

Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators

Andreas Wagner

Institute for Advanced Study ("Wissenschaftskolleg") Berlin, Wallotstrasse 19, D-14193 Berlin, Germany

Received: 18 July 1995/Accepted in revised form: 9 January 1996

Abstract. In various organisms loss-of-function mutations of individual genes with unexpectedly weak or no phenotypic effects in the homozygous state have been observed. In several of these cases, independent evidence shows that the respective gene products do have essential biological functions. An explanation emerging from detailed biochemical and genetic studies on such genes is that two or more genetically redundant genes contribute to that function, i.e., a group of genes that is able to substitute partially for a loss of function in one member of that group. The often-observed sequence similarity among redundant genes suggests gene duplications as a frequent source of genetic redundancy. Aside from this observation, the evolution of genetic redundancy is poorly understood. Genetic redundancy is potentially of great relevance to organismal evolution, since it may (i) 'protect' organisms from potentially harmful mutations, and (ii) maintain pools of functionally similar, yet diverse gene products, and thus represent a source of evolutionary novelty at the biochemical level. The question of how genetic redundancy evolves should ideally be answered by experimentation. However, the large time scales involved and insufficient quantitative understanding of the underlying regulatory pathways are likely to preclude such an approach in the foreseeable future. Preliminary answers are sought here by using a biochemically motivated model of a small but central part of a developmental pathway. Sets of transcription regulators are modeled that mutually regulate each other's expression and thereby form stable gene expression patterns. It is then studied how genetic redundancy caused by gene duplications might evolve in such networks. The results obtained suggest that redundancy may, at least in some cases, be a global property of gene interactions within a regulatory pathway, rather than a local property of genes in that pathway. They also raise the possibility that duplications of a whole regulatory gene network, as may have taken place during the evolution of *HOM/Hox* genes in chor-

dates, are less likely to be reversible (by gene deletions) than duplications of individual network genes. These findings are discussed with reference to experimental evidence on the evolution of *HOM/Hox* genes.

1 Introduction

Probably the most important way of controlling gene expression in eukaryotes is regulation of transcription initiation by RNA polymerase II, exerted via transcription regulators that bind to promoter or enhancer sequences near transcription units. Such regulators interact with the basal transcription machinery at the site of transcription initiation, thereby activating or repressing initiation by mechanisms that are still incompletely understood (Johnson and McKnight 1989; Mermelstein et al. 1989). Transcription regulators are also centrally involved in organismal development. Experimental evidence suggests that groups of genes encoding transcription regulators frequently interact in a network-like fashion of cross- and autoregulation, thereby establishing patterns of gene expression which, in turn, determine basic *Bauplan* features of the organism. Their important role makes such networks and their constituent genes an important and popular subject of research in evolutionary and developmental biology (Seeger and Kaufman 1987; Ingham 1988; Kappen et al. 1989; McGinnis et al. 1990; Olson 1990; Rosenfeld 1991; Sommer and Tautz 1991; Stuart et al. 1991; McGinnis and Krumlauf 1992; Schubert et al. 1993). To comprehend the regulatory superstructure represented by a network of transcription regulators and its evolution is a daunting task, given our incomplete understanding of transcription regulation. The modest approach pursued here can only be justified by the importance of the problem. This approach uses a mathematical model for networks of transcription regulators to study their evolution (see also Wagner 1994, 1996).

Loss-of-function mutations with only weak or no phenotypic effects in the homozygous state are frequent

for various genes and in various organisms. While the possibility can sometimes not be excluded that expression of the respective genes in the wild-type is without functional significance, there are several cases where independent evidence clearly demonstrates that the affected genes have an essential biological function. An explanation emerging from detailed biochemical and genetic studies on such genes is that two or more genetically redundant genes contribute to that function, i.e., a group of genes is able partially to substitute for a loss of function in one member of that group (e.g., Joyner et al. 1991; for reviews see Gatherer 1993; Tautz 1992; Thomas 1993). Only a double, triple, etc., mutation would then exhibit the expected strong phenotypic effect. Evidence for genetic redundancy caused by gene duplications is also available for genes that are presumably parts of a network of transcriptional regulators, e.g., the multiple *HOM/Hox* gene clusters of mouse (McGinnis and Krumlauf 1992). Some of this evidence stems from the phenotypes of loss-of-function mutants. Further, more indirect evidence is provided by the apparent deletion of several homeobox genes from putative gene clusters in various taxa (Cartwright et al. 1993; Krumlauf 1994; Misof and Wagner 1995).

Various reasons for the existence of genetic redundancy are conceivable. On one hand, it might only be the remnant of past gene duplication events. On the other hand, genetic redundancy may have adaptive value, since it may 'protect' a developmental pathway from some otherwise deleterious mutations. Thus, redundancy might be maintained by natural selection [cf. Wagner (1996) for a comparable mechanism in the evolution of genetic canalization]. The finding that genetic redundancy is often observed among genes with considerable sequence similarity (e.g., Joyner et al. 1991; McGinnis and Krumlauf 1992) suggests gene duplications as an important source of redundancy. A typical scenario might involve duplication of a gene and subsequent genetic divergence of original and duplicate. Immediately after duplication, the pair of gene copies is completely redundant, implying that a loss-of-function mutation in one gene (e.g., a gene deletion) has no phenotypic effect. Over time, as the two copies come to acquire different functions, redundancy is reduced. This process may generate a reservoir of genes with distinct but overlapping functions, a pool of genetic variation that is a potential source of evolutionary innovations on the biochemical level.

2 The model

The model to be used here was motivated in greater detail by Wagner (1994, 1996). It is concerned with the expression pattern of transcription factor genes in one developmental stage and in one cell or body region that shares an expression pattern, such as, for example, a set of nuclei in part of a *Drosophila* blastoderm expressing a specific subset of segmentation genes (Ingham 1988). In the model, a network of transcription regulators is represented by a dynamical system whose state variables cor-

respond to expression states of network genes. These states are denoted as

$$\vec{S}(t) := (S_1(t), \dots, S_N(t)) \quad (1)$$

where $S_i(t)$ is the expression state of the i th gene, G_i , at some time $t \geq 0$ during some developmental process in which the network acts. For reasons of computational simplicity, it is assumed that $S_i(t)$ can only assume two values, namely $(+1)$ and (-1) , corresponding to a situation in which the gene G_i is expressed or not expressed, respectively, at time t . The gene expression state $\vec{S}(0)$ at time $t = 0$ is called the initial expression state. It can be conceptualized as being imposed onto the network by the products of one or more 'upstream' genes that are not themselves part of the network.

