Copyright © 2001 by the Genetics Society of America

The Probability of Preservation of a Newly Arisen Gene Duplicate

Michael Lynch,* Martin O'Hely,† Bruce Walsh‡ and Allan Force§

**Department of Biology, Indiana University, Bloomington, Indiana 47405,* †*Department of Integrative Biology, University of California, Berkeley, California 94720,* ‡*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721 and* §*Virginia Mason Research Center, Benaroya Research Institute, Seattle, Washington 98101*

Manuscript received April 2, 2001
Accepted for publication August 27, 2001

ABSTRACT

Newly emerging data from genome sequencing projects suggest that gene duplication, often accompanied by genetic map changes, is a common and ongoing feature of all genomes. This raises the possibility that differential expansion/contraction of various genomic sequences may be just as important a mechanism of phenotypic evolution as changes at the nucleotide level. However, the population-genetic mechanisms responsible for the success *vs.* failure of newly arisen gene duplicates are poorly understood. We examine the influence of various aspects of gene structure, mutation rates, degree of linkage, and population size (N) on the joint fate of a newly arisen duplicate gene and its ancestral locus. Unless there is active selection against duplicate genes, the probability of permanent establishment of such genes is usually no less than $1/(4N)$ (half of the neutral expectation), and it can be orders of magnitude greater if neofunctionalizing mutations are common. The probability of a map change (reassignment of a key function of an ancestral locus to a new chromosomal location) induced by a newly arisen duplicate is also generally $>1/(4N)$ for unlinked duplicates, suggesting that recurrent gene duplication and alternative silencing may be a common mechanism for generating microchromosomal rearrangements responsible for postreproductive isolating barriers among species. Relative to subfunctionalization, neofunctionalization is expected to become a progressively more important mechanism of duplicate-gene preservation in populations with increasing size. However, even in large populations, the probability of neofunctionalization scales only with the square of the selective advantage. Tight linkage also influences the probability of duplicate-gene preservation, increasing the probability of subfunctionalization but decreasing the probability of neofunctionalization.

FOSTERED in part by the belief that gene duplication is a major contributor to the origin of evolutionary novelties, substantial theoretical and empirical attention has been given to the evolutionary fates of gene duplicates. The traditional view has been that a gene duplicate will ultimately suffer one of two fates: either one copy will be silenced by degenerative mutations (nonfunctionalization) or one copy will evolve a new beneficial function (neofunctionalization) that permanently preserves it in the population (Haldane 1933; Fisher 1935; Ohno 1970; Nei and Roychoudhury 1973; Christiansen and Frydenberg 1977; Bailey *et al.* 1978; Takahata and Maruyama 1979; Li 1980; Watterson 1983; Walsh 1995). Under this model, the alternative copy always retains the original function. However, a third possible fate has recently been recognized: both copies may be reciprocally preserved through the fixation of complementary loss-of-subfunction mutations (subfunctionalization), which results in a partitioning of the tasks of the ancestral gene (Force *et al.* 1999; Lynch and Force 2000a; Stoltzfus 2000; Wagner 2000). Such a partitioning of ancestral-gene tasks may also be driven by a form of positive Darwinian selection, the acquisition of copy-specific mutational refinements to alternative gene subfunctions previously kept at suboptimal levels by pleiotropic constraints (Piatigorsky and Wistow 1991; Hughes 1994). Finally, it has been suggested that redundancy may be directly advantageous as a mechanism for minimizing the phenotypic effects of null alleles and/or developmental accidents (Clark 1994; Nowak *et al.* 1997; Krakauer and Nowak 1999; Wagner 1999).

As pointed out by Spofford (1969), a significant gap in our understanding of gene duplication concerns the critical initial phase during which a single copy of a duplicated gene must rise to a high enough frequency in the population to become subject to the mutational processes noted above. Almost all of the existing theory for the evolution of duplicate genes starts with the assumption that all members of the base population carry two fully functional genes at both loci. This is perhaps a reasonable scenario for a newly established polyploid species, but an alternative approach is required to explain the establishment of single-gene duplicates originating by more common processes such as replicative translocation or tandem duplication.

Our focus is on the ultimate fate of a pair of duplicate

Corresponding author: Michael Lynch, Department of Biology, Indiana University, Bloomington, IN 47405.
E-mail: mlynch@bio.indiana.edu

loci, one of which (the ancestral copy) carries active alleles in all members of the population and the other of which (the descendant copy) is initially represented by a single gene in a single (heterozygous) individual, all other individuals at this latter locus being effectively null homozygotes. We restrict our attention to whole-gene duplication, so that processed pseudogenes or partial duplications are not considered, and we assume that there is no intrinsic disadvantage to duplicates as might arise if gene-dosage issues were important. Given these starting conditions, several potential outcomes can be envisioned:

First, as with any newly arisen mutation, there is a high probability that the new copy will be rapidly lost by random genetic drift. If there is no selective advantage for the new copy, this probability will be equal to $\lambda = 1 - [(1/(2N)]$, where N denotes the population size. Upon such an outcome, all evidence of the duplication event will be eliminated from the population.

Second, in the rare event that the new duplicate rises to high frequency, it may randomly accumulate a higher load of degenerative mutations than the ancestral copy and in the absence of any selective advantage may eventually become nonfunctionalized. In this case, the ancestral gene copy is permanently retained, while a semipermanent record of the duplication event may transiently remain in the form of a pseudogene.

Third, if functional alleles rise by chance to high frequency at the new duplicate locus, it is possible that the ancestral copy will become a nonfunctional pseudogene. In this case, the population is again returned to the single-gene state of the ancestral population, but the genomic location of the functional gene will have changed (Haldane 1933; Walsh 1995).

Finally, both copies of the locus may become permanently preserved either by subfunctionalization, with each copy carrying out a unique set of subfunctions (or both being mutationally reduced to the level of expression of the single-copy ancestral gene), or by neofunctionalization, with one copy evolving a new beneficial function at the expense of the original function (which is retained by the other copy). A change in map position will result if the two loci become subfunctionalized or if the original locus becomes neofunctionalized.

The evolutionary outcome of a gene-duplication event relates to three issues of potentially broad evolutionary significance. First, the mechanisms by which gene duplicates become permanently preserved have a bearing on the evolutionary potential of a species. For example, a neofunctionalizing mutation is equivalent to the origin of an evolutionary novelty, while subfunctionalizing mutations can provide new evolutionary flexibility by releasing an ancestral gene from pleiotropic constraints. We refer to the probability that a newly arisen gene duplicate becomes permanently preserved as Θ. Second, complete or partial silencing of an ancestral gene results in chromosomal repatterning, equivalent to a change in the genetic map, assuming the loci are not completely linked. Such changes are of relevance to the speciation process, as they passively induce postzygotic genomic incompatibilities in hybrid progeny (Werth and Windham 1991; Lynch and Force 2000b). We refer to the probability that a newly arisen gene duplicate induces a map change as Δ. Third, if duplicate genes become fixed in a population more frequently than their parental loci are lost, an expansion of the genome must occur. We refer to the probability that a newly arisen gene duplicate results in a permanent expansion of the genome size as Γ. This is equivalent to the probability of joint preservation of a pair of duplicates.

The development of a comprehensive theory for the evolution of duplicate genes raises formidable technical difficulties because the process involves two multiallelic loci with epistatic interactions. We have been successful in deriving some analytical approximations that help provide insight into the mechanisms governing the dynamics of duplicate-gene evolution, but to establish the validity of the theory it has also been necessary to rely extensively on computer simulations.

PRESERVATION BY DEGENERATIVE MUTATIONS

The situation in which mutations to novel beneficial functions are sufficiently rare to be ignored provides a useful null model for interpreting the fates of duplicate genes because the evolutionary dynamics are governed entirely by random genetic drift and degenerative mutation. Under this model, a newly arisen gene duplicate has three possible fates: (1) The new copy may simply be lost by random genetic drift and/or silenced by the accumulation of degenerative mutations; (2) the new copy may become permanently fixed in the population, with the original locus subsequently being silenced by degenerative mutations; or (3) both loci may become mutually preserved by subfunctionalization (Figure 1). The probability of preservation of the duplicate gene and, in the case of unlinked duplicates, the probability of a map change are equal to the sum of probabilities of fates 2 and 3, while the rate of genome expansion is equal to the probability of fate 3. To accommodate the fact that all of these probabilities decline rapidly with increasing N [because the probability of initial establishment is on the order of $1/(2N)$], we scale the three summary statistics (Θ, Δ, and Γ) by multiplying by $2N$. Letting $P_{\text{non,o}}$ denote the probability of silencing of the original locus and P_{sub} denote the probability of subfunctionalization,

$$\Theta = \Delta = 2N(P_{\text{non,o}} + P_{\text{sub}}) \tag{1a}$$

and

$$\Gamma = 2NP_{\text{sub}}. \tag{1b}$$

With this scaling, $\Theta = 1$ implies that the probability of preservation of a newly arisen gene duplicate is equivalent to the rate of fixation of a neutral mutation, $1/(2N)$.

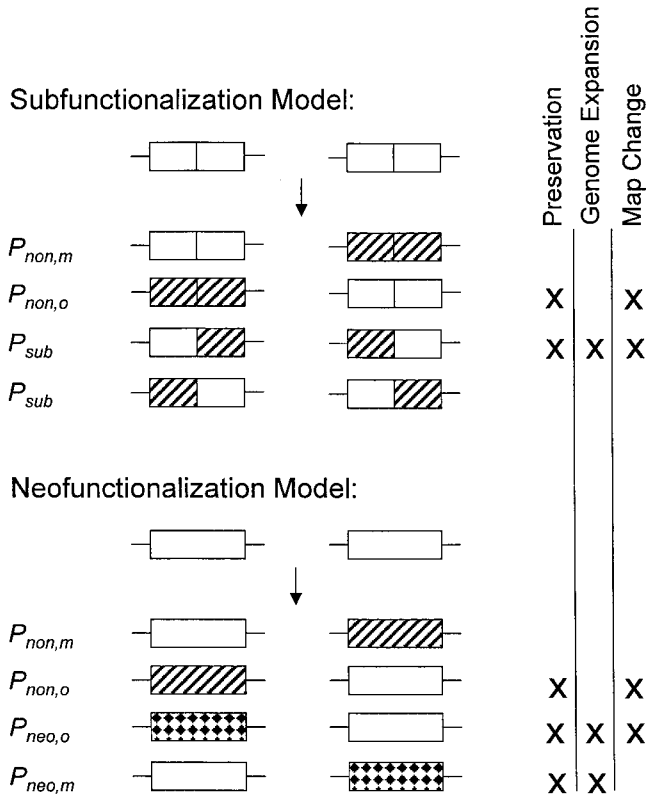

FIGURE 1.—Schematic for the alternative stable outcomes of the gene-duplication process for the subfunctionalization and neofunctionalization models. For both cases, the ancestral gene is on the left and the newly arisen duplicate is on the right. For the subfunctionalization model, the gene is divided into two sections, each one denoting an independently mutable subfunction. Diagonal lines denote loss of function or subfunction; diamonds denote neofunctionalization (with an accompanying loss of the original function). The probabilities of the alternative fates are listed on the left: non, nonfunctionalization; sub, subfunctionalization; neo, neofunctionalization; and o and m, the original and newly arisen locus, respectively. The genomic consequences of the various fates are marked on the right.

