EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION
PUBLISHED BY
THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 60 **May 2006** **No. 5**

Evolution, 60(5), 2006, pp. 881–892

ALLELIC DIVERGENCE PRECEDES AND PROMOTES GENE DUPLICATION

STEPHEN R. PROULX[1,2,3] AND PATRICK C. PHILLIPS[1]
[1]*Center for Ecology and Evolutionary Biology, 5289 University of Oregon, Eugene, Oregon 97403*
[3]*E-mail: proulx@proulxresearch.org*

Abstract.—One of the striking observations from recent whole-genome comparisons is that changes in the number of specialized genes in existing gene families, as opposed to novel taxon-specific gene families, are responsible for the majority of the difference in genome composition between major taxa. Previous models of duplicate gene evolution focused primarily on the role that neutral processes can play in evolutionary divergence after the duplicates are already fixed in the population. By instead including the entire cycle of duplication and divergence, we show that specialized functions are most likely to evolve through strong selection acting on segregating alleles at a single locus, even before the duplicate arises. We show that the fitness relationships that allow divergent alleles to evolve at a single locus largely overlap with the conditions that allow divergence of previously duplicated genes. Thus, a solution to the paradox of the origin of organismal complexity via the expansion of gene families exists in the form of the deterministic spread of novel duplicates via natural selection.

Key words.—Gene duplication, heterozygote advantage, natural selection, multigene family, polymorphism.

Received September 6, 2005. Accepted February 23, 2006.

Recent whole-genome comparative studies have revealed striking changes in the size of gene families associated with major taxonomic transitions (Chervitz et al. 1998; Lander et al. 2001; Holt et al. 2002; Young et al. 2003; Spaethe and Briscoe 2004). Gene families can include thousands of members and are involved in every aspect of organismal function (Hughes 1999). The large-scale evolution of such gene families is enigmatic because existing theories of the evolution of gene duplication primarily rely on either neutral divergence (Ohno 1970) or mutational degradation of existing duplicates (Force et al. 1999; Lynch and Conery 2000, 2003a,b; Lynch and Force 2000; Lynch et al. 2001). Before a duplication can reach fixation, however, it must arise in a single individual and be governed by population genetic processes that depend on the fitness of individuals with one or two copies of the gene. Previous theories have typically assumed that the evolution of gene families follows a two-step process. First, gene duplications can become fixed in a population by drifting to fixation (Lynch and Force 2000) or through weak selection acting on redundancy alone (Clark 1994; Wagner 1999; Otto and Yong 2002). Second, after the duplication has been fixed in the population, the pair of duplicate loci may diverge and take on different functions or one copy may be lost. During this phase, duplication is assumed to generate functional redundancy that frees the duplicate to explore genotypic space. In refinements of this theory, the duplicates may diverge by gaining a novel function (neofunctionalization) or through subfunctionalization, where each of the duplicate genes takes on some of the functions that the original gene performed (Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001). The presence of a rare neofunctional allele before duplication can generate weak positive selection for duplication (Lynch and Force 2000; Walsh 2003). Because these theories rely on random drift or weak selection to produce duplicates, they are insufficient to explain the rapid evolution of large gene families.

While the existing scenarios for gene duplicate evolution are similar in their focus on the role that postduplication events play in the divergence of duplicate genes, they differ in their assumptions about when divergence begins and what types of mutations lead to divergence. In the neofunctionalization scenario, mutations can lead to null alleles or alleles with a new function (Walsh 1995; Walsh 2003). Here, the outcome depends on a race between the process of psuedogene formation and the fixation of a beneficial mutation. In contrast, several scenarios have been proposed where duplicate genes take on alternative subfunctions after duplication. This can occur when a single gene codes for a protein that has multiple biological functions (gene sharing; Piatigorsky and Wistow 1991; Hughes 1994) or is expressed in multiple tissues (as in the duplication-degeneration-complementation, or DDC, theory; Force et al. 1999). Walsh (2003) provided a synthetic review of a number of these subfunctionalization models. His analysis shows that subfunctionalization and

[2] Present address: Department of Ecology, Evolution and Organismal Biology, Iowa State University, 253 Bessey Hall, Ames, Iowa 50011.

neofunctionalization can interfere with one another and that conditions that limit drift favor neofunctionlization, while small population size and a high subfunctionalization mutation rate favor subfunctionalization. Because the DDC model relies primarily on drift to set the stage for adaptive maintenance of duplicate alleles, it is most likely to have an effect in small populations.