Starting from the initial gene expression pattern, $\vec{S}(0)$, cross- and autoregulatory interactions among network genes cause the expression state to change. These changes are modeled by the set of difference equations

$$S_i(t + \tau) = \sigma \left[\sum_{j=1}^N w_{ij} S_j(t) \right] = \sigma [h_i(t)] \quad (2)$$

Here, the expression state of gene G_i at time $t + \tau$, $S_i(t + \tau)$, is a function of a weighted sum, $h_i(t)$, of the expression state of all network genes at time t . $h_i(t)$ represents the sum of the regulatory effects of all network genes on gene G_i . $\sigma(x)$ is the sign function ($\sigma(x) = -1$ for $x < 0$, $\sigma(x) = +1$ for $x > 0$ and $\sigma(0) = 0$), and τ is a time constant characteristic for the process under consideration. The real constants w_{ij} represent the 'strength' of regulatory interaction of the product of G_j with G_i , i.e., the degree of transcriptional activation ($w_{ij} > 0$) or repression ($w_{ij} < 0$) that the transcription regulator produced by gene G_j has on gene G_i . Such regulatory interactions are known to be mediated by regulatory (enhancer) DNA elements. In biological terms, individual values of w_{ij} can be thought of as some compound measure of the binding constant and the transcription activation (repression) ability of the factor produced by G_j at the enhancer element that mediates its interaction with G_i . Mutations of regulatory DNA elements are modeled by changes in individual w_{ij} values. The 'connectivity matrix' $w = (w_{ij})$ that the constants w_{ij} define corresponds to all regulatory DNA elements relevant to regulatory interactions among network genes. The quantity $-\sigma[\sum_{j=1}^N w_{ij}]$ can be viewed as the activation state of gene i if no other genes in the network are expressed (corresponding to the basal rate of transcription in a model with continuous state space). Some (or most) entries of w may be zero. The fewer non-zero entries w has, the fewer regulatory interactions exist among network genes. An important model parameter is therefore the fraction of entries different from zero, denoted by c ($c \in (0, 1)$), which will be called the 'connectivity density' of the network. The discrete-time dynamical system (2) can also be viewed as the limiting case of a system of differential equations, in which concentrations of gene products, rather than binary (on-off) gene expression states change (Wagner 1994), but computational limitations prohibited the use of such a system

here. For the same reason, a haploid model is used. However, all results will apply to a diploid situation, provided that there is little allelic variation in network genes. It should be noted that the structure of (2) is similar to 'spin glass' (Binder and Young 1986) or neural net (Hopfield and Tank 1986; Amit 1989) type models of gene networks first introduced by Kauffman (1969, 1993). However, (2) is conceptually different from Kauffman's models in that a specific type of gene interaction, namely transcription regulation, is considered. Recently, models conceptually similar to (2) have been successfully used to describe and predict regulatory gene interactions in early *Drosophila* embryogenesis (Mjolsness et al. 1991; Reinitz et al. 1995; Reinitz and Sharp 1995).

The dynamics of (2) will lead to the attainment of an equilibrium gene expression state, which may be a fixed point of (2) or a limit cycle. For reasons of tractability, only fixed point equilibria will be considered here. They are denoted by $\bar{S}(\infty)$. Genes expressed in the equilibrium state will affect the expression of genes outside ('downstream') of the network. Many experimental genetic studies (e.g., McGinnis and Krumlauf 1992) suggest that deviations in the expression pattern of a gene network from the wild-type pattern causes developmental perturbations that often lead to deleterious effects on the adult phenotype. In a biological population, such individuals would be eliminated by natural selection. These observations motivate the assumption that an optimal equilibrium gene expression state, denoted as $\bar{S}^{\text{opt}}(\infty)$, exists for networks acting in a developmental process. If a network attains an equilibrium state $\bar{S}(\infty)$ different from this optimal state, developmental perturbations will result and the fitness of the resulting adult organism will be reduced. The fitness of an organism (network) is modeled with a Gaussian fitness function

$$\exp\left(-\frac{d_h[\bar{S}^{\text{opt}}(\infty), \bar{S}(\infty)]^2}{s}\right) \quad (3)$$

where d_h denotes the normalized Hamming distance (e.g., Amit, 1989) of the two states, and $s(s > 0)$ represents the strength of selection. Deviations from an optimal state may occur for a variety of reasons, one of them being mutations in genes within or outside the network. The pattern of regulatory interactions within a gene network represents its most interesting 'organizational' properties, and thus mutations in regulatory DNA regions, represented by changes in the matrix w , will be the focus of this study.

Gene duplications, mutations and deletions are conceptualized as illustrated in the following example. Figure 1a shows a matrix w of regulatory interactions within a network of three genes: G_1 , G_2 and G_3 . Figure 1b shows the regulatory interactions of this network after duplication of gene G_1 and gene G_3 . The duplication event adds two genes to the network, denoted as G_1^* and G_3^* , as well as two rows to the matrix of Fig. 1a. These two rows, rows 2 and 5, correspond to the promoters of G_1^* and G_3^* , respectively. Duplicated matrix elements within any given row correspond to the influence of the two copies of a duplicated gene on the regulation of the

$$\begin{array}{l} \text{a} \\ \begin{matrix} G_1 \\ G_2 \\ G_3 \end{matrix} \begin{pmatrix} w_{11} & 0 & w_{13} \\ 0 & w_{22} & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix} \\ \\ \text{b} \\ \begin{matrix} G_1 \\ G_1^* \\ G_2 \\ G_3 \\ G_3^* \end{matrix} \begin{pmatrix} w_{11} & w_{11} & 0 & w_{13} & w_{13} \\ w_{11} & w_{11} & 0 & w_{13} & w_{13} \\ 0 & 0 & w_{22} & w_{23} & w_{23} \\ w_{31} & w_{31} & w_{32} & 0 & 0 \\ w_{31} & w_{31} & w_{32} & 0 & 0 \end{pmatrix} \\ \\ \text{c} \\ \begin{matrix} G_1 \\ G_1^* \\ G_2 \\ G_3 \\ G_3^* \end{matrix} \begin{pmatrix} w_{11} & w_{11} & 0 & w'_{13} & w'_{13} \\ w_{11} & w_{11} & 0 & w_{13} & w_{13} \\ 0 & 0 & w'_{22} & w_{23} & w_{23} \\ w_{31} & w_{31} & w_{32} & 0 & 0 \\ w_{31} & w_{31} & w_{32} & 0 & 0 \end{pmatrix} \\ \\ \text{d} \\ \begin{matrix} G_1^* \\ G_2 \\ G_3 \\ G_3^* \end{matrix} \begin{pmatrix} w_{11} & 0 & w_{13} & w_{13} \\ 0 & w'_{22} & w_{23} & w_{23} \\ w_{31} & w_{32} & 0 & 0 \\ w_{31} & w_{32} & 0 & 0 \end{pmatrix} \end{array}$$

Fig. 1a–d. Changes in connectivity matrices caused by gene duplications, mutations and gene deletions. **a** The connectivity matrix of a network with three genes (G_1 , G_2 and G_3) and connectivity density $c = 2/3$ is shown. Symbols w_{ij} represent non-zero connectivities (regulatory interactions). **b** The connectivity matrix of the network in **a** after duplication of G_1 and G_3 . Rows 2 and 5 correspond to the promoters of the newly created genes, G_1^* and G_3^* . Note the duplication of all matrix entries that mediate regulatory effects of the duplicated genes, reflecting the increased regulatory influence of duplicated genes. **c** The network in **b** after two mutation events have changed regulatory interactions within the network. **d** A network resulting from deletion of G_1 from the network shown in **c**

gene whose promoter this row represents, but not to two separate promoter sites. For example, the two elements w_{23} in row three of Fig. 1b represent the increase in relative regulatory influence on the expression of gene G_2 that gene G_3 has gained after its duplication. One might thus say that gene dosage effects enter this model indirectly through changes in the balance of regulatory interactions of the network. Immediately after duplication, original and duplicate have identical promoters and will therefore have identical dynamics with respect to (2). Mutations can break this symmetry, since they affect promoter elements of the two genes independently, as exemplified by Fig. 1c. The matrix shown in this figure is derived from that in Fig. 1b. Two mutations occurred, symbolized by primed symbols. Because both elements w_{13} represent one regulatory interaction, they are changed jointly to w'_{13} . Note also that the promoters of gene G_1 and its duplicate are different now, which may lead to differential expression of the two genes. Deletions of some or all of the previously duplicated genes can occur. These deletions are assumed to affect original gene and duplicate with equal probability. Figure 1d shows an

example of a deletion of G_1 , leading to a change in the matrix of Fig. 1c. The duplicated gene, G_1^* , is left intact. Such events will again change the regulatory balance within a network and may lead to changes in a gene expression pattern. A direct comparison of a gene expression state before and after deletion is impossible, because these events change the dimension of the network's state space. However, if original and duplicate genes show little functional divergence, as in the scenario considered here, this problem has a straightforward solution, which is outlined in the Appendix.