Definitions of these and all additional terms associated with this model are summarized in Table 1.

As in most other theoretical investigations of the evolution of duplicate genes, we initially consider the double-null recessive model, whereby all two-locus genotypes have equal fitness except for the inviable double-null homozygotes that completely lack a particular function (or subfunction). Nonfunctionalizing mutations, which eliminate all gene function, arise at each locus at rate μ_c per gene copy per generation, and, when a gene has independently mutable subfunctions, each subfunction is subject to silencing at rate μ_r. We restrict our attention to the situation in which genes have either a single function (in which case $\mu_r = 0$) or two independently mutable subfunctions (each with the same μ_r). Such subfunctions may be physically defined in a number of ways, including tissue-specific regulatory elements, alternative functional domains of a protein, and/or alternative splice variants. We consider the two extreme situations in which the duplicate loci are either completely linked (*i.e.*, a tandem pair) or freely recombining.

As there is no reason to expect the mutation process to be altered upon gene duplication, we assume that the initial locus has allele frequencies expected under selection-mutation-drift equilibrium prior to duplication. The new locus is then randomly initiated with a single copy of either a fully functional allele or a subfunctional allele, with the probabilities of initial status being defined by the relative equilibrium frequencies of the classes of active alleles at the original locus. We also assume that the founding allele for the new locus is carried initially in a gamete containing its ancestral type at the original locus. In the case of complete linkage, because a duplicate is permanently associated with its parental source, a newly arisen subfunctional gene cannot proceed to fixation, as this would result in the loss of the alternative subfunction. In the case of free recombination, the ancestral locus is guaranteed to be preserved in the event the new locus is founded by a subfunctional allele.

It is well known that the equilibrium frequency of a recessive lethal (nonfunctional) allele for a gene with a single function is $\sqrt{\mu_c}$ in large populations ($N\mu_c > 1$), and this frequency declines in smaller populations (Figure 2). The equilibrium frequency of nonfunctional alleles is reduced when genes have independently mutable subfunctions, but this is more than offset by the frequency of subfunctional alleles (Figure 2). For example, at large N with $\mu_c = \mu_r = 10^{-5}$, each of the two types of subfunctional alleles have equilibrium frequencies of 0.0025, while the null allele has frequency 0.0015. Thus, provided $N > 10^3$, some subfunctional alleles are expected to be segregating at the initial locus unless $\mu_r \ll \mu_c$.

To evaluate the probabilities of the three alternative fates ($P_{non,o}$, $P_{non,m}$, and P_{sub}) under this model over a range of population sizes, we performed stochastic simulations of a gamete-based model, which we have previously shown to yield equivalent results to individual-based simulations (LYNCH and FORCE 2000a). An effectively infinite gamete pool is assumed so that recombination and mutation can be treated as deterministic processes. Given the expected frequencies of gamete types in any generation, the expected frequencies of zygote genotypes after random mating and selection are determined, and then the actual zygote frequencies are obtained by random sampling of N genotypes. This cycle of events is continued until the final fate of the pair of duplicates has been determined, *i.e.*, when either one locus completely lacks functional alleles (nonfunctionalization) or when each locus has completely lost a unique subfunction (subfunctionalization). For any set of mutational parameters, we typically performed enough simulations so that at least 2500 runs would lead to the gene duplicate becoming well-established in the population by random genetic drift. This required as many as 10^9 replicate runs

TABLE 1

Terms associated with the model incorporating only degenerative mutations

N	Effective population size, assumed to be equal to actual population size.
λ	$1 - [1/(2N)]$.
Θ	Probability that the new duplicate is permanently preserved ($\times 2N$).
Γ	Probability that both loci are jointly preserved ($\times 2N$).
Δ	Probability that a key (sub)function of the ancestral locus is reassigned to a new genomic location ($\times 2N$).
μ_r	Rate of mutations eliminating single subfunctions.
μ_c	Rate of mutations eliminating total gene function.
α	$\mu_r/(\mu_c + 2\mu_r)$.
p_f	Initial frequency of fully functional alleles at the ancestral locus.
$P_{non,o}$	Probability that the original locus is silenced.
$P_{non,m}$	Probability that the descendant locus is silenced.
P_{sub}	Probability that the two loci are preserved by subfunctionalization.
$P'_{non,o}$	Probability that the original locus is silenced, conditional upon prior fixation of the new unlinked duplicate.
P'_{sub}	Probability that the two loci are preserved by subfunctionalization, conditional upon prior fixation of the new unlinked duplicate.
$P'_{sub,f}$	Probability that two unlinked duplicates are preserved by subfunctionalization, conditional upon the new locus being founded by a functional allele.
$P'_{sub,s}$	Probability that two unlinked duplicates are preserved by subfunctionalization, conditional upon the new locus being founded by a subfunctional allele.
P_0	Probability that a copy of a functional allele at a new unlinked locus remains intact after $4N$ generations.
P_1	Probability that a copy of a functional allele at a new unlinked locus has lost a single subfunction after $4N$ generations.
P_2	Probability that a copy of a subfunctional allele at a new unlinked locus remains intact after $4N$ generations.

at large N, and we employed no fewer than 5×10^6 runs at small N.

Linked loci: Cases of absolute linkage can be treated formally as a single-locus model, and in this case we refer to a linked pair of duplicates as a two-copy allele.

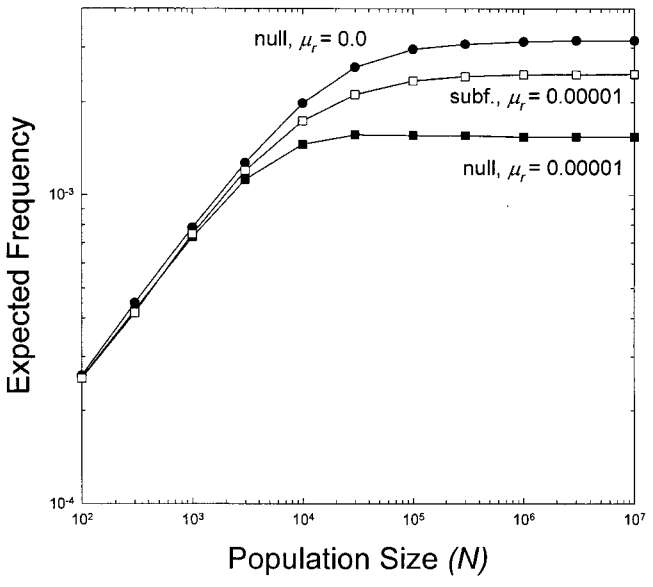

FIGURE 2.—Expected equilibrium frequencies of null and subfunctional alleles at the initial locus at various population sizes, under drift-mutation-selection balance. Results were obtained by computer simulation with the mutation rate to nulls being $\mu_c = 10^{-5}$ and the gene either having a single function ($\mu_r = 0$) or two independently mutable subfunctions with $\mu_r = 10^{-5}$. In the latter case, each of the two possible types of subfunctional alleles has expected frequencies equal to the plotted values.

Functional two-copy alleles have a slight selective advantage over their single-copy counterparts during the initial phase of establishment because single-copy alleles that experience either subfunctionalizing or nonfunctionalizing mutations can never go to fixation, whereas a mutated two-copy allele can fix as long as the two component genes cover all subfunctions. In small populations, this advantage is negligible because the two-copy allele is either lost or fixed by random genetic drift before a significant probability of mutation has accrued, and the probability that the new duplicate initially drifts to fixation is very close to its initial frequency, $1/(2N)$. Letting $P'_{non,o}$ and P'_{sub} denote the subsequent fate probabilities conditional on the two-copy allele having become established, then because nonfunctionalization will occur randomly at one locus or the other, $P'_{non,o} = (1 - P'_{sub})/2$, and

$$\Theta = 2N \cdot \frac{1}{2N} \cdot \left(\frac{1 - P'_{sub}}{2} + P'_{sub} \right) = \frac{1 + P'_{sub}}{2}, \quad (2a)$$

$$\Gamma = P'_{sub}. \quad (2b)$$

To obtain an expression for P'_{sub}, we note that the probability that the first mutation to be fixed in a two-copy lineage is of a subfunctionalizing type is $2\mu_r/(\mu_c + 2\mu_r)$. Conditional on this occurring, joint preservation of the two genes by subfunctionalization is expected to occur with probability $\alpha = \mu_r/(\mu_c + 2\mu_r)$, because following the loss of one subfunction from one locus, the subfunctional locus is still free to fix subsequent mutations at rate $\mu_r + \mu_c$ (resulting in nonfunctionalization), while the intact locus may only fix a mutation for the alternative subfunction (at rate μ_r, resulting in subfunctionali-

zation; Force *et al.* 1999). Thus, for small N, we expect $P'_{sub} \simeq 2\alpha^2$, and hence $\Theta \simeq 0.5 + \alpha^2$ and $\Gamma \simeq 2\alpha^2$.

With increasing population size, there is an increasing probability that single-copy alleles will mutate during the long sojourn of a two-copy allele through the population, putting the former at a slight selective disadvantage. Consider, for example, the case of genes with a single function. At the limit as $N \to \infty$, the expected frequency of descendants of the initial two-copy gene among the total pool of functional genes increases from the initial level of $1/(2N)$ to a stable level of $1/N$ (appendix). This transient behavior occurs because the initial mutations experienced by two-copy alleles are completely neutral, which causes their descendants to increase at the expense of one-copy alleles. The increase continues until all two-copy alleles have acquired a mutation in at least one copy, at which point they are selectively equivalent to functional single-copy alleles. These results suggest that at large N a completely linked pair of duplicate genes (in this case, assumed to be incapable of subfunctionalization or neofunctionalization) will fix with probability $1/N$, with a random member of the pair becoming silenced, which further implies $\Theta \to 2N \cdot (1/N) \cdot 0.5 = 1.0$ as $N \to \infty$. The temporal dynamics outlined in the appendix suggest that this large-population approximation should apply provided $N\mu_c > 2$. Using the approach outlined in the appendix, after considerable analysis, we also obtained results that suggest that $\Theta \to 1.0$ as $N \to \infty$ when there are two independently mutable subfunctions.