In contrast to these postduplication analyses, several authors have considered the role that existing polymorphism at a single locus contributes to the selection on duplicates of that locus (Spofford 1969; Ohno 1970; Hammerstein 1996; Otto and Yong 2002; Walsh 2003). These studies have shown that, when alternative alleles are maintained because of heterozygote advantage, duplications can rapidly spread through the population. After a random duplication event, one of the duplicate genes will become fixed for a single allele, while the other locus will become fixed for the alternative allele. This allows individuals to act as genetic heterozygotes yet breed true. When fitness is assumed to depend on having at least one copy of each allele, but dosage effects are not present, then the gene duplicate will always be under positive selection to spread (Otto and Yong 2002). Tight linkage and strong selection against homozygotes increases the rate of spread of the gene duplicate. This effect can be quite strong when the two homozygotes have similar fitness effects, but can even create selection for the spread of a duplication when one homozygote is lethal and the other differs from the heterozygote by only a small amount (Walsh 2003).

Genetic redundancy has been considered to be the feature that allows duplicate genes to diverge (Ohno 1970; Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001), but redundancy precedes duplication. In diploid species, redundancy is not unique to duplicated genes because dominance creates the same kind of redundancy at single-copy genes (Phillips and Johnson 1998). It is therefore the mode of inheritance that most distinguishes duplicate loci from single-copy genes because segregation prevents the fixed inheritance of alternative allelic variants at a single locus (Spofford 1969; Hammerstein 1996; Otto and Yong 2002; Proulx and Phillips 2005). Put another way, heterozygotes at a single locus are broken up by segregation, whereas for duplicate loci an individual can carry copies of alternative alleles at separate loci. In the Ohno (Ohno 1970) framework, it is actually this difference in heredity that allows duplicate loci to mutate and drift to high frequency, not a change in the nature of redundancy. Therefore, to understand how gene families arise, we need to examine the conditions that lead to the initial divergence of alleles descended from a common ancestor, whether at one locus or two.

Because the process of the evolution of gene duplicates has typically been separated into a duplication phase and a divergence phase and focused on two- or few-allele models, the adaptive dynamics of allelic change have been largely overlooked. Here, we use tools from adaptive dynamics and evolutionary theory to create a model of the evolution of gene duplicates that considers evolutionary dynamics before, during, and after duplication (Abrams et al. 1993; Dieckmann and Law 1996; Metz et al. 1996; Geritz et al. 1998; Kisdi and Geritz 1999; Geritz and Kisdi 2000; Van Dooren 2000). In our model, the same selective forces that can act to cause duplicate genes to diverge are present before duplication, when the gene exists as a single diploid locus. By comparing the conditions that cause divergence to occur at a single locus to the conditions that cause duplicate genes to diverge, we can predict when divergence will precede duplication. We first develop a general model of this process that shows that the evolution of novel genetic function is likely to both precede and promote gene duplication. The central result is that most genotypic fitness landscapes that allow divergence of genes at a recently duplicated locus also allow divergence of alleles at a single locus. Under most adaptive landscapes that allow divergence of duplicate gene copies, populations with only a single gene copy will travel over the evolutionary landscape to points where heterozygote advantage evolves and gene duplication is favored. We then apply this approach to simulations of a specific model of metabolic or regulatory flux, which confirms the predictions of the general model and illustrates the temporal dynamics of gene duplication. These simulations show that the time required for the allelic divergence process can be quite short compared with the time required for gene duplication to proceed via neutral processes.

The Models

A General Model for Allelic Divergence

Fitness relationships

We employ an adaptive dynamics approach (Abrams et al. 1993; Dieckmann and Law 1996; Metz et al. 1996; Geritz et al. 1998) to model divergence of alleles at either a single diploid locus or at a pair of duplicate loci in an infinite population. This approach has been used to study the evolutionary divergence of genotypes when ecological pressures create disruptive selection through spatial heterogeneity or resource competition (Doebeli 1996; Dieckmann and Doebeli 1999; Kisdi and Geritz 1999; Geritz and Kisdi 2000; Van Dooren 2000; Kisdi 2001). These studies have found that evolutionary branching occurs under the same conditions in diploid models, but with more complicated dynamics (Geritz and Kisdi 2000). This modeling framework assumes that a continuum of possible alleles exists and that these alleles are linked by mutation. The population will evolve in time because of both population genetic changes in the allele frequencies and through the addition of new alleles by mutation and the loss of existing alleles by selection. The method approximates the evolutionary trajectory by assuming that a fixed number of resident alleles are present in the population and have reached their equilibrium frequencies. A mutant allele is then introduced into the population and is followed until it goes extinct or reaches an equilibrium frequency. Thus, population genetic dynamics act on a fast time scale, whereas mutational dynamics act on a slow time scale. This framework allows us to investigate the long-term evolutionary outcome in a way that is not possible using only two-allele population genetic models.