Consider a network in which a gene was duplicated without a (deleterious) change in the network's gene expression pattern. As time passes, mutations alter regulatory interactions within the network, the regulatory roles of original and duplicate gene copies may diverge, and deletions of some or all of the previously duplicated genes may alter the gene expression pattern of the network. The likelihood of such a change, denoted as $r(d)$, is taken as a measure of genetic redundancy. How it depends on d , a measure of the extent to which a network's pattern of regulatory interactions has diverged between duplication and deletion, is one of the questions studied here. In the numerical approach taken here, a population of (initially) identical gene networks is generated from one 'founder' network in which one or more network genes are duplicated. The optimal gene expression state within the population is the equilibrium gene expression state of this founder network. Then, networks within the population are allowed to diverge by subjecting them to mutation, while exerting stabilizing selection on the optimal gene expression pattern. In the course of this process, the effects of gene deletions on the network's gene expression patterns are studied.

Little is known empirically about 'typical' initial and equilibrium expression states, as well as about patterns of regulatory interactions within gene networks. Starting a population simulation with one type of network that has a prespecified initial, equilibrium state and connectivity matrix would therefore require many *ad hoc* assumptions for these parameters. This is in part circumvented here by pursuing a statistical approach involving infinite sets E ('ensembles') of networks, each with its own initial state, equilibrium state and matrix w of regulatory interactions. Each network within an ensemble is the starting point of an evolution scenario as outlined above, during which evolution of stability in a population of organisms is monitored. This permits one to assess how sensitive results are to variations in network features. In this approach, the ensemble is a more suitable level for characterizing network structure than the individual network, and network characterization is therefore statistical in nature. The following parameters are used to characterize a network ensemble. First, all networks within an ensemble have the same number of genes, N . Second, among networks within an ensemble, the number of genes expressed in the initial state follows a binomial distribution $B(N, p)$. This implies that within the ensemble the mean number of genes expressed in the initial state is given by Np . Genes expressed in the equilibrium state follow the same distribution, but are

stochastically independent of those expressed in the initial state. Third, connectivity matrices w within the ensemble are defined by a probability distribution $\rho(w)$ of regulatory interaction strengths (specified below), with a mean fraction c of connectivities different from zero among networks within the ensemble. The number of non-zero connectivities varies from network to network. Since a numerical analysis is carried out here, only finite samples of a network ensemble can be studied. Such samples are generated by a random search in the space of all possible gene networks with given N , p , c and ρ . A separate population simulation is carried out for each member of a sample of E , using the member as a population's founder network in the sense outlined above. Results are then averaged over all population simulations. Two scenarios will be discussed independently, since they best illustrate the principles involved. In the first scenario, only one gene is duplicated. In the second scenario, all N network genes are duplicated simultaneously. One gene and N genes are subsequently deleted, respectively. In other words, the same number of genes that was duplicated previously is deleted in one single gene deletion event. The important question regarding robustness of results to variation in model features is addressed by numerically generating several ensemble samples with different values of N , p , c and ρ .

3 Numerical methods

The numerical analysis carried out here involves two principal steps. First, a sample of a network ensemble E was generated numerically. For each member of the sample, a prespecified number of genes was duplicated, and it was tested whether $\bar{S}(\infty)$ is changed by this duplication, i.e., whether $\bar{S}(\infty) \neq \pi^{-1}[\bar{S}^{du}(\infty)]$ (see Appendix), given $\pi[\bar{S}(0)]$ as initial state after duplication. The subset of E that meets this condition is called E^d .

3.1 Generation of E^d

Each network in an ensemble (sample) has its own initial state $\bar{S}(0)$, equilibrium state $\bar{S}(\infty)$, and connectivity matrix w . Because sample members were generated independently from each other, it will only be described how one member was generated.

First, two binary pseudorandom arrays in $\{-1, 1\}^N$, corresponding to $\bar{S}(0)$ and \bar{S}^{opt} , were generated. Individual entries of these arrays are stochastically independent both within each array and among arrays. Each entry was chosen according to the probabilistic rule $P(S_i = 1) = p$, p being a real number in $(0, 1)$. Then, a $N \times N$ pseudorandom matrix $w = (w_{ij})$ of regulatory interaction strengths was generated. Its entries w_{ij} are independently and identically distributed with probability density $(1 - c)\delta(w_{ij}) + c\rho(w_{ij})$. Here, ρ denotes a continuous probability density. Gaussian ($\rho(x) = 1/(\sqrt{2\pi}\sigma)\exp[-x^2/(2\sigma^2)]$) or 'reflected gamma' densities ($\rho(x) = [2\Gamma(a)]^{-1}e^{-|x|}|x|^{a-1}$, $x \in \mathbb{R} \setminus \{0\}$, $a > 0$) were used, since they cover a wide range of qualitatively different

shapes of symmetric distributions. $\delta(x)$ denotes the Dirac delta function and $c \in (0, 1)$ is a real number which determines the mean number cN^2 of entries of w that are different from zero. It will be referred to as the ‘density’ of regulatory interactions within the network. Using $\bar{S}(0)$ as initial state, the network’s dynamics (2) was evaluated numerically for up to $2.5N$ time steps with this random matrix. If the network did not attain the equilibrium state $\bar{S}(\infty)$ as a fixed point, the matrix w was ‘discarded’, a new matrix w was generated in the same way as the old one, and the network’s dynamics was evaluated again. New matrices were generated in this way until a matrix had been found for which the network attained $\bar{S}(\infty)$ as a fixed point, or until 5×2^{2N} matrices had been generated, whichever came first. If no matrix had been found after 5×2^{2N} trials, the pair of states $\bar{S}(0)$ and $\bar{S}(\infty)$ was discarded, a new pair of state arrays $\bar{S}(0)$ and $\bar{S}(\infty)$ was generated in the same way as the old pair, and a stochastic search for a matrix w was carried out in the same way as for the old pair of states. Once a matrix was found for which (2) attained $\bar{S}(\infty)$ as a fixed point with $\bar{S}(0)$ as initial state, the triplet $(\bar{S}(0), \bar{S}(\infty), w)$ was considered a member of the ensemble sample. Further details and a motivation of this sample generation procedure can be found in Wagner (1996). The parameters N , c and p , as well as the distribution ρ , were not changed while generating one ensemble sample. However, for different ensemble samples, different parameter values were used ($0.1 < p < 0.9$, $0.25 < c < 1.0$, different distribution types for ρ) to assess the robustness of the results to parameter variation. An ensemble member found in this way was subjected to a gene duplication event. If only one gene was to be duplicated, gene 1 was chosen (without loss of generality). If gene duplication had not changed the equilibrium expression state, the triplet $(\bar{S}(0), \bar{S}(\infty), w)$ was considered a network in E^d .