The preceding analytical approximations are in close agreement with observations from computer simulations (Figures 3 and 4). At small N, $\alpha = 0.0$ when there is only a single-gene function, yielding $\Theta \simeq 0.5$ and $\Gamma = 0$, whereas $\alpha = 0.333$ when $\mu_r = \mu_c$, yielding $\Theta \simeq 0.611$ and $\Gamma \simeq 0.222$. As $N \to \infty$, $\Theta \to 1.0$ under the conditions of one or two subfunctions, and $\Gamma \to 0$.

Unlinked duplicates: For freely recombining loci, the selective advantage of a newly arisen duplicate is negligible due to the fact that it does not remain associated with a functional partner. The key issue then becomes whether the newly arisen gene is capable of drifting to fixation in an intact state. As pointed out in Lynch and Force (2000a), the probability of subfunctionalization of unlinked duplicates declines with increasing population size because the accumulation of secondary mutations can eventually silence a subfunctional allele during the long ($\sim 4N$ generation; Kimura and Ohta 1969) sojourn to fixation. To account for this behavior, we present the following approximations, first for a fully functional newborn gene duplicate and then for a subfunctional newborn.

Under the assumption of negligible selection, an initially fully functional allele retains full functionality after $4N$ generations with probability

$$P_0 = e^{-4N(\mu_c + 2\mu_r)} \tag{3}$$

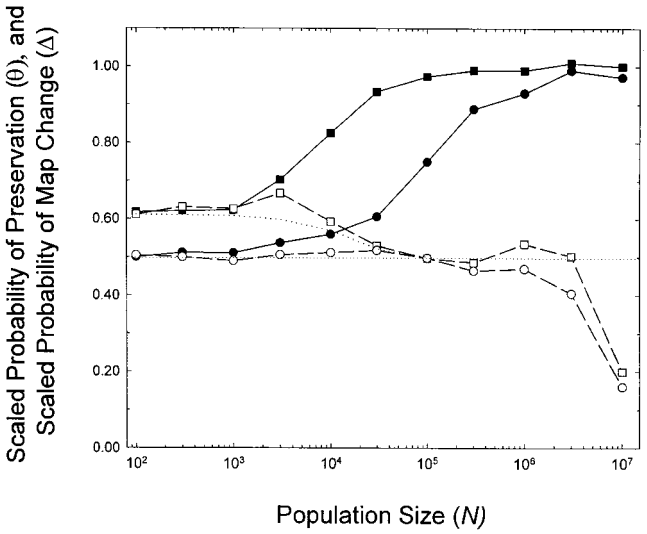

Figure 3.—The scaled probability of preservation of a duplicate gene (also equal to the scaled probability of a map change) for the situation in which the rate of mutation to novel functions is negligible. Open and solid symbols denote results for freely recombining and completely linked loci, respectively. Squares denote the results for the situation in which there are two independently mutable subfunctions, each with mutation rate $\mu_r = 10^{-5}$, and the circles denote the case in which there is a single function ($\mu_r = 0$). In both cases, the rate of origin of mutations that eliminate all function is $\mu_c = 10^{-5}$. The dotted lines denote the analytical approximations for the case of unlinked genes obtained by use of Equations 2a, 3, 4, 6, and 8.

(again, assuming two independently mutable subfunctions) and will have lost a single subfunction with probability

$$P_1 = 2(1 - e^{-4N\mu_r})\, e^{-4N(\mu_c + \mu_r)}. \tag{4}$$

Having reached the latter state (with the original locus still intact), joint preservation of the two loci by subfunctionalization will occur with probability α, following the logic outlined above. Noting that subsequent fixation events are expected to occur approximately every $4N$ generations on average and that $P_1 P_0^{t-1}$ is the probability that an initially intact gene has lost a single subfunction $4Nt$ generations following fixation, then the probability of subfunctionalization, conditional on the initial establishment of a duplicate, is

$$P'_{sub,f} = \alpha P_1 \sum_{t=0}^{\infty} P_0^t = \frac{\alpha P_1}{1 - P_0}. \tag{5}$$

If, on the other hand, the newly arisen duplicate is a copy of a subfunctional allele, then the probability that it is intact after the expected $4N$ generations required for establishment is

$$P_2 = e^{-4N(\mu_c + \mu_r)}, \tag{6}$$

and

$$P'_{sub,s} = \alpha P_2 \tag{7}$$

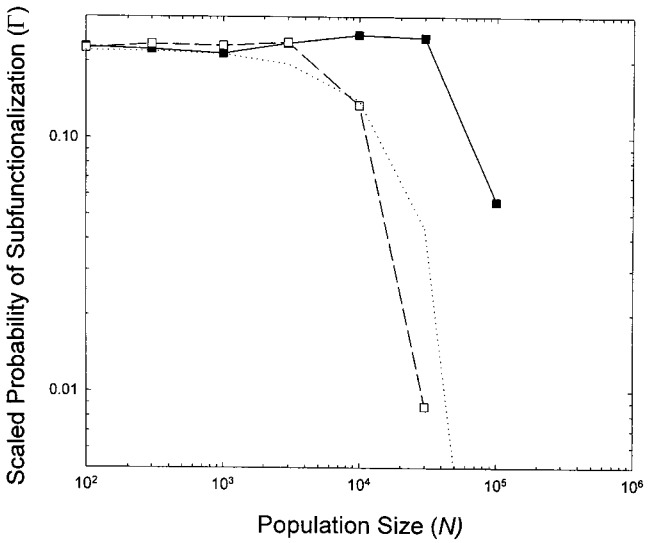

FIGURE 4.—The scaled probability of duplicate-gene preservation by subfunctionalization for the situation in which there are two independently mutable subfunctions and the rate of mutation to novel functions is negligible. Open and solid symbols denote results for freely recombining and completely linked loci, respectively. The mutation rates are $\mu_r = \mu_c = 10^{-5}$. The dotted line denotes the analytical approximation for the case of freely recombining loci, obtained by use of Equations 2b, 3, 4, 6, and 8.

is the conditional probability of subfunctionalization. Letting p_f denote the expected initial frequency of the fully functional allele at the original locus, then the weighted conditional probability of subfunctionalization is

$$P'_{sub} = \alpha[(p_f P_1/(1 - P_0)) + ((1 - p_f)P_2)]. \quad (8)$$

For small N, $p_f \simeq 1$ and $P_1/(1 - P_0) \rightarrow 2\alpha$, yielding $P'_{sub} \simeq 2\alpha^2$, and from Equations 2a and 2b, $\Theta = \Delta \simeq 0.5 + \alpha^2$ and $\Gamma \simeq 2\alpha^2$. These results are identical to the expectations for linked duplicates. As $N \rightarrow \infty$, $P'_{sub} \rightarrow 0$, implying $\Theta = \Delta \rightarrow 0.5$ and $\Gamma \rightarrow 0$. This suggests that the probability of duplicate-gene preservation at large N is twofold lower in unlinked than in linked duplicates.

Provided $N\mu_c < 10$, these analytical approximations for unlinked duplicates yield results that are quite compatible with those obtained by computer simulation (Figures 3 and 4). There are three fairly distinct regions of response to increasing N. First, for $N\mu_c \ll 1$, $\Theta = \Delta \simeq 0.5 + \alpha^2$ and $\Gamma \simeq 2\alpha^2$ as predicted by the theory for small N. Second, for $1 < N\mu_c < 10$, $\Theta = \Delta \simeq 0.5$ and $\Gamma \simeq 0$ as predicted by the theory for large N. Third, as $N\mu_c$ increases beyond 10, $\Theta = \Delta$ gradually approaches zero. Although this latter phase is unaccounted for by the theory, it presumably occurs because when $N\mu_c > 1$ there is a significant probability that all of the descendants of a newly arisen duplicate become silenced by mutations prior to the initial establishment of the lineage. In any event, contrary to the situation for linked duplicates, the probability of preservation of unlinked duplicates declines with increasing population size, although, provided $N\mu_c < 10$, this probability still equals or exceeds $1/4N$.

PRESERVATION BY NEOFUNCTIONALIZATION

We now consider the situation in which mutations with phenotypic effects either silence a gene or introduce a new beneficial function at the expense of the original function (Figure 1). The fitness landscape is assumed to be one in which individuals that carry no alleles with the original function have zero fitness, with the remaining genotypes having fitnesses equal to $1 + ns$, where $n = 0$, 1, 2, or 3 is the number of neofunctional alleles carried. Silencing mutations are assumed to arise at rate μ_c per gene copy for both types of active alleles, whereas alleles of the "ancestral" type (hereafter referred to as wild type) can also mutate to the neofunctionalized state at rate μ_b.

To evaluate the probabilities of the alternative fates of a pair of duplicate loci subject to beneficial mutations, we employed a simulation approach identical in structure to that described in the previous section, starting with a single-copy locus with allele frequencies equal to the simulated expectations under selection-mutation-drift equilibrium. The newly arisen duplicate was initiated as a single copy randomly recruited from the pool of wild-type and neofunctional alleles at the original locus, and the generation-to-generation cycle of events was continued until the final fate of the pair of duplicates had been established. It is straightforward to identify nonfunctionalization as a final stable state, as this simply requires that one locus becomes fixed for null alleles. Identification of neofunctionalization as a fate is slightly more subjective because, in a finite population, there is always a very small possibility that a neofunctionalized locus may become lost in the future (because it carries a beneficial but nonessential function and is subject to nonfunctionalizing mutations). We considered neofunctionalization to have occurred when one locus had completely lost the wild-type allele and acquired a high enough frequency of the neofunctionalized allele to ensure a probability of fixation of the latter of at least 0.99. Using the diffusion approximation for the fixation probability of a beneficial allele with additive effects (KIMURA 1962), this critical frequency is equal to

$$p^* = -\frac{1}{4Ns}\ln[1 - 0.99(1 - e^{-4Ns})], \quad (9)$$

which for large Ns reduces to $p^* \simeq 1.15/(Ns)$. (For the case of completely linked duplicates, this critical frequency must be applied to pairs of two-copy alleles with one neofunctional and one wild-type member, because neofunctional single-copy genes cannot become fixed in the population.) In the simulations that we