In our model, an allele is defined by a continuous variable that reflects its effect on gene function and therefore on organismal fitness. Thus, $W(x_1, x_2)$ is the fitness of an individual carrying one allele with value x_1 and another with value x_2. In the case of an individual carrying a duplicate copy of the

gene in question, fitness is a function of four variables $W(x_1, x_2, x_3, x_4)$. We assume that the duplicate and original locus behave identically, so the order of the variables has no effect on fitness.

The first step in defining an evolutionary model that incorporates gene duplication is to define fitness functions that relate the action of multiple alleles at single diploid loci to multiple alleles at a pair of duplicate loci. To insure that direct selection on copy number does not drive our results, we assume that fitness depends only on the relative frequency of alleles. This will occur, for instance, if total transcription levels are regulated or if the interaction between structural gene products depends on stochiometry. This assumption ensures that

$$W(x_1, x_2) = W(x_1, x_1, x_2, x_2), \tag{1}$$

which equates the fitness of an individual with only a single copy of the gene who has one copy of allele x_1 and another of x_2 with the fitness of an individual with a duplicate copy of the gene with two copies of allele x_1 and two copies of allele x_2. Starting only with this relationship, several features of the fitness function are apparent. First, there will not be any direct selection on duplication per se, because increasing copy number does not have an effect on fitness. Second, the relationship between heterozygote and homozygote fitness in the single-locus case and individuals carrying copies of different alleles in the duplicate case is related to the way that fitness changes when one allele is substituted at a single locus. Equation (1) expresses the relationship between fitness in individuals with single copies of the gene in question to fitness in individuals with two copies of the gene. We can use equation (1) to understand how new mutations that cause quantitative change in gene function will affect fitness in individuals with either a single copy of the gene or a duplicate pair of genes. Taking implicit derivatives of equation (1) yields the useful relationships

$$W_{\langle 0,1\rangle} = 2W_{\langle 0,0,0,1\rangle}, \tag{2}$$

$$W_{\langle 0,2\rangle} = 2(W_{\langle 0,0,0,2\rangle} + W_{\langle 0,0,1,1\rangle}), \quad \text{and} \tag{3}$$

$$W_{\langle 1,1\rangle} = 4W_{\langle 0,0,1,1\rangle}, \tag{4}$$

where for the moment we adopt the notational convention that the partial derivatives of the fitness function are denoted with subscripts for the number of partial derivatives with respect to that variable: $\partial W(x_1, x_2)/\partial x_1 = W_{\langle 1,0\rangle}$ and $\partial^2 W(x_1, x_2, x_3, x_4)/\partial x_3 \partial x_4 = W_{\langle 0,0,1,1\rangle}$.

Equations (2–4) relate changes in allele function in individuals with single copies of the gene to the same changes in allele function in individuals with duplicate copies of the gene. For instance, equation (1) says that replacing one copy of an allele in an individual with a single copy of the gene will have twice the effect on fitness as the same replacement (of only one allele) in an individual with duplicate copies of the gene. These derivatives determine whether novel mutations that have small fitness effects will spread in the population for a given genetic system.

Evolutionary attractors

A straightforward application of some classical results reveals that both the single-locus and duplicate-locus system will first evolve to the same monomorphic genetic state. This analysis follows the assumption that new mutations arise and segregate in a population until they either go extinct or become fixed. An evolutionary attractor at a single locus is distinguished by two criteria (Abrams et al. 1993; Metz et al. 1996; Geritz et al. 1998),

$$W_{\langle 1,0\rangle} = 0 \quad \text{and} \tag{5}$$

$$W_{\langle 1,1\rangle} + W_{\langle 0,2\rangle} < 0, \tag{6}$$

where all derivatives are evaluated at the evolutionary attractor. The first criterion implies that among mutations of small effect, none produce homozygotes with higher fitness than homozygotes of the attractor allele. The second criterion implies that the attractor allele can invade populations that are fixed for alleles on either side of the attractor allele. Incidentally, inequality (6) also implies that the attractor allele is at a local maximum for homozygous fitness, that is, a local maximum of the function $W(x, x)$. An allele that meets these conditions is an evolutionary attractor in that a population that is monomorphic for this allele cannot evolve to be monomorphic for another allele and conversely that the attracting allele will increase from low frequency in any population that is fixed for an allele near the attractor.