The second principal step of the numerical analysis, a simulated evolution process, was carried out only for members of E^d and separately for each member. It will thus be described only for one member.

3.2 Simulated evolution

A ‘population’ of 100 identical copies of the ensemble member was generated and subjected up to 400 times to the following two steps in the given order. Each iteration of these steps was considered one ‘generation’ of simulated evolution.

1. *Mutation.* Exactly one randomly chosen non-zero entry of the connectivity matrix of each member of the population was replaced by a pseudorandom number distributed as $\rho(w_{ij})$. This approach is essentially the house-of-cards approximation from population genetics (see e.g., Turelli 1985; Zeng and Cockerham, 1993). The entry chosen for mutation was selected randomly with equal probability among all entries and independently from entries mutated in previous generations.

2. *Fitness evaluation and selection.* Each of the 100 networks (‘individuals’) was subjected to the dynamics (2) with initial state $\pi[\bar{S}(0)]$. If a network had not reached equilibrium after $2.5N$ time steps [see Wagner (1996) for

a motivation for this value], its fitness was assigned the smallest possible value of $\exp(-1/s)$, thus making it very unlikely that the network would survive the subsequent ‘selection’ step. Fitness of networks that attained an equilibrium state was calculated using (3) with $s = 0.01$, a value that ensures strong stabilizing selection. A network was chosen at random with uniform distribution and a pseudorandom number, r , with uniform distribution on $(0, 1)$ was generated. If r was smaller than the fitness of the network, the network ‘survived’. This process was repeated until a new population of the same size as the old population had been generated, i.e., sampling of networks was carried out with replacement, corresponding to ‘soft selection’ (e.g., Hartl and Clark, 1989, p 525).

During simulated evolution (every generation for the first 21 generations and every five generations thereafter), population members that attained the optimal equilibrium state, i.e., where $\pi^{-1}[\bar{S}^{du}(\infty)] = \bar{S}^{opt}(\infty)$, were subjected to simulated gene deletion events. Each of these networks was subjected to deletion of the previously duplicated gene(s). The fraction of networks that still attained the optimal equilibrium state after deletion was calculated and used as a measure of genetic redundancy. A high fraction of such networks indicates a high degree of redundancy. The results reported below are averages over the independent simulations carried out for populations derived from each ensemble member.

Small ensemble samples (200 networks) and population sizes (100 individuals) were used here because of computational limitations. However, individual exploratory simulations (not shown) with up to a sevenfold larger sample size and a fivefold larger population size were carried out, and the results obtained did not differ qualitatively from those obtained with the smaller values, suggesting that they are robust to changes in these parameters.

4 Results

Figure 2a shows the evolution of redundancy for networks subject to two different duplication events as a function of genetic divergence. Divergence, d , is measured by the number of mutations in regulatory DNA sequences (connectivities) that accumulated per network between duplication and deletion event. The measure of redundancy is the probability that deletion of a number of genes equal to those previously duplicated does not lead to a (deleterious) change in a network’s gene expression pattern. High redundancy means that a deletion event is unlikely to have deleterious effects. The upper set of points in Fig. 2a represents redundancy, $r_1(d)$, in a network where one gene was duplicated; the lower set of points represents the corresponding quantity, $r_N(d)$, for a network in which all genes were duplicated. Both sets were generated with densely connected ($c = 1$) networks of $N = 10$ in which 50% of network genes were expressed ($p = 0.5$). Figure 2a shows that r_1 decreases rapidly from its initial value of one (i.e., all deletions are neutral) immediately after duplication, to an equilibrium of

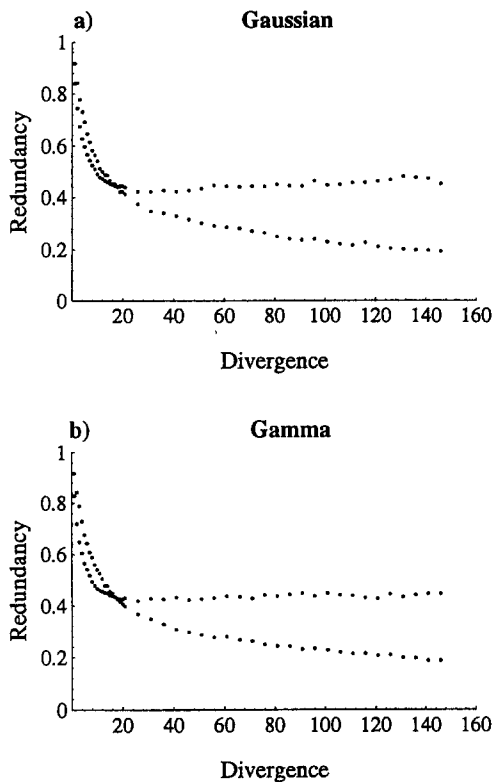


Fig. 2a, b. Evolution of redundancy and the distribution of regulatory interactions. The *upper* and *lower graphs* in **a** and **b** represent redundancies r_1 and r_N , respectively, plotted every generation for the first 21 generations after duplication, and every five generations thereafter, which results in the differential spacing of the data points. Divergence is measured by the number of mutations per network after duplication. **a** Gaussian distribution of connectivities with $\sigma = 1$; **b** reflected gamma distribution with $a = 2$ (see Sect. 3). Further parameter values: $N = 10$, $c = 1$, $p_0 = p_{eq} = 0.5$

$r \approx 0.45$. It is noteworthy in and by itself that non-neutral deletions are possible, because only divergence in regulatory DNA regions, but no functional divergence in gene products was modeled. This implies that genetic redundancy in a regulatory pathway need not be solely determined by the similarity of functions of individual gene products. How individual gene products interact to arrive at gene expression patterns may play an important role. Figure 2a also shows that deletions of N genes are always more likely to be deleterious than deletions of one gene, except for a short initial period after duplication. While r_1 has reached an equilibrium long before 150 generations have elapsed, redundancy for the N gene case continues to decrease slowly. Results (not shown) from one exploratory simulation ($N = 10$, $p = 0.5$, $c = 1$, ρ Gaussian) demonstrate that r_N does not decline to zero, but equilibrates at a value of $r_N \approx 0.08$ after about 700 generations. Because of the enormous computational cost of the numerical approach used here, only 150 generations of simulated evolution were carried out in all other simulations. It is noteworthy that equilibrium redundancy in the two cases differs only by a factor of 5.6 ($r_1/r_N = 0.45/0.08$). If, as one might assume, individual

duplicated genes contribute independently to the level of redundancy in a network, r_N would decline to a much lower value of $0.45^9 \approx 8 \times 10^{-4}$.

Figure 2b shows results completely analogous to Fig. 2a, except that individual regulatory interaction strengths follow a reflected gamma distribution ($a = 2$; see Sect. 3) instead of a normal distribution. It can be seen that redundancy evolves in a qualitatively identical way in the two cases. Other distribution parameters yield very similar results (not shown). This indicates that the results reported here are largely independent of how regulatory interaction strengths are distributed within a gene network, and therefore, to a limited extent, of the kind of changes mutations cause in regulatory DNA elements.