TABLE 2

Additional terms associated with the neofunctionalization model

μ_b	Rate of origin of neofunctional mutations.
s	Selective advantage of a neofunctional allele at the original locus on a background containing at least one functional allele.
s_n	Initial marginal selective advantage of a neofunctional allele at the new locus.
s_f	Initial marginal selective advantage of a wild-type allele at the new locus.
$P_{neo,o}$	Probability that the original locus is neofunctionalized.
$P_{neo,m}$	Probability that the descendant locus is neofunctionalized.
$P'_{neo,o}$	Probability that the original locus is neofunctionalized, conditional upon prior fixation of the new linked duplicate.
$P'_{neo,m}$	Probability that the descendant locus is neofunctionalized, conditional upon prior fixation of the new linked duplicate.
$P''_{neo,o}(t)$	Probability that the original locus is neofunctionalized, conditional upon the new linked duplicate being initially destined for loss.
$P''_{neo,m}(t)$	Probability that the descendant locus is neofunctionalized, conditional upon the new linked duplicate being initially destined for loss.
$\hat{p}_n$	Expected frequency of neofunctional alleles at the ancestral locus under selection-mutation-drift balance.
$\hat{p}_0$	Expected frequency of null alleles at the ancestral locus under selection-mutation-drift balance.
p_n	$\hat{p}_n/(1 - \hat{p}_0)$
p_f	$1 - p_n$
u_F	Fixation probability for a beneficial mutation with initial frequency $1/(2N)$.
$u_F(s)$	Fixation probability for neofunctional alleles at the original locus, following the fixation of wild-type alleles at the new locus.
$u_F(s_f)$	Fixation probability for a new wild-type duplicate with initial frequency $1/(2N)$ and initial marginal fitness s_f.
$u_F(s_n)$	Fixation probability for a new neofunctional duplicate with initial frequency $1/(2N)$ and initial marginal fitness s_n.
β	$2Nu_F\mu_b/(\mu_c + 2Nu_F\mu_b)$
$n_m(t)$	Expected number of two-copy alleles in a population in generation t, conditional on not yet having been lost or rescued.
$u_L(t)$	Probability that a newly arisen locus has been lost by drift by generation t in a population assumed to be effectively infinite in size.
$r(t)$	Probability that a newly arisen locus, initially destined to be lost by drift, acquires a neofunctional mutation in generation t that carries it to fixation.
$\ell(t)$	Probability that the fate of a two-copy allele has not been determined by generation t.
$p_L(t)$	Probability that an effectively neutral allele destined to be lost by drift is lost in generation t.

performed, we assumed that the rate of mutation to neofunctional alleles (10^{-9} per gene per generation) is much smaller than the mutation rate to nulls (10^{-5} per gene per generation, as in the previous section), and s was 0.001, 0.01, or 0.1.

Under this model, a newly arisen gene duplicate can be regarded as preserved in the population if neofunctionalization occurs at either locus or if the original locus becomes nonfunctionalized. Thus, the scaled probability of preservation is

$$\Theta = 2N(P_{neo,m} + P_{neo,o} + P_{non,o}), \tag{10}$$

with the component terms being defined in Tables 1 and 2. For genes that are not completely linked, a map change occurs if the original locus becomes silenced or neofunctionalized, so the scaled probability of a map change is

$$\Delta = 2N(P_{neo,o} + P_{non,o}). \tag{11}$$

Finally, a new gene is added to the genome whenever one member of the pair is neofunctionalized, as this results in joint preservation of both copies. Hence,

$$\Gamma = 2N(P_{neo,o} + P_{neo,m}). \tag{12}$$

A key feature of this model of gene duplication is that the original locus (prior to duplication) can exhibit a balanced polymorphism due to the recurrent input of mutations and to heterozygote superiority. Although neofunctional alleles have zero fitness when in the homozygous state, they have a heterozygote advantage of s when associated with wild-type alleles. For large N, a set of standard recursion equations for allele frequencies (ignoring drift) yields the approximate equilibrium frequencies of the neofunctional (n) and null (0) alleles. For $\mu_c < [s/(1 + s)]^2$,

$$\hat{p}_n \simeq \frac{s^2 - \mu_c(1 + s)^2}{s(1 + 2s)}, \tag{13a}$$

$$\hat{p}_0 \simeq \frac{\mu_c(1 + s)}{s}, \tag{13b}$$

whereas for $\mu_c > [s/(1 + s)]^2$,

$$\hat{p}_n \simeq 0, \tag{14a}$$

$$\hat{p}_0 \simeq \sqrt{\mu_c}. \tag{14b}$$

These results, combined with observations from computer simulations (Figure 5), illustrate two key points. First, for sufficiently weak positive selection ($\mu_c > [s/(1 + s)]^2$), the mutation pressure against a neofunctional allele overwhelms the selective advantage, maintaining the frequency of neofunctional alleles at the

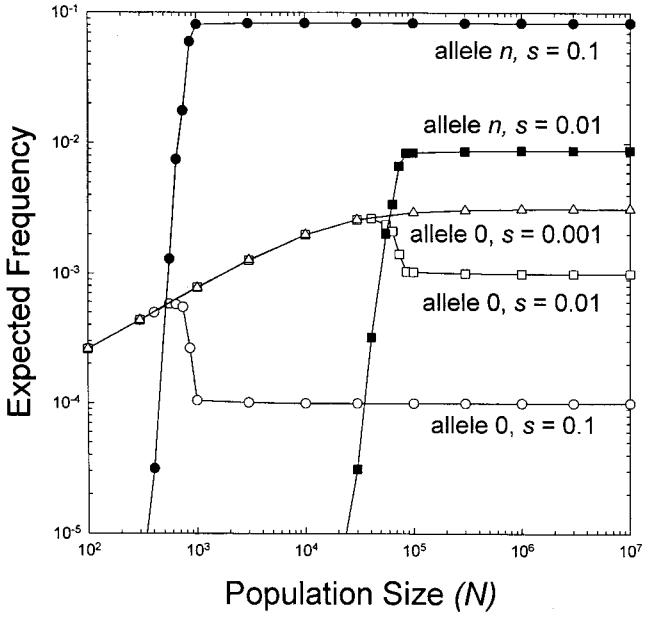

FIGURE 5.—Expected equilibrium frequencies of neofunctional (n) and nonfunctional (null, 0) alleles at the initial locus at various population sizes, under drift-mutation-selection balance, obtained by computer simulation.

original locus at negligible levels. For example, with $s = 0.001$ and $\mu_c = 10^{-5}$, $\hat{p}_n$ asymptotically approaches $\sim\mu_b/(2s) \simeq 5 \times 10^{-7}$ at large N. In this case, a new duplicate locus will almost always be initiated with a wild-type allele, and neofunctionalization will require mutation to new neofunctional alleles subsequent to the duplication process. Second, when selection is stronger ($\mu_c < [s/(1 + s)]^2$), the expected frequency of neofunctional alleles residing at the original locus is nearly a threshold function of population size, being closely approximated by Equation 13a, provided $Ns^2 > 4$, and rapidly dropping to negligible values ($<1/2N$) for N below the threshold. For example, as $N \rightarrow \infty$, with $\mu_c = 10^{-5}$, $\hat{p}_n \rightarrow 0.0088$ when $s = 0.01$, and $\hat{p}_n \rightarrow 0.083$ when $s = 0.1$. This means that at large population sizes with unlinked loci, neofunctionalization need not rely on the rare occurrence of beneficial mutations but can be poised to move forward if (1) the new locus is founded with a neofunctional allele or (2) the new locus is founded with a wild-type allele that subsequently acquires a sufficiently high frequency that the neofunctional alleles at the original locus become subject to directional, rather than balancing, selection.

Linked loci: In the case of complete linkage, a newly arisen gene duplicate must be of wild type to have any chance of permanent preservation, because under the assumptions of the model a linked pair of neofunctional genes is lethal in the homozygous state. So for linked duplicates, we considered only the case in which the initial duplicate carried the essential ancestral function. In this case, permanent preservation of both loci occurs when the founding two-copy allele goes to fixation and one member evolves a new function. This outcome yields a state of fixed heterozygosity, in the sense that each gamete carries one allele with the ancestral function and another with the new function (SPOFFORD 1969).

As noted above, the case of completely linked duplicates can be treated as a single-locus model with two classes of alleles, single copy and two copy. Ignoring the weak directional forces of selection, a newly arisen linked pair of gene duplicates (*i.e.*, a two-copy allele carrying only wild-type genes) will initially be destined to go to fixation with probability $1/(2N)$ and otherwise to become lost with probability λ. Should the two-copy allele proceed down the path toward fixation, one member of the pair will ultimately become either silenced or neofunctionalized. For fully redundant genes, silencing mutations go to fixation at the rate of μ_c per locus, since the number of newly arising mutations is $2N\mu_c$ per locus and the probability of a fixation of a neutral allele is $1/(2N)$, whereas beneficial mutations to a novel function go to fixation at the rate of $2Nu_F\mu_b$, as there are again $2N$ gene copies per locus, each mutating at rate μ_b and in this case fixing with probability u_F. We rely on the diffusion approximation for the probability of fixation of a newly arisen beneficial mutation with additive effects,

$$u_F = \frac{1 - e^{-2s}}{1 - e^{-4Ns}} \tag{15}$$

(KIMURA 1962). Letting $\beta = 2Nu_F\mu_b/(\mu_c + 2Nu_F\mu_b)$ denote the relative probability of neofunctionalization, the conditional probabilities of the four possible fates of linked duplicates destined to fixation are

$$P'_{non,m} = P'_{non,o} = (1 - \beta)/4N, \tag{16a}$$

$$P'_{neo,m} = P'_{neo,o} = \beta/4N. \tag{16b}$$

Were these the only paths to the preservation of a new duplicate, one would expect the upper limit for Θ and Γ to equal 1, because $\beta \leq 1.0$. However, we must also consider the possibility of the appearance of a neofunctionalizing mutation in a two-copy allele that is otherwise destined to be lost by random genetic drift, as this can alter the course of events.

To quantify the probability of such a rescue effect, we need to know the number of alleles that are available targets for neofunctionalizing mutations. The expected number of two-copy alleles in the population in generation t, conditional on not having yet been lost or having been rescued, can be shown to be

$$n_m(t) = \frac{e^{-t/(2N)}}{1 - u_L(t)}, \tag{17}$$

where $u_L(t)$ is the probability that the locus has been lost by drift by generation t. Because we are focusing

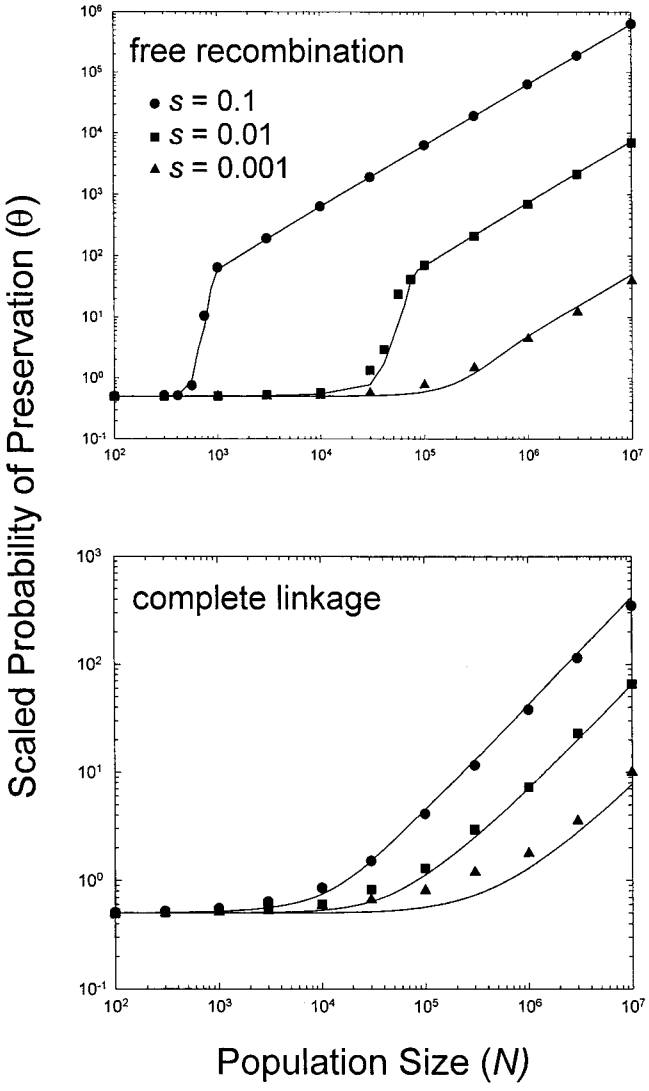

FIGURE 6.—The scaled probability of preservation of a duplicate gene for the situation in which mutations either completely silence a gene or endow it with a new function at the expense of the old function. Solid lines are the predictions derived from the theory outlined in the text.

