in an ensemble of 50 populations after 5000 generations of evolution. The simulation was started at a mean overlap of 0.28 which is the mean overlap at mutation-selection balance in a very large population, as predicted by (3) for the parameters used ( $\lambda_r = 0.09, \lambda_w = 0, \sigma_{mr} = 0.05, \mu = 0.25$ ). Black bars correspond to a population of constant size ( $N_e = N$ ) such that  $N_e\mu$  has the value indicated on the abscissa. White bars correspond to populations of fluctuating size with the same value of  $N_e\mu$  as for the adjacent black bar, where each population spent  $t_L = 250$  generations in a regime of large population size ( $N_L = 10000$ ) and  $t_S = 250$  generations in a regime of small population size,  $N_S$ .  $N_S$  was calculated from (9) as  $N_S = fN_eN_L/(N_L - N_e(1 - f))$ , with  $f = t_S/(t_L + t_S)$ . Mean overlap for  $N_e\mu = 5$  and constant population size was equal to  $1.5 \times 10^{-5}$ , a value too small to appear on the plot.

**Fig. 5. Influence of fluctuations in  $N\mu$  on functional overlap.** Dots and bars correspond to mean and standard deviation of functional overlap  $\langle r \rangle$  among two genes in an ensemble of 50 populations after  $10^5$  generations of simulated evolution with fluctuating population size. The simulation was started at a mean overlap corresponding to the upper horizontal line, which is the mean overlap at mutation-selection balance for a very large population from (3). The lower horizontal line is the value to which functional overlap would decline from this initial value after  $10^5$  generations if the population ensemble was under the sole influence of drift for the parameters used here. In the simulation, the population ensemble spent alternately  $t_S$  generations in a regime of low  $N\mu$  (dominated by drift) and  $t_L$  generations in a regime of high  $N\mu$  (dominated by selection). This cycling was repeated until  $10^5$  generations had elapsed and was continued after that until the end of the next period of small  $N\mu$  was reached. At that time, ensemble overlap statistics were evaluated. The dashed curve represents the value of  $N_e\mu$  as a function of  $t_S$  and  $t_L$ , where  $N_e$  is the inbreeding effective population size calculated from  $t_S$  and  $t_L$ . For the entire range of this plot, the values of  $N_e\mu$  suggest that the evolutionary dynamics should be dominated by drift, such that the mean ensemble overlap should be close to the lower horizontal line (Wagner, 1999, Fig. 7). However, as long as more than 50 generations are spent in each cycle of large population size ( $t_L \geq 50$ ) the influence of

selection continues to be appreciable, even if only a small fraction of time is spent in the large population regime. Only for smaller  $t_L =$  does drift dominate if  $t_S/t_L > 50$ . For reasons of computational feasibility, it was the mutation rate and not the population size that was varied.  $N = 1000, \mu_S = 0.0005, \mu_L = 0.5, \lambda_w = 0, \lambda_r = 0.9, \sigma_{mr} = 0.05$ .

**Fig. 6. Pleiotropy slows the decay of functional overlap.** (a) The function  $f(r) = r^l$  illustrates how the probability of neutral mutations for two genes of a given functional overlap  $r$  is influenced by the number of functions  $l$  affected per mutation. (b) Evolution of functional overlap  $\langle r \rangle$  in a population ensemble (20 populations) under the influence of drift as a function of the degree of pleiotropy  $l$ . Dots and bars represent means and standard deviations as obtained from a Monte Carlo simulation. The solid lines are numerical solutions of (10). ( $\lambda_r = 0.5, \sigma_{mr} = 0.05, \lambda_w = 0, \mu = 2.5 \times 10^{-6}$ ). As  $l$  is increased from one to three, the decay of  $r$  slows. (c) Times  $\tau_x$  until  $\langle r \rangle$  decreases from one to  $x$  as a function of the degree of pleiotropy  $l$  as calculated from (15). A linear increase in  $l$  leads to an exponential increase in  $\tau_x$ . ( $\lambda_r = 0.9, \sigma_{mr} = 0.05, \lambda_w = 0, \mu = 2.5 \times 10^{-6}, N = 100$ ).