Unfortunately, the dependence of r on the number of network genes could not be explored adequately, because computational requirements increase rapidly as the number of genes, N , is increased. However, a qualitative statement about the dependence of redundancy on N in the case of one duplicated gene can be made. Results reported by Wagner (1994) suggest that duplications of individual genes are less and less likely to have deleterious effects as N is increased. Since deletion of previously duplicated genes is a process symmetrical to duplications in this model, genetic redundancy is likely to approach the maximum value of one (almost all deletions being neutral) as N becomes very large, regardless of the amount of divergence accumulated.

Figure 3 shows how the evolution of redundancy depends on the density of regulatory interactions, c , i.e., the number of genes that regulate the expression of any given gene in a network. From this figure, it can be seen that a decrease in c does not affect the evolution of r_N . However, r_1 increases as c increases, i.e., deletions of one gene become less and less likely to be deleterious. This finding has a quite intuitive explanation. Assume that gene i is duplicated, and consider the regulatory DNA elements, represented by w_{ji} , $1 \leq j \leq N$, that are directly affected by this duplication. Most of these elements will be zero if c is small, and thus, the duplication of only one gene will not even affect the regulatory 'input' at the promoters of most other network genes. As a consequence, it becomes more likely that the regulatory dynamics within the network is not perturbed to an extent that leads to a change in a gene expression pattern if c is decreased.

Figure 4 shows how the evolution of redundancy depends on p , the fraction of expressed genes in the network. In Fig. 5, redundancy values attained at the end of the simulations shown in Fig. 4 and Fig. 2a are summarized. 'Error' bars correspond to one standard deviation of redundancy over the 200 independent population simulations carried out for each parameter. These figures show that r_1 does not change significantly as p is changed. r_N , on the other hand, increases markedly as the number of expressed genes becomes very small. Thus, in networks with very few expressed genes, a deletion of N genes is more likely to be neutral than in a network with many expressed genes. A simple analytical argument given in the Appendix shows that this is a consequence of

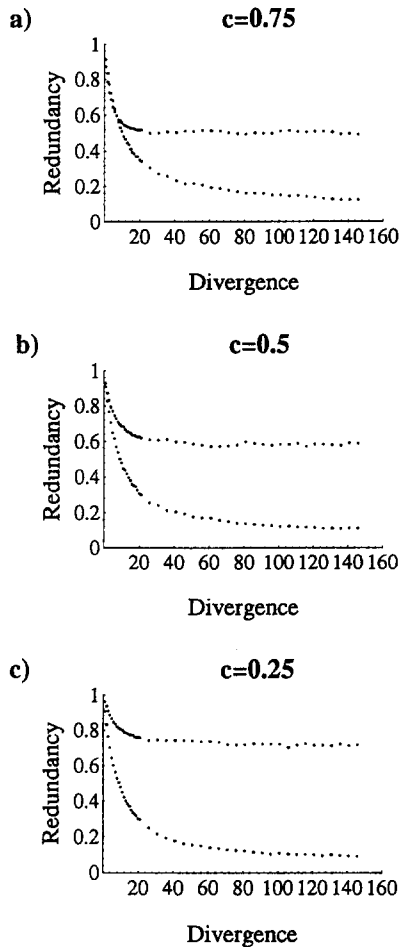


Fig. 3a–c. Dependence of redundancy on c . The *upper* and *lower* graphs in a–c represent r_1 and r_N , respectively, plotted every generation for the first 21 generations after duplication, and every five generations thereafter. Divergence is measured by the number of mutations per network after duplication. Parameter values are as in Fig. 2a, except for c : **a** $c = 0.75$; **b** $c = 0.5$; **c** $c = 0.25$

the number of original/duplicate gene pairs, in which one of the genes is expressed while the other is not.

To summarize, one can say that (i) genetic redundancy between two or more genes need not necessarily be determined by the functional similarity/dissimilarity of the respective gene products; exactly how regulatory interactions involving multiple network genes lead to the attainment of a given gene expression pattern may be an important determinant of the amount of redundancy in a regulatory network; (ii) redundancy is, in general, lower if many genes are duplicated and subsequently deleted than if few genes are duplicated and deleted; (iii) decreasing the ‘density’ of regulatory interactions within a network makes deletions of one gene less likely to be deleterious; (iv) if few genes are expressed in a network, deletions of N previously duplicated gene become less likely to be deleterious; (v) networks with many genes are probably more resilient to duplications/deletions of individual genes than networks with few genes.

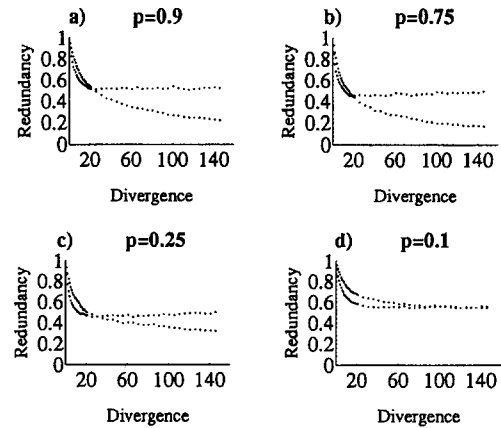


Fig 4a–d. Dependence of redundancy on the mean number of expressed genes, Np . The *upper* and *lower* graphs in a–d represent r_1 and r_N , respectively, plotted every generation for the first 21 generations after duplication, and every five generations thereafter. Divergence is measured by the number of mutations per network after duplication. Parameter value are as in Fig. 2a, except for p : **a** $p = 0.9$; **b** $p = 0.75$; **c** $p = 0.25$; **d** $p = 0.1$

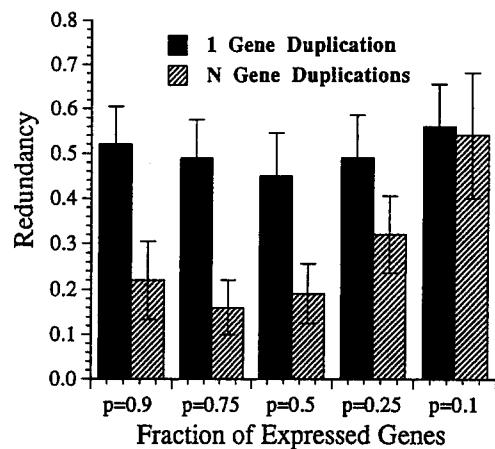


Fig. 5. Redundancy and Np : summary of Figs. 2a and 4. Shown are r_1 and r_N at the end of the simulations shown in Figs. 2a and 4 ‘Error’ bars represent one standard deviation of redundancy over the 200 independent population simulations carried out for each case. Except for varying p , all parameter values are as given in Fig. 2a

5 Discussion

To determine evolutionary conservation, increase or decay of genetic redundancy experimentally would mean (i) identifying groups of genes that perform comparable and redundant functions in different taxa, (ii) defining a suitable measure of redundancy among these genes and (iii) comparing the level of redundancy in these taxa. Available experimental data do not yet provide the necessary level of resolution, and a modeling approach may therefore be a useful first step. For the purpose of this paper, genetic redundancy is defined as the absence of a phenotypic effect of gene deletions (loss-of-function mutations) in gene networks encoding for transcription regulators that perform essential biological functions. Also, it is assumed that redundancy is caused by gene

duplications, a prominent, although not the only conceivable source of genetic redundancy (Tautz 1992).