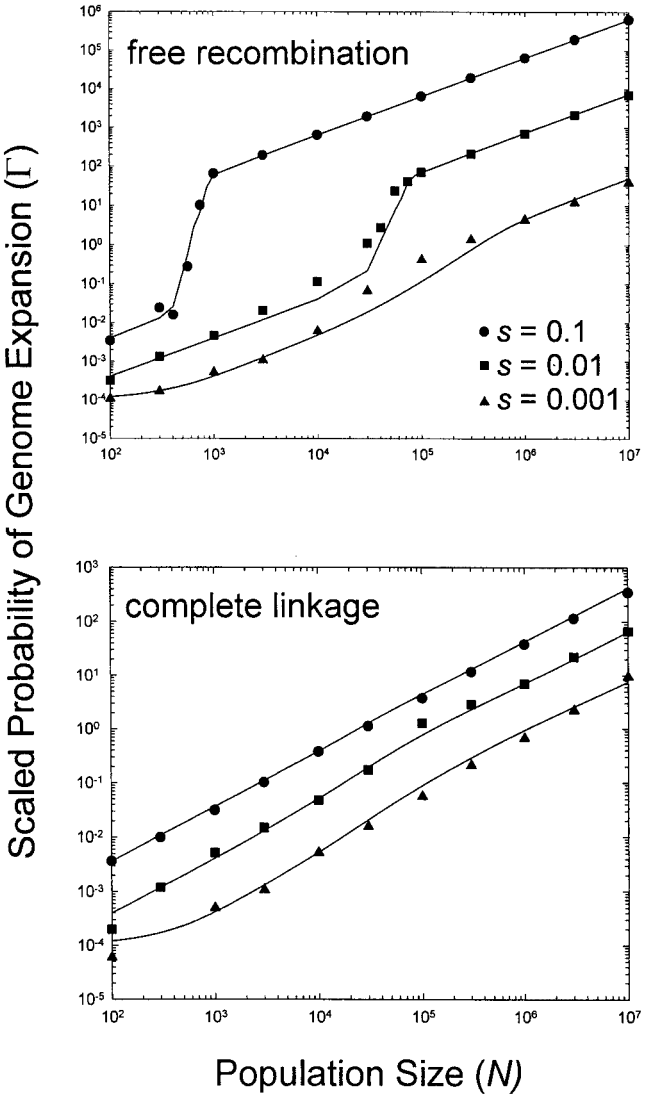

FIGURE 7.—The scaled probability of genome expansion per newly arisen gene duplicate for the situation in which mutations either completely silence a gene or endow it with a new function at the expense of the old function. Solid lines are the predictions derived from the theory outlined in the text.

on a large-population phenomenon, $u_L(t)$ can be approximated with FISHER's (1922) recursion for a mutant allele initially present in a single copy,

$$u_L(t) = e^{u_L(t-1)-1}, \tag{18}$$

starting with $u_L(0) = 0$. The probability that a two-copy allele otherwise destined to be lost acquires a neofunctionalizing mutation in generation t that will carry it to fixation is then

$$r(t) = 1 - e^{-2\mu_b u_F n_m(t) e^{-\mu_c t}}, \tag{19}$$

the 2 accounting for the two copies of the ancestral gene per two-copy allele, and the term $e^{-\mu_c t}$ being the probability that a gene within the pair has not acquired a silencing mutation by time t. Letting

$$p_L(t) = 1 - \frac{1 - u_L(t+1)}{1 - \mu_L(t)} \tag{20}$$

be the probability that an effectively neutral allele destined to eventual loss is lost in generation t and $\ell(t)$ be the probability that the fate of two-copy alleles has not been determined by generation t, then the partition of the contributions to alternative fates for the λ cases in which a two-copy allele is initially destined to become lost is

$$P''_{neo,m}(t) = P''_{neo,o}(t) = 0.5\lambda\ell(t)r(t), \tag{21a}$$

$$P''_{non,m}(t) = \lambda\ell(t)p_L(t)[1 - r(t)], \tag{21b}$$

with

$$\ell(t+1) = \ell(t) - P''_{\text{neo,m}}(t) - P''_{\text{neo,o}}(t) - P''_{\text{non,m}}(t). \tag{22}$$

The final probabilities of the four alternative fates are given by

$$P_{\text{non,m}} = P'_{\text{non,m}} + \sum_{t=0}^{\infty} P''_{\text{non,m}}(t), \tag{23a}$$

$$P_{\text{non,o}} = P'_{\text{non,o}}, \tag{23b}$$

$$P_{\text{neo,m}} = P'_{\text{neo,m}} + \sum_{t=0}^{\infty} P''_{\text{neo,m}}(t), \tag{23c}$$

$$P_{\text{neo,o}} = P'_{\text{neo,o}} + \sum_{t=0}^{\infty} P''_{\text{neo,o}}(t). \tag{23d}$$

(For the reader's convenience, we summarized the definitions of all terms associated with the neofunctionalization model in Table 2.)

For the most part, these expressions are in good agreement with the simulated data (Figures 6 and 7). At small population sizes, there is a negligible likelihood of a beneficial mutation resurrecting a two-copy locus destined to be lost by drift, so from Equations 16a and 16b alone, $\Theta \simeq (1+\beta)/2$ and $\Gamma \simeq \beta$. At the very smallest population sizes ($N < 10^3$), β asymptotically approaches $\mu_b/(\mu_c + \mu_b)$, which for $\mu_b \ll \mu_c$ results in $\Theta \to 0.5 + (\mu_b/\mu_c)$ and $\Gamma \to \mu_b/\mu_c$. On the other hand, in the limit as $N \to \infty$, the chance of the original locus becoming silenced is negligible, which results in $\Gamma \simeq \Theta$ scaling nearly linearly with population size.

Unlinked loci: The probability of neofunctionalization can be greatly enhanced in the case of freely recombining loci because a new duplicate locus that is founded by a neofunctionalized allele is free to move toward fixation and because the fates of subsequent mutations at one locus are less influenced by those at the other. Given that the equilibrium allele frequencies at the original locus are related to N and s in a threshold manner (Equations 13 and 14 and Figure 5), two alternative sets of analytical approximations appear to be necessary.

We first consider the situation in which neofunctionalized alleles are likely to be segregating at nonnegligible frequencies, $\mu_c < [s/(1+s)]^2$, which for the parameters that we examined holds for $s = 0.1$ and 0.01. To have any chance of establishing itself permanently, a newly arisen duplicate locus must be founded by either a neofunctionalized (n) or wild-type (f) allele, the probabilities of which are

$$p_n = \hat{p}_n/(1-\hat{p}_0), \tag{24a}$$

$$p_f = 1 - p_n, \tag{24b}$$

where $\hat{p}_n$ and $\hat{p}_0$ are defined by the values in Figure 5. If the founder allele is of the neofunctional type, the probability of fixation is given by Equation 15 with selection coefficient

$$s_n = s(1-\hat{p}_n-\hat{p}_0)/(1+\hat{p}_n+\hat{p}_0), \tag{25a}$$

and, conditional upon such fixation, the original locus must maintain the original function. If the founder allele is wild type, the probability of fixation is a function of the relative fitnesses of the *ff*, *f*0, and 00 genotypes at the new locus induced by the presence of 00, *n*0, and *nn* genotypes at the original locus, where 0 denotes a nonfunctional allele. The latter genotypes have zero fitness if the genotype at the new locus is 00 but respective fitnesses of 1, $1+s$, and $1+2s$ if the genotype at the new locus is *ff* or *f*0. Scaling the fitness of the 00 genotype at the new locus to be equal to one, the initial expected selective advantage of both the *ff* and *f*0 genotypes is equal to

$$s_f = (\hat{p}_0 + \hat{p}_n)(2\hat{p}_n s + \hat{p}_0 + \hat{p}_n), \tag{25b}$$

which for large N and $\mu_c < [s/(1+s)]^2$ simplifies to $s_f \simeq s^2/(1+2s)$. Wright (1969, p. 382) provides a series approximation for the probability of fixation of a dominant beneficial mutation, but for the values of s that we employed this yields results that are very close to the values obtained with Equation 15 after substituting s_f for s. Conditional upon fixation of the *f* allele at the new locus, the neofunctional alleles residing at the original locus may proceed to fixation with probability

$$u_F(s) = \frac{1-e^{-4Ns\hat{p}_n}}{1-e^{-4Ns}}, \tag{26}$$

and in the event that this does not occur, one of the two loci is expected to become neofunctionalized via new mutations with probability β. Summing up the various paths, the probabilities of the four alternative fates of the gene pair are then given by

$$P_{\text{neo,m}} = [p_f u_F(s_f)(1-u_F(s))\beta/2] + [p_n u_F(s_n)], \tag{27a}$$

$$P_{\text{neo,o}} = p_f u_F(s_f)[u_F(s) + ((1-u_F(s))\beta/2)], \tag{27b}$$

$$P_{\text{non,o}} = p_f u_F(s_f)(1-u_F(s))(1-\beta)/2, \tag{27c}$$

$$P_{\text{non,m}} = 1 - P_{\text{neo,m}} - P_{\text{neo,o}} - P_{\text{non,m}}, \tag{27d}$$

where $u_F(s_f)$ and $u_F(s_n)$ are obtained from Equation 15 after substituting for s. In the limit for large N, $\beta \to 1$, $p_n \to s/(1+2s)$, $u_F(s_f) \to 2s^2/(1+2s)$, and $u_F(s_n) \to 2s(1+s)^2/(1+2s)^2$, leading to $\Theta = \Gamma \simeq 4Ns^2(2+3s)(1+s)/(1+2s)^2$ and $\Delta \simeq \Theta/(2+3s)$. Provided $s < 0.1$, these large-N/large-s approximations reduce further to $\Theta = \Gamma \simeq 8Ns^2$ and $\Delta \simeq 4Ns^2$, showing that all three statistics increase linearly with N (implying that the probabilities of these fates are independent of N) and with the square of s.