Attractors at a duplicate locus can be determined in a similar way by evaluating derivates of fitness in the two-locus model to yield:

$$W_{\langle 0,0,0,1\rangle} = 0 \quad \text{and} \tag{7}$$

$$3W_{\langle 0,0,1,1\rangle} + W_{\langle 0,0,0,2\rangle} < 0. \tag{8}$$

Substituting the relationships in equation (2–4) into expressions (7) and (8) produces the single-locus criteria of expressions (5) and (6). Thus, any allele that is an evolutionary attractor in a single-locus model is also an attractor after duplication. This means that the starting point for the creation of gene families through allelic diversification will be the same, regardless of whether duplication has occurred. Even though an attractor represents a monomorhpic state that cannot be replaced by other alleles, it is the evolutionary dynamics around such evolutionary attractors that allow evolutionary diversification (Metz et al. 1996; Geritz et al. 1998).

Genetic divergence

Three types of genetic divergence can occur following evolution toward an attractor. If the gene is currently present as only a single locus, then allelic divergence can occur, resulting in a stable polymorphism where two divergent specialized alleles are present. Simultaneous divergence can occur in a population in which a single copy of the gene is initially present, and then a duplication carrying a slightly diverged allele is selected for such that the duplication spreads and the alleles diverge simultaneously. Although in a deterministic model the likelihood of these two events happening simultaneously is infinitesimally small, some genetic diversity will always accumulate through drift. The simultaneous divergence criterion reveals when this accumulated variance can be used to drive the spread of a duplicate gene. This method allows us to understand how the spread of a duplication via drift can contribute to genetic divergence between newly duplicated loci.

If duplication has already occurred before any substantial divergence, then duplicate divergence can proceed, resulting in each locus becoming monomorphic for a different specialized allele. Divergence will occur if, at the evolutionary attractor, nearby alleles can invade, but not replace, the attractor. We derived the conditions for each type of divergence (allelic eq. 9, simultaneous eq. 10, duplicate eq. 11) by performing local stability analyses and used the relationship in equations (1–4) to relate them to the diploid fitness function (see Appendix 1 available online only at http://dx.doi.org/10.1554/05-507.1.s1):

$$\frac{\partial^2 W}{\partial x_2^2} > 0, \tag{9}$$

$$\frac{\partial^2 W}{\partial x_2^2} - \frac{1}{3}\frac{\partial^2 W}{\partial x_1 \partial x_2} > 0, \quad \text{and} \tag{10}$$

$$\frac{\partial^2 W}{\partial x_2^2} - \frac{1}{2}\frac{\partial^2 W}{\partial x_1 \partial x_2} > 0, \tag{11}$$

where the derivatives are all evaluated at the attractor. These inequalities determine when alleles near the evolutionary attractor can invade, but they also imply that the attractor allele will be lost (see Appendix 1 available online). Inequality (9) is equivalent to the generic result from evolutionary theory that simply states that disruptive selection at an attractor leads to divergence of alleles in the population (Abrams et al. 1993; Dieckmann and Law 1996; Metz et al. 1996; Geritz et al. 1998). Inequalities (10) and (11) show how this generic result is modified by the additional redundancy contributed by duplication. Because these expressions are evaluated at an evolutionary attractor, the additional condition that $\partial^2 W/\partial x_2^2 + \partial^2 W/(\partial x_1 \partial x_2) < 0$ (inequality 6) must hold and means that these conditions (9–11) are nested, with the requirements for allelic divergence being the most stringent and the requirements duplicate divergence the most relaxed (Fig. 1). This implies that, in general, the very same conditions that select for divergence of alleles at a single locus select for divergence of genes at duplicate gene pairs. Having additional loci allows finer scale exploration of the adaptive landscape. With two loci, the ratio of the alleles can be adjusted in intervals of 25% as compared to 50% for single-copy genes. In fact, as the number