Aside from their central role in development, genes encoding for proteins involved in networks of transcription regulators also include examples of genetically redundant genes (e.g., Joyner et al. 1991; see also below). This motivates the approach taken here, in which a dynamical system was used to model the evolution of redundancy in networks of transcriptional regulators. Using this model, it was studied how the evolution of regulatory DNA sequences after a gene duplication event influences genetic redundancy. Some time after duplication of one or more network genes, previously duplicated genes were deleted. The greater the time that elapses between duplication and deletion, the more the regulatory sequences diverge from those governing the regulatory interactions immediately after the duplication event, and the more likely it is that a gene deletion event causes a deleterious change in the gene expression pattern of the modeled network.

The evolution of regulatory interactions among network genes was studied because they represent the central feature of a network's organization. Recent experimental findings regarding the three *Drosophila* genes *paired*, *gooseberry* and *gooseberry neuro*, underscore the importance of regulatory gene interactions for the evolution of redundancy. Li and Noll (1994) report that these three genes, which encode for transcription regulators and evolved from the same ancestral gene by repeated gene duplication, have distinct developmental functions only by virtue of their different *cis*-regulatory regions. It is an advantage of the modeling approach pursued here that the evolution of transcription units and of regulatory sequences can be conceptually separated. To further justify modeling the evolution of regulatory sequences, one might also argue that regulatory DNA sequences are likely to evolve at a faster rate than transcription units within the network. This is because the gene products considered here are likely to affect the expression of many genes, causing strong pleiotropic effects of mutations in their transcription units. Mutations in regulatory DNA sequences, on the other hand, will affect the expression of only one gene, and their effect may even be compensated by the overall regulatory dynamics of a network.

The results reported above also show that the amount of redundancy in a network need not be proportional to the functional similarity of the coding region of a gene pair (duplicate/original). Here, the reason is that global interaction patterns within the network may determine the extent of redundancy, such that even functionally identical genes need not be completely redundant. This 'non-local' property of redundancy for particular genes is not surprising if one thinks of a gene deletion as causing a perturbation in a parameter (gene dosage) of some dynamical system that represents part of a developmental pathway. However, it is sometimes overlooked in experimental studies, in which redundancy is implicitly attributed to the genes under consideration. Whether the phenomenon observed here is frequent in regulatory pathways remains to be seen. However, a po-

tentially relevant analogy exists in many formal genetic studies demonstrating that the phenotypic effect of an allele (a mutant) often depends on its genetic background (Rendel 1979; references in Wagner et al. 1994), and is therefore not only a property of the allele, but the result of interaction effects among many different gene loci.

HOM/Hox gene clusters are a prominent example of a class of genes in which presumably both gene duplications and gene deletions occurred. The manifold regulatory interdependencies that exist between these genes suggest that they also represent gene networks in the sense used here (e.g., Lewis 1978; Bienz and Tremml 1988; González-Reyes and Morata 1990). Since they are a candidate network whose evolution is well studied, the results obtained above will be discussed mostly with reference to them.

Two gene duplication scenarios were examined here. In the first, one network gene is duplicated and, subsequently, deleted. Such duplications of individual network genes are likely to have occurred frequently in evolution, such as in the case of the *HOM/Hox* gene clusters of mice, which have their origin in a small number of precursor genes (Schubert et al. 1993). It was examined how the likelihood, $r_1(d)$, that a gene deletion does not have a deleterious effect depends on the amount of genetic divergence, d , accumulated between duplication and the deletion event. It was found that this measure of redundancy decreases rapidly from its initial value $r_1(0) = 1$ (complete redundancy) immediately after the duplication event. It reaches an equilibrium, whose precise value depends on network features such as the number of genes that are 'connected' by regulatory interactions. A second scenario concerned duplication of all genes in a network, and subsequent deletion of an equal number of genes. Redundancy was again defined as the likelihood, $r_N(d)$, that such a gene deletion event does not have a deleterious effect. Duplications of whole *HOM/Hox* gene clusters, which seem to have occurred at least twice in the evolution of *HOM/Hox* genes (Schughart et al. 1989), represent likely examples for such network duplications. The biological question addressed with the second scenario thus regards the reversibility of network duplications: can a *HOM/Hox* gene cluster be deleted once duplicated without deleterious effects? In general, $r_N(d)$ is substantially lower than $r_1(d)$, implying that such cluster deletions are more likely to be deleterious. This prediction is consistent with current experimental evidence, since no support for a decrease in the number of *HOM/Hox* clusters during evolution has so far been found from phylogenetic studies. The prediction could be tested more rigorously if it is found that independent *HOM/Hox* gene cluster duplications have occurred more than once in evolution. Preliminary evidence for this possibility comes from a study on the chelicerate *Limulus polyphemus*, consistent with the existence of four *HOM/Hox* gene clusters (Cartwright et al. 1993). Any differences in redundancy for the two duplication scenarios will also be massively enhanced if genetic redundancy can evolve differently in different expression domains of a network. [The model (2) is only concerned with one such domain.] This possibility is consistent with

results of genetic studies indicating that loss-of-function mutations in a presumably cell-autonomous gene need not affect all regions in which the gene is expressed (e.g., Chisaka and Capecchi 1991). Whereas gene expression without functional significance may account for some of these cases, differential evolution of redundancy in different tissues or ontogenetic stages will probably be responsible in others.

These findings are based on a quite general model of a complex and poorly understood biochemical process, and the fact that they are qualitatively independent of particular model parameters is reassuring. It suggests that they are generic properties of gene networks, conserved over a broad class of network architectures. The only exception regards a case in which very few network genes are expressed. Here r_1 and r_N become very similar. However, a very low number of expressed genes would defeat the purpose of a regulatory network, and is therefore of questionable biological significance.

Caution is appropriate when approaching the issue of single gene deletions in groups of genes that presumably form a network, because more than one interpretation is possible in cases where a gene occurring in a group of taxa is apparently missing in other taxa related to that group. First, the gene may not be as structurally conserved as one would expect from its function, if this function is known. Its coding sequence may have evolved so rapidly that it cannot be cloned (PCR-amplified) by utilizing sequence similarity to putative homologues in related taxa. Second, the gene may not have been deleted in some of the taxa under consideration, but rather represent a new acquisition in the group of taxa in which it has been found. Being able to distinguish between the latter two possibilities requires detailed knowledge of phylogenetic relations in the groups under consideration. The *Drosophila* pair-rule segmentation gene *fushi tarazu* (*ftz*) may represent an example of either of these two cases. Initial attempts to identify an *ftz* homologue in the beetle *Tribolium* and in the honeybee *Apis mellifera* failed (Walldorf et al. 1989; Stuart et al. 1991), raising the possibility that *ftz* represents a recent addition to the *Drosophila* group, or that it was lost in other groups. However, more recent evidence suggests that *ftz* may have diverged rapidly, and that an *ftz* homologue exists in the locust *Schistocerca*, a taxon that is more distantly related to *Drosophila* than either *Tribolium* or *Apis* (Dawes et al. 1994; Ferrier and Akam 1994). Given these caveats, no clear case for loss of individual genes in organisms that presumably have only one cluster of *HOM/Hox* genes has been reported to date. The situation is different for organisms with more than one *HOM/Hox* gene cluster. For example, *HOM/Hox* genes in humans and mouse are organized into four clusters that presumably arose by two or more duplications of one ancestral cluster (Schughart et al. 1989). By the conserved gene organization in each cluster, as well as by criteria of sequence conservation, it is possible to identify closely related groups of genes on different clusters that may have a common ancestor in the original cluster. These groups are called cognate groups, and the mouse has 13 such groups, numbered 1(3') through 13(5') (Scott

1992). Most remarkably, only two of these 13 cognate groups (i.e., groups 4 and 9) have four member genes, as one would expect if the ancestral cluster had 13 member genes. In eight cognate groups, one gene is missing (groups 1, 3, 5, 6, 8, 10, 11, 13) and in three more cognate groups, two genes are missing (groups 2, 7, 12). The 5'-most groups 10 through 13 will be excluded here, since their organization (Scott 1992) indicates that they represent phylogenetically recent gene duplications. Even under this conservative assumption, nine of 36 genes are missing in cognate groups 1 through 9. Homology to the genes of the *Drosophila* Ant-C and BX-C clusters (McGinnis and Krumlauf 1992) indicates that most of these nine cognate groups were represented in the ancestral cluster. This strongly suggests that numerous gene deletions occurred after duplication of the ancestral cluster, and argues against the possibility that new acquisitions rather than gene deletions took place in these cases. It can of course not be strictly excluded that some of these genes have not been deleted, but were placed outside the cluster and have undergone rapid sequence evolution. However, given the extensive surveys for Antennapedia class homeobox genes that were carried out on the mouse genome (e.g., Murtha et al. 1991), it seems likely that even distant relatives of the apparently missing genes would have been detected, if they exist. Somewhat more preliminary evidence along the same lines is available from recent PCR surveys of Antennapedia class homeobox genes in several metazoan taxa. For example, evidence exists consistent with four clusters (nine cognate groups) of *HOM/Hox* genes in the Chelicerate *Limulus polyphemus*, where as many as 9 of the 36 expected genes may be missing (Cartwright et al. 1993). In the teleost fish *Fundulus heteroclitus*, a total of 22 members from eight cognate groups were found, consistent with as many as ten of 32 genes missing (Misof and Wagner 1996). In the zebrafish, evidence is available for four *HOM/Hox* gene clusters, with as many as ten missing genes in cognate groups 1 through 9 (Misof et al. 1996). Taken together, these findings are consistent with the hypothesis that larger gene networks can more readily tolerate deletions of individual genes. If these deletions had occurred shortly after duplications of *HOM/Hox* gene clusters, redundancy among cognates would probably have further facilitated such neutral deletions. More definitive conclusions clearly require the availability of linkage maps for these taxa. However, once such maps are established, it will be instructive to see whether the putative gene deletions follow a common pattern in different cognate groups, or whether they have occurred independently in each lineage.

HOM/Hox genes are not the only example of gene networks, but phylogenetic information and linkage data from a variety of taxa are needed before rigorous tests of qualitative model predictions can be made for any other candidate group of genes, especially given that the modeling approach used here is a statistical one, i.e., no particular deletion event is 'prohibited' by the model.

Acknowledgements. I would like to thank L. W. Buss, J. Kim, J. Reinitz, R. Vaisnys and G. P. Wagner for providing constructive criticism and

invaluable suggestions on earlier drafts of the manuscript. This paper is communication number 19 from the Centre for Computational Ecology at the Yale Institute for Biospheric Studies.

Appendix

Comparison of gene expression states before and after gene duplication

Only the case of gene duplications will be explicated here, because generalization to gene deletions is completely analogous. Consider a network with initial expression state $\vec{S}(0)$ that attains an equilibrium state $\vec{S}(\infty)$, and assume, without loss of generality, that genes G_1 through G_k are duplicated. The gene expression state of the network is then expanded according to a function π :

$$\pi: \{-1, +1\}^N \rightarrow \{-1, +1\}^{N+k}$$

$$(S_1, \dots, S_k, S_{k+1}, \dots, S_N) \rightarrow (S_1, S_1, \dots, S_k, S_k, S_{k+1}, \dots, S_N)$$

The initial state of the network after duplications becomes $\pi[\vec{S}(0)]$. This new initial state together with the changed matrix w determine the dynamics of (2) after duplication. If an equilibrium state is attained by this network, it is called $\vec{S}^{du}(\infty)$. In general $\vec{S}^{du}(\infty) \neq \pi[\vec{S}(\infty)]$. Now define π^{-1} as

$$\pi^{-1}: \{-1, +1\}^{N+k} \rightarrow \{-1, +1\}^N$$

$$(S_1, S'_1, \dots, S_k, S'_k, S_{k+1}, \dots, S_N)$$

$$\rightarrow (\sigma[S_1 + S'_1 + 1], \dots, \sigma[S_k + S'_k + 1], S_{k+1}, \dots, S_N)$$

Primed variables denote possibly different expression states of original gene and duplicate. $\sigma[S_i + S'_i + 1]$ is a Boolean OR function. It corresponds to the assumption that at least one copy of the duplicated gene has to be expressed in the equilibrium state in order to say that the corresponding biochemical 'function' is expressed. It will be said that a gene duplication event has no effect on the network, or that it does not change the expression pattern generated by the network, iff $\pi^{-1}[\vec{S}^{du}(\infty)] = \vec{S}(\infty)$.

Dependence of redundancy on the fraction, p , of expressed genes

Consider first the case of N gene duplications, and denote as $r_N(\infty)$ the likelihood that a gene deletion occurring an infinitely long time after the duplication event does not change $\vec{S}(\infty)$. Figures 4 and 5 indicate that, as p increases from $p = 0.5$, $r_N(\infty)$ remains unchanged. However, as p decreases from $p = 0.5$ and becomes very small, $r_N(\infty)$ increases, i.e., gene deletions become more likely to be neutral. Two biologically irrelevant, but analytically instructive cases, namely $p = 0$ and $p = 1$, help to understand this behavior. In these cases, it is trivial to characterize the connectivity matrices of optimal networks because $\vec{S}(\infty) = \vec{S}(\tau) = \sigma[h_i(0)]$. The dynamical system (2) is reduced to a set of linear inequalities for entries of w . Consider first the case of $p = 0$. It follows from (2), that in all members of the population, and for

any original/duplicate pair G_i and G_i^* , the respective pair of 'inputs', $h_i(0)$ and $h_i^*(0)$, have negative sign. Now note that in the case of N duplicated genes, the sign of any $h_i(0)$ and $h_i^*(0)$ does not change as a consequence of a gene deletion. Biologically speaking, this means that the deletion event does not cause changes in gene dosages relative to other network genes. Thus, gene deletions will not affect the equilibrium gene expression pattern at all. No deletion will cause genes to be expressed if they were not expressed before deletion. The case $p = 1$ is fundamentally different. If $p = 1$, $S_i(\infty) = 1$ for all i , but for each i it is only required that one of the two genes, original or copy, be expressed in the equilibrium state. Thus, the situation is more complicated, in that either of the following three conditions can hold: (i) $h_i(0) > 0$, $h_i^*(0) > 0$; (ii) $h_i(0) > 0$, $h_i^*(0) < 0$; (iii) $h_i(0) < 0$, $h_i^*(0) > 0$. If N genes are deleted, the signs of individual $h_i(0)$ values are unaffected. Thus, for a gene G_i , for which case (i) holds, deletion will have no effect. However, if condition (ii) holds when G_i is deleted, the remaining gene, G_i^* , will not be expressed in the equilibrium state. Similarly, if (iii) holds and G_i^* is deleted, the remaining gene, G_i , will not be expressed. The latter two situations will alter the (previously optimal) gene expression pattern. If evolving populations are large, and if many mutations have accumulated between duplication and deletion, the probability of a deletion of N genes having no effect on the expression pattern is given by $(\frac{1}{3} + \frac{2}{3}\frac{1}{2})^N = \frac{2}{3}^N$. For $N = 10$ genes, it follows that $r_N(\infty) \approx 0.017$, far from $r_N(\infty) = 1$ for $p = 0$.