We now turn to the situation in which $\mu_c > [s/(1+s)]^2$, which for the parameters that we examined holds for $s = 0.001$, and in which case there is a negligible chance of the new locus being initially founded with a neofunctional allele. We again take a cohort approach, similar to that used in the case of linked loci, noting that the founder allele at the new locus is initially destined to fix with probability $1/(2N)$ and otherwise to be lost with probability λ. In the former case, one of the loci is expected to eventually become neofunctionalized with

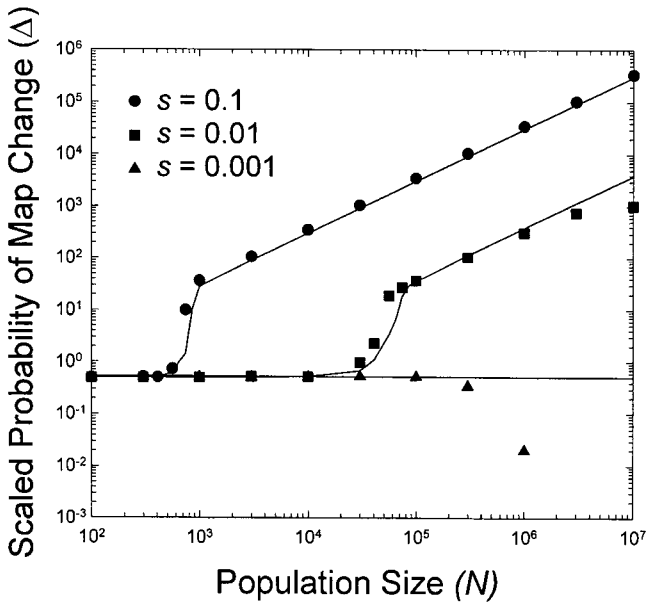

FIGURE 8.—The scaled probability of a map change (for unlinked duplicates) per newly arisen gene duplicate for the situation in which mutations either completely silence a gene or endow it with a new function at the expense of the old function. Solid lines are the predictions derived from the theory outlined in the text.

probability β or to become nonfunctionalized with probability 1 − β. In the latter case, we must account for the possibility that the new locus, otherwise destined to be lost, will be rescued with a neofunctionalizing mutation. The probability of rescue in generation t is given by

$$r(t) = 1 - e^{-\mu_b u_F n_m(t)}, \tag{28}$$

with u_F defined by Equation 15 and $n_m(t)$ by Equation 17, and the generation-specific contributions to alternative fates for the cases in which the founder allele is initially destined to loss are

$$P''_{neo,m}(t) = \lambda \ell(t) r(t), \tag{29a}$$

$$P''_{non,m}(t) = \lambda \ell(t) p_L(t)[1 - r(t)], \tag{29b}$$

where $p_L(t)$ is defined by Equation 20, and

$$\ell(t+1) = \ell(t) - P''_{neo,m}(t) - P''_{non,m}(t). \tag{30}$$

We then have

$$P_{neo,m} = (\beta/4N) + \sum_{t=0}^{\infty} P''_{neo,m}(t), \tag{31a}$$

$$P_{neo,o} = \beta/4N, \tag{31b}$$

$$P_{non,o} = (1 - \beta)/4N, \tag{31c}$$

$$P_{non,m} = 1 - P_{neo,m} - P_{neo,o} - P_{non,m}. \tag{31d}$$

As can be seen in Figures 6–8, the theory for freely recombining duplicates is in fairly close agreement with the values of Θ, Γ, and Δ observed over the full range of N and s, the main exception being the overestimation of Δ at large N when selection is weak. When N is small, $\Theta = \Delta \simeq 0.5$ independent of s. This is again a consequence of the fact that the probability of fixation of a newly arisen locus is equal to $1/(2N)$ and that one of the loci will then almost always become silenced, because of the negligible probability of neofunctionalization. On the other hand, once N exceeds a threshold value (depending on s and μ_c), Θ scales linearly with N and approximately linearly with s^2 in agreement with the asymptotic expressions given above. A similar scaling with N and s^2 is seen for Γ at large N. The abrupt change in the behavior of Θ, Γ, and Δ at intermediate N and strong selection ($s = 0.01$ and 0.1) corresponds precisely with the abrupt change in frequency of neofunctional alleles at the original locus (Figure 5).

DISCUSSION

These results demonstrate that the evolutionary trajectories of duplicate genes are not just functions of intrinsic organismal properties such as gene structure, regulatory-region complexity, distribution of mutational effects, etc., but are also highly dependent on the effective size of a population. This view suggests that the mechanisms influencing the fates of duplicate genes may vary dramatically among species (and even within the history of individual species lineages) depending on the population size prevailing during the initial appearance of a duplicate gene. Population size influences the evolution of duplicate genes in two ways. First, larger populations are more likely to harbor segregating subfunctional or neofunctional alleles at the ancestral locus prior to duplication, raising the possibility that the newly arisen locus may be founded by an allele other than the wild type and also the possibility that the ancestral locus can rapidly become neofunctionalized (without the reliance on new beneficial mutations) if the new locus becomes established with wild-type alleles. Second, because the time to fixation (and loss) increases with increasing population size, the potential fates of duplicate genes can be altered during the long period in which they drift through large populations and acquire secondary mutations. For example, subfunctional alleles at a new locus may become completely silenced by degenerative mutations prior to fixation, whereas functional alleles that are otherwise destined to be lost by drift can on occasion be rescued by a beneficial mutation. Thus, attempts to understand the evolution of the duplicate genes (and by extrapolation, other aspects of genome expansion/contraction) are not likely to be successful unless they are considered in the context of the genetic properties of finite populations.

Preservation of the new copy: Two rather different models, one incorporating only degenerative mutations and the other also including beneficial mutations, suggest that the probability of preservation of a newly arisen duplicate gene is generally no less than half of its initial frequency (*i.e.*, $\Theta > 0.5$) regardless of the degree of

linkage (Figures 3 and 6). Thus, unless there is active selection against a duplicate gene, its probability of permanent establishment is at least one-half the expected fixation probability of a neutral allele, *i.e.*, $\geq 1/(4N)$. Moreover, in the absence of an appreciable likelihood of fixation of beneficial mutations (either because the rate of mutation to such alleles is too low, the beneficial effects are too small, or the population size is insufficiently large), the probability of preservation is unlikely to exceed $1/(2N)$. On the other hand, in sufficiently large populations, neofunctionalization can lead to probabilities of preservation (per duplication event) that are independent of N and orders of magnitude greater than possible under a scenario dominated by degenerative mutations. Provided the null mutation rate is sufficiently small relative to the strength of selection ($\mu_c < [s/(1 + s)]^2$) and the effective population size is sufficiently large ($Ns^2 > 4$; Figure 5), most cases of neofunctionalization following gene duplication are expected to be driven by neofunctional alleles preexisting at the ancestral locus rather than by mutations arising subsequent to the duplication event. If the new locus is founded by a wild-type allele that reaches sufficiently high frequency, natural selection will promote the neofunctional alleles segregating at the original locus. Alternatively, the new locus may be founded by a neofunctional allele that goes to fixation, in which case the original gene function will be maintained at the ancestral locus.

Although our results suggest that subfunctionalization will be a more common mechanism of duplicate-gene preservation in small populations, with neofunctionalization becoming progressively more common as N increases, the exact population size at which neofunctionalization begins to exceed subfunctionalization as a preservational mechanism will depend on the relative rates of origin of the two types of preservational mutations (μ_r and μ_b) and on the selective advantage of neofunctional alleles. For the case of neofunctionalization, it is noteworthy that $\Theta(= \Delta)$ scales not with s, as would normally be expected for an unconditionally advantageous allele at a single locus, but with the square of s. This scaling can be understood most easily by considering the case of unlinked duplicates at large N. If the founding allele at the new locus is wild type, its main initial advantage (relative to "absentee" alleles at the new locus) arises in backgrounds where the genotype at the ancestral locus is of type nn, $n0$, or 00, and from Equations 13a and 13b it can be seen that the most abundant of these genotypes, nn, has an expected frequency $\simeq [s/(1 + 2s)]^2$. On the other hand, if the founding allele is of the neofunctional type, it will go to fixation with probability $\simeq 2s_f$, and from Equation 25b it can be seen that $s_f \simeq s^2/(1 + 2s)$. Thus, regardless of the nature of the founder allele, its probability of fixation scales approximately with s^2 at large N. If subfunctionalizing mutations greatly outnumber neofunctionalizing mutations and s is typically small, neither of which seems unlikely, then the majority of successful gene duplicates may owe their preservation to subfunctionalization. Not included in our analyses is the possibility that many duplicates may be subfunctionalized at birth via the duplication process itself, due, for example, to the failure of the duplicated region to cover the full ancestral gene sequence (Averof *et al.* 1996). Such conditions would further increase the relative incidence of subfunctionalization as a preservational process.

In large populations, the degree of linkage between duplicate genes can substantially influence the probability of preservation of a new gene copy (Figures 3 and 6). When degenerative mutations dominate the process, a linked pair of functional duplicates has a weak transient selective advantage over a single-copy allele, because the former requires at least two mutations to be silenced. This results in an increase in the probability of preservation from $1/(4N)$ at small N to an asymptotic level of $1/(2N)$ at large N. Thus, in the absence of beneficial mutations, a linked pair of duplicates fixes at the neutral rate at large N despite the fact that the underlying process is non-neutral. This behavior contrasts with that of an unlinked duplicate, which, in the absence of beneficial mutations, is prevented from becoming permanently established in very large populations by saturation with silencing mutations by the time the lineage fixes in the population. In contrast, when neofunctionalizing mutations become a prominent influence, linkage reduces the probability of preservation of gene duplicates. Free recombination facilitates the neofunctionalization process because a pair of completely linked neofunctional genes (or a pair containing one neofunctional and one nonfunctional copy) is prevented from going to fixation by the lack of the critical ancestral gene function.

These results suggest the hypothesis that duplicate genes that are preserved by neofunctionalization will tend to be unlinked, whereas those preserved by subfunctionalization (or silencing of the ancestral gene) will tend to be more closely linked (at least during the period of preservation). It should be noted, however, that although duplicate genes often arise in tandem association with the parental locus, they are frequently recruited to new locations at an early stage of their history (Lynch and Conery 2000). The influence of linkage on the fate of a duplicate pair will clearly depend on the timing of such translocation events.

Evolution of genome size: Although the preservation of duplicate genes often leads to an expansion in genome size, this is not necessarily the case because the preservation of a new gene copy may be balanced by the loss of the ancestral copy. For example, in sufficiently small populations, where the likelihood of neofunctionalization is reduced to negligible levels, a new duplicate may still become preserved if it drifts to fixation and the original locus becomes nonfunctionalized,

but in this case there is no net change in genome size. Any pressure toward genome-size expansion is expected to come from subfunctionalization until a critical population size has been reached and neofunctionalization becomes more dominant, the exact threshold population size again depending on μ_r, μ_b, s^2, and the degree of linkage between ancestral and descendant loci.

Like nucleotide substitutions, insertions, and deletions, gene duplication appears to be a common attribute of all genomes. For example, analysis of the complete genomic sequences of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* suggests that new duplications may typically become established in populations at rates on the order of 10^{-3}–10^{-2} per gene per million years (Lynch and Conery 2000). These are probably conservative estimates as they do not include duplicates arising in large multigene families. Thus, on a per-locus basis, the rate of gene duplication appears to be of the same order of magnitude as nucleotide substitution. With the typical eukaryotic genome containing on the order of 10^4–10^5 genes, it appears (very roughly) that 10–1000 new gene duplicates may become established at high frequency per genome on a timescale of 1 million years, with their subsequent long-term fates then depending on the mutational mechanisms outlined above.

Because subfunctionalizing and neofunctionalizing mechanisms will generally ensure an innate tendency toward a net accumulation of new genes, stability in genome size requires selection against too many gene duplicates and/or molecular mechanisms that stochastically delete additional copies. In the absence of such opposing forces, one might expect the expansion of genome size to be a self-accelerating process, as the accumulation of more genes provides more substrate for future duplications. However, the opportunities for preservation by subfunctionalization are expected to be reduced as members of a gene family partition up the tasks of the ancestral gene, and, under the neofunctionalization model, the likelihood of establishing a new beneficial function may decline with an increase in organismal complexity; *i.e.*, both μ_r and μ_b may decline with increasing genome size. These design limitations alone may constrain the indefinite expansion of genome size, but mutational mechanisms almost certainly play an additional role. For example, nonessential DNA appears to have a half-life of ~14 million years in Drosophila and ~880 million years in mammals (Petrov and Hartl 1998), and comparative analyses have consistently indicated a tendency for the rate of deletion of DNA to exceed that of insertion (de Jong and Ryden 1981; Gu and Li 1995; Lynch 1996). Although numerous mechanisms may counteract the innate tendency toward genome expansion generated by gene duplication, it is unlikely that these opposing forces will ever be perfectly balanced. Rather, the genome sizes of individual species may typically undergo stochastic phases of expansion and contraction depending on the prevailing aspects of population size and selection regime.

The mechanisms that we have suggested for the expansion of genome size via duplicate genes need not be all inclusive. For example, it has been suggested that genomic redundancies may be selectively maintained to mask the consequences of null homozygotes or errors in transcription and translation (Clark 1994; Nowak *et al.* 1997; Krakauer and Nowak 1999; Wagner 1999). Although these types of buffering models are diverse in terms of assumptions, they are most closely related to our analyses in which both the neofunctionalizing and subfunctionalizing mutation rates are equal to zero. In this case, any selective advantage of a newly arisen duplicate gene is entirely derived from masking the effects of the null homozygote at the original locus, whose frequency approaches μ_e when N is large. However, under this simple model, we find that one member of a duplicate pair is always eventually lost by random genetic drift, even at very large population sizes. This seems to result from the fact that the selective advantage of a duplicate gene under this model (the equilibrium frequency of null homozygotes at the original locus) is less than or equal to the silencing mutation rate. Thus, the permanent preservation of duplicate genes by a buffering mechanism appears to require both very large N and a frequency of null phenotypes elevated above the genetic expectation by errors in intracellular processing.

Alterations of the genetic map: Gene duplication may be of as much relevance to the origin of new species as it is to the origin of evolutionary novelty within species (Werth and Windham 1991; Lynch and Force 2000b). As noted above, for unlinked duplicates, the probability of a map change for gene function (or subfunction) is generally no less than $1/(4N)$ per gene-duplication event, and, in large populations, neofunctionalization can magnify this probability by several orders of magnitude (Figure 8). One consequence of a map change is that double-null homozygotes segregate out with frequency 1/16 in the progeny of F_1 hybrids, and additional problems can arise when nulls are not completely recessive, when genomic imprinting occurs, when one member of a pair resides on a sex chromosome, and when the haploid phase of the genome is transcriptionally active (Lynch and Force 2000b). If we accept that the incremental rate of origin of new gene duplicates in a population is somewhere in the range of 10–1000 per million years, then on the order of a dozen to a few hundred potential map changes can be expected to arise in two lineages separated for this time period, the actual number depending on the fraction of newly arisen duplicates that are either unlinked at the time of origin or soon become unlinked by subsequent chromosomal events. Consistent with this view, recent work in comparative genomics indicates that even when gross chromosomal gene order remains roughly stable between species, microchromosomal rearrangements (in-

cluding reassignments of individual genes to new chromosomal locations associated with duplication events) are quite common among closely related species (Kent and Zahler 2000; Bancroft 2001; Dehal *et al.* 2001). An indirect consequence of gene duplication for the origin of map changes that we have not considered here is homologous recombination between duplicated loci, which can produce reciprocal translocations (Ryu *et al.* 1998). Thus, there is little question that duplication-induced map changes are a common genomic property, and the key remaining questions concern the degree to which these, as opposed to other mechanisms (*e.g.*, changes within genes), dominate the process of reproductive isolation.

Although the origin of new species is often viewed as a small-population phenomenon, our results demonstrate how reproductive incompatibilities can passively arise between very large isolated populations. Because Δ increases with increasing s, reproductive incompatibilities induced by gene duplication may be accompanied by the origin of new adaptive functions. However, such an association is a simple consequence of the change in map position that frequently accompanies the origin of genes with new functions, not a result of the adaptive changes themselves. It is noteworthy as well that map displacements of divergently resolved gene duplicates will cause the superficial appearance of negative epistatic interactions in the genetic analysis of hybrid progeny, even in the absence of any interactions between the gene products contributing to novelties in the sister taxa. In this sense, studies of reproductive isolating barriers that do not identify mechanisms to the gene level may be quite deceiving. As emphasized elsewhere (Lynch and Conery 2000; Lynch and Force 2000b), the gene-duplication model for the origin of genomic incompatibility is consistent with both the leading genetic models for the origin of reproductive isolation (the epistasis model of Dobzhansky 1936 and Muller 1940 and the chromosomal rearrangement model of White 1978 and others), while invoking fewer assumptions than either. Our results also raise the hypothesis that divergent resolution of gene duplicates following a genome-wide or chromosomal duplication event may promote the origin of many nested reproductive isolation events in descendant lineages, with adaptive radiations following as a secondary consequence.

Future work: The theory developed in this article is meant to provide some heuristic guidance to our understanding of the mechanisms that lead to the preservation *vs.* silencing of duplicate genes, and by necessity a number of assumptions have been made. For example, we have focused on nonfunctionalizing and subfunctionalizing mutations of large effects (as have most previous theoretical investigations in this area). However, our earlier work (Lynch and Force 2000a) suggests that additional subfunctions or mutations of minor effect will simply increase the probability of duplicate-gene preservation to a level of $1/(2N)$ when N is small, and limited simulations at large N suggest the same. In addition, we have ignored issues of dosage, which may play a significant role with genes whose products must be in the correct stoichiometric ratios with those of their interacting partners (Force *et al.* 1999; Shimeld 1999). Except in the case of duplications involving entire genomes, such effects would impose negative selection against newly arisen duplicates. Finally, in our models involving neofunctionalization, we assumed that a mutant allele with a gain of function fails to perform its original function. One can envision a range of additional models involving neofunctionalizing mutations, the opposite extreme being the case in which neofunctionalization has no impact on the ancestral gene function. In the latter case, however, one would imagine that such unconditionally beneficial mutations would have ample opportunity to arise at the original locus (where virtually all of the mutational substrate resides). We have, therefore, chosen to focus on mutant alleles that depend on the duplication process to provide the freedom necessary to move toward fixation.

These issues aside, it is clear that a definitive understanding of the forces that dictate the fates of duplicate genes will require careful work at the empirical level. Such studies will need to focus on pairs of loci that are relatively early in their phase of establishment because the mutations responsible for the initial preservation of such genes may be substantially different from those that are incurred during subsequent evolutionary history. Unfortunately, almost all existing studies of the biology of duplicate genes have focused on pairs that have been established for so long that it is impossible to identify the mutations that were responsible for their initial preservation. A fundamental issue that remains to be resolved is the extent to which newborn duplicate genes share the full spectrum of functions and efficiencies of their ancestral copy. Although the preceding theory assumes complete functional redundancy, there is no reason why duplicated gene regions should always provide full coverage of upstream and downstream regulatory regions. Less than full coverage will almost certainly modify the potential evolutionary trajectories of newly arisen duplicates, most likely increasing the probability of subfunctionalization, but perhaps providing new opportunities for neofunctionalization as well.

For newly arising pairs of loci, it will be most instructive to know the incidence of active *vs.* partially or completely silenced alleles at both the original and the descendant locus, as well as the incidence of absenteeism at the new locus. Silent nucleotide sites should help reveal the relative ages of pairs of duplicates (assuming problems with gene conversion are minor), and careful studies of the rate of substitution at silent *vs.* replacement sites may clarify whether different gene regions are evolving in a neutral fashion, are being maintained by purifying selection, or are in the process of being

transformed to new beneficial functions. A series of such studies with loci of different ages could then provide at least a qualitative glimpse into the factors that determine the fates of a typical pair of gene duplicates and the timescale over which these are established. Dermitzakis and Clark (2001) recently proposed a phylogenetic method for testing whether the two members of a duplicate pair evolve in a similar manner over all of their protein-coding domains, showing how significant differences between paralogues can be used to identify the potential footprints of subfunctionalization. In principle, their approach can be extended to regulatory-region DNA, and the *conceptual* power of the method may be greatly enhanced by the inclusion of an outgroup species containing a single-copy gene. The primary caveat here is that the *statistical* power of phylogenetic comparison is relatively weak unless the phylogeny is deep enough to contain substantial numbers of nucleotide substitutions, so the method of Dermitzakis and Clark (2001) may be of limited utility in studies of the earliest stages of gene duplication.