Note that analogous arguments cannot be applied to r_1 : there is always a finite probability that changes in $h_i(t)$ occur after deletion. Thus, even in the extreme cases of $p = 0$ or $p = 1$, deletion of one gene will not be neutral.

References

- Amit DJ (1989) Modeling brain function. Cambridge University Press, Cambridge
- Bienz M, Tremml G (1988) Domain of *Ultrabithorax* expression in *Drosophila* visceral mesoderm from autoregulation and exclusion. *Nature* 333:576–578
- Cartwright P, Dick M, Buss LW (1993) *HOM/Hox* type homeoboxes in the chelicerate *Limulus polyphemus*. *Mol Phylogenet Evol* 2:185–192
- Chisaka O, Capecchi MR (1991) Regionally restricted developmental defects result from targeted disruption of the mouse homeobox gene *hox-1.5*. *Nature* 350:473–479
- Dawes R, Dawson I, Falciani F, Tear G, Akam M (1994) *Dax*, a locust Hox gene related to *fushi-tarazu* but showing no pair rule expression. *Development* 120:1561–1572
- Ferrier DEK, Akam ME (1994) The *HOM/Hox* cluster of the locust. *Proceedings of Spring Meeting of the British Society for Developmental Biology*, 5–8 April 1994, Edinburgh, UK
- Gatherer D (1993) Gene knockouts and murine development. *Dev Growth Differ* 35:365–370
- González-Reyes A, Morata G (1990) The developmental effect of overexpressing a *Ubx* product in *Drosophila* embryos is dependent on its interaction with other homeotic products. *Cell* 61:515–522
- Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer, Sunderland, Mass
- Hopfield JJ, Tank DW (1986) Computing with neural circuits: a model. *Science* 233:625–633
- Ingham W (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* 335:25–32

- Johnson PF, McKnight SL (1989) Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* 58:799–839
- Joyner AL, Herrup K, Auerbach BA, Davis CA, Rossant J (1991) Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the *en-2* homeobox. *Science* 251:1239–1243
- Kappen C, Schughart K, Ruddle FH (1989) Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. *Proc Natl Acad Sci USA* 86:5459–5463
- Kauffman SA (1969) Metabolic stability and epigenesis in randomly connected nets. *J Theor Biol* 22:437
- Kauffman SA (1993) *The origins of order*. Oxford University Press, New York
- Krumlauf R (1994) Hox genes in vertebrate development. *Cell* 78:191–201
- Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570
- Li X, Noll M (1994) Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature* 367:83–87
- McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. *Cell* 68:283–302
- McGinnis W, Kuziora MA, McGinnis W (1990) Human *hox-4.2* and *Drosophila deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63:969–976
- Mermelstein FH, Flores O, Reinberg D (1989) Initiation of transcription by RNA polymerase II. *Biochim Biophys Acta* 1009:1–10
- Misof BY, Wagner GP (1996) Evidence for four Hox clusters in the Killifish, *Fundulus heteroclitus* (Teleostei). *Mol Phylogenet Evol* (in press)
- Misof BY, Blanco MJ, Wagner GP (1996) Hox genes of the Zebrafish. *Genomics* (submitted)
- Mjolsness E, Sharp DH, Reinitz J (1991) A connectionist model of development. *J Theor Biol* 152:429–453
- Murtha MM, Leckman JF, Ruddle FH (1991) Detection of homeobox genes in development and evolution. *Proc Natl Acad Sci USA* 88:10711–10715
- Nusse R, Varmus HE (1992) *Wnt* genes. *Cell* 69:1073–1087
- Olson E (1990) MyoD family: a paradigm for development. *Genes Dev* 4:1454–1461
- Reinitz J, Sharp D (1995) Mechanism of formation of *eve* stripes. *Mech Dev* 49:133–158
- Reinitz J, Mjolsness E, Sharp D (1995) Cooperative control of positional information in *Drosophila* by *bicoid* and maternal *hunchback*. *J Exp Zool* 271:47–56
- Rendel JM (1979) Canalisation and selection. In: Thompson JN Jr, Thoday JM (eds) *Quantitative genetic variation*. Academic Press, New York
- Rosenfeld MG (1991) POU-domain transcription factors: pou-er-ful developmental regulators. *Genes Dev* 5:897–907
- Schubert FR, Nieselt-Struwe K, Gruss P (1993) The Antennapedia-type homeobox genes have evolved from three precursors early in metazoan evolution. *Proc Natl Acad Sci USA* 90:143–147
- Schughart K, Kappen C, Ruddle FH (1989) Duplication of large genomic region during the evolution of vertebrate homeobox genes. *Proc Natl Acad Sci USA* 86:7067–7071
- Scott MP (1992) Vertebrate homeobox gene nomenclature. *Cell* 71:551–553
- Seeger MA, Kaufman TC (1987) Homeotic genes of the Antennapedia complex (ANT-C) and their molecular variability in the phylogeny of the Drosophilidae. In: Raff RA, Raff EA (eds) *Development as an evolutionary process*. Alan Press, New York, pp 179–202
- Sommer R, Tautz D (1991) Segmentation gene expression in the housefly *Musca domestica*. *Development* 113:419–430
- Stuart JJ, Brown SJ, Beeman RW, Denell RE (1991) A deficiency of the homeotic complex of the beetle *Tribolium*. *Nature* 350:71–74
- Sucov HM, Murakami KK, Evans RM (1990) Characterization of an autoregulated response element in the mouse retinoic receptor type β gene. *Proc Natl Acad Sci USA* 87:5392–5396
- Tautz D (1992) Redundancies, development and the flow of information. *BioEssays* 14:263–266
- Thomas JH (1993) Thinking about genetic redundancy. *Trends Genet* 9:395–398
- Turelli M (1985) Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics* 111:165–195
- Wagner A (1994) Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci USA* 91:4387–4391
- Wagner A (1996) Does evolutionary plasticity evolve? *Evolution* (in press)
- Wagner A, Wagner GP, Similon P (1994) Epistasis can facilitate the evolution of reproductive isolation by peak shifts: a two-locus two-allele model. *Genetics* 138:533–545
- Walldorf U, Fleig R, Gehring WJ (1989) Comparison of the homeobox containing genes of the honeybee and *Drosophila*. *Proc Natl Acad Sci USA* 86:9971–9975
- Zeng Z-B, Cockerham CC (1993) Mutation models and quantitative genetic variation. *Genetics* 133:729–736