As whole genome sequences have emerged for a diversity of species, the identification of newly arisen pairs of duplicates has become quite feasible (Lynch and Conery 2000), and it is also clear that duplications still in the process of spreading through a population can be located. An example of such a study is the recent investigation of the α-amylase gene cluster in the *D. melanogaster* complex (Robin *et al.* 2000). Phylogenetic analysis suggests that one member of this cluster is fixed as a pseudogene in *D. melanogaster* (a victim of nonfunctionalization), whereas its orthologues remain active and apparently under purifying selection in the closely related species *D. simulans* and *D. yakuba*. It seems very likely that this locus contained at least some active alleles in the common ancestor of these three species but had not yet arrived at a stable state. Under this interpretation, the alternative states that have arisen in the descendant lineages may simply be stochastic outcomes of the mutation process and allelic sorting by random genetic drift (as in our simulations). It remains to be seen whether the new locus has been preserved by subfunctionalization or neofunctionalization in the *D. simulans* and *D. yakuba* lineages or whether it is still in a phase of resolution (in fact, only a single allele was examined in these two taxa). Several other examples of presence/absence polymorphisms of duplicate genes are known in Drosophila, including methallothionein in *D. melanogaster* (Lange *et al.* 1990), urate oxidase in *D. virilis* (Lootens *et al.* 1993), and alcohol dehydrogenase in *D. funebris* (Amador and Juan 1999).

Finally, we note that our results have not entirely clarified the conditions influencing the likelihood of successful gene-duplication events in extremely large populations. On the one hand, neofunctionalizing mutations are most likely to become permanently established in large populations (Figure 6). On the other hand, if the preservational process is largely driven by degenerative mutations or if the selective advantage of a neofunctional allele is sufficiently small, when $N\mu_c \gg 10$ and the loci are unlinked, it is almost certain that all of the descendants of a newly arisen duplicate will be silenced by the time its lineage is fixed (Figure 3). It is, therefore, at least plausible that the increased genome size of vertebrates (mouse and human) relative to invertebrates (flies and worms), of *C. elegans* relative to *D. melanogaster*, and perhaps even eukaryotes relative to prokaryotes is largely an indirect consequence of differences in effective population size. This view does not deny the possibility that increases in genome size may ultimately facilitate the evolution of organismal complexity by natural selection, but it does raise the possibility that nonselective forces, most notably random genetic drift and degenerative mutation, set the initial stage upon which such evolutionary changes can subsequently take place.

We thank Kevin Higgins for help with computational procedures. This research was supported by National Institutes of Health (NIH) grant RO1-GM36827 to M.L.; by graduate fellowships to A.F. funded by a National Science Foundation (NSF) training grant in genetic mechanisms of evolution and in evolution and by an NIH training grant in developmental biology; and by a postdoctoral fellowship to A.F. funded by an NSF IGERT training grant in evolution, development, and genomics.

LITERATURE CITED

Amador, A., and E. Juan, 1999 Nonfixed duplication containing the *Adh* gene and a truncated form of the *Adhr* gene in the *Drosophila funebris* species group: different modes of evolution of *Adh* relative to *Adhr* in *Drosophila*. Mol. Biol. Evol. **16:** 1439–1456.

Averof, M., R. Dawes and D. Ferrier, 1996 Diversification of arthropod *Hox* genes as a paradigm for the evolution of gene functions. Semin. Cell Dev. Biol. **7:** 539–551.

Bailey, G. S., R. T. M. Poulter and P. A. Stockwell, 1978 Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. Proc. Natl. Acad. Sci. USA **75:** 5575–5579.

Bancroft, I., 2001 Duplicate and diverge: the evolution of plant genome microstructure. Trends Genet. **17:** 89–93.

Christiansen, F. B., and O. Frydenberg, 1977 Selection-mutation balance for two nonallelic recessives producing an inferior double homozygote. Am. J. Hum. Genet. **29:** 195–207.

Clark, A. G., 1994 Invasion and maintenance of a gene duplication. Proc. Natl. Acad. Sci. USA **91:** 2950–2954.

Dehal, P., P. Predki, A. S. Olsen, A. Kobayashi, P. Folta *et al.*, 2001 Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. Science **293:** 104–111.

de Jong, W. W., and L. Ryden, 1981 Causes of more frequent deletions than insertions in mutations and protein evolution. Nature **290:** 157–159.

Dermitzakis, E. T., and A. G. Clark, 2001 Differential selection after duplication in mammalian developmental genes. Mol. Biol. Evol. **18:** 557–562.

Dobzhansky, T., 1936 Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21:** 113–135.

Fisher, R. A., 1922 On the dominance ratio. Proc. R. Soc. Edinb. **52:** 399–433.

Fisher, R. A., 1935 The sheltering of lethals. Am. Nat. **69:** 446–455.

Force, A., M. Lynch, B. Pickett, A. Amores, Y.-L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Gu, X., and W.-H. Li, 1995 The size distribution of insertions and

deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequencing alignment. J. Mol. Evol. **40:** 464–473.

Haldane, J. B. S., 1933 The part played by recurrent mutation in evolution. Am. Nat. **67:** 5–9.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. Ser. B **256:** 119–124.

Kent, W. J., and A. M. Zahler, 2000 Conservation, regulation, synteny, and introns in a large-scale *C. briggsae-C. elegans* genomic alignment. Genome Res. **10:** 1115–1125.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47:** 713–719.

Kimura, M., and T. Ohta, 1969 The average number of generations until fixation of a mutant gene in a finite population. Genetics **61:** 763–771.

Krakauer, D. C., and M. A. Nowak, 1999 Evolutionary preservation of redundant duplicated genes. Semin. Cell Dev. Biol. **10:** 555–559.

Lange, B. W., C. H. Langley and W. Stephan, 1990 Molecular evolution of Drosophila metallothionein genes. Genetics **126:** 921–932.

Li, W.-H., 1980 Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics **95:** 237–258.

Lootens, S., J. Burnett and T. B. Friedman, 1993 An intraspecific gene duplication polymorphism of the urate oxidase gene of *Drosophila virilis:* a genetic and molecular analysis. Mol. Biol. Evol. **10:** 635–646.

Lynch, M., 1996 Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13:** 209–220.

Lynch, M., and J. C. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1154.

Lynch, M., and A. Force, 2000a The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459–473.

Lynch, M., and A. Force, 2000b The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. **156:** 590–605.

Muller, H. J., 1940 Bearing of the *Drosophila* work on systematics, pp. 185–268 in *The New Systematics*, edited by J. S. Huxley. Clarendon Press, Oxford.

Nei, M., and A. K. Roychoudhury, 1973 Probability of fixation of nonfunctional genes at duplicate loci. Am. Nat. **107:** 362–372.

Nowak, M. A., M. C. Boerlijst, J. Cooke and J. Maynard Smith, 1997 Evolution of genetic redundancy. Nature **388:** 167–170.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Petrov, D. A., and D. L. Hartl, 1998 High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* groups. Mol. Biol. Evol. **15:** 293–302.

Piatigorsky, J., and G. Wistow, 1991 The recruitment of crystallins: new functions precede gene duplication. Science **252:** 1078–1079.

Robin, G. C. de Q., R. J. Russell, D. J. Cutler and J. G. Oakshott, 2000 The evolution of an α-esterase pseudogene inactivated in the *Drosophila melanogaster* lineage. Mol. Biol. Evol. **17:** 563–575.

Ryu, S.-L., Y. Murooka and Y. Kaneko, 1998 Reciprocal translocation at duplicated *RPL2* loci might cause speciation of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. Curr. Genet. **33:** 345–351.

Shimeld, S. M., 1999 Gene function, gene networks and the fate of duplicated genes. Semin. Cell Dev. Biol. **10:** 549–553.

Spofford, J. B., 1969 Heterosis and the evolution of duplications. Am. Nat. **103:** 407–432.

Stoltzfus, A., 2000 On the possibility of constructive neutral evolution. J. Mol. Biol. **49:** 169–181.

Takahata, N., and T. Maruyama, 1979 Polymorphism and loss of duplicate gene expression: a theoretical study with application to tetraploid fish. Proc. Natl. Acad. Sci. USA **76:** 4521–4525.

Wagner, A., 1999 Redundant gene functions and natural selection. J. Evol. Biol. **12:** 1–16.

Wagner, A., 2000 The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. Genetics **154:** 1389–1401.

Walsh, J. B., 1995 How often do duplicated genes evolve new functions? Genetics **110:** 345–364.

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

Werth, C. R., and M. D. Windham, 1991 A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am. Nat. **137:** 515–526.

White, M. J. D., 1978 *Modes of Speciation*. Freeman, San Francisco.

Wright, S., 1969 *Evolution and the Genetics of Populations, Vol. 2, The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Communicating editor: M. A. Asmussen

APPENDIX

Assuming linked duplicates with a single function, we designate the null and functional single-copy alleles as 0 and *f*, respectively, whereas the four possible two-copy alleles are designated as 00, 0*f*, *f*0, and *ff*. Under the double-null-homozygote model, alleles 0 and 00 are equally viable, and we define their joint frequency to be P_0, which implies an absolute fitness for these alleles of $W_0 = 1 - p_0$. All other alleles have absolute fitnesses equal to 1, so that mean population fitness is $\overline{W} = 1 - p_0^2$. The set of recursion equations for allele frequencies under the assumption of an infinite population size is

$$\Delta p_0 = (1/\overline{W})[(W_0 - \overline{W})p_0 + \mu_c(p_f + p_{0f} + p_{f0})],$$
$$\Delta p_f = (1/\overline{W})[(1 - \mu_c - \overline{W})p_f],$$
$$\Delta p_{0f} = (1/\overline{W})[(1 - \mu_c - \overline{W})p_{0f} + \mu_c p_{ff}],$$
$$\Delta p_{f0} = (1/\overline{W})[(1 - \mu_c - \overline{W})p_{f0} + \mu_c p_{ff}],$$
$$\Delta p_{ff} = (1/\overline{W})[(1 - 2\mu_c - \overline{W})p_{ff}].$$

To transform these difference equations into a solvable set of differential equations, we (1) assume p_0 remains at its initial equilibrium value for a one-locus system, $\sqrt{\mu_c}$ (in reality, there is a very slight initial decline in p_0 when a functional two-copy allele appears, as this slightly reduces the input into the 0 class); (2) use $1/\overline{W} \simeq 1 + p_0^2 = 1 + \mu_c$; and (3) ignore terms of order μ_c^2. The frequencies of the four classes of active alleles then change according to

$$dp_f/dt \simeq 0,$$
$$dp_{0f}/dt = dp_{f0}/dt \simeq \mu_c p_{ff},$$
$$dp_{ff}/dt \simeq -\mu_c p_{ff}.$$

Noting that the initial frequencies are $p_f = 1 - (1/2N) - \sqrt{\mu_c}$, $p_{0f} = p_{f0} = 0$, and $p_{ff} = 1/2N$, the solutions of the above equations are

$$p_f(t) \simeq 1 - (1/2N) - \sqrt{\mu_c},$$
$$p_{0f}(t) = p_{f0}(t) \simeq (1/2N)(1 - e^{-\mu_c t}),$$
$$p_{ff}(t) \simeq (1/2N)e^{-\mu_c t},$$

which shows that as $t \to \infty$, the descendants of the founding duplicate rise in frequency from $p_{ff}(0) = 1/2N$ to $p_{0f}(\infty) + p_{f0}(\infty) = 1/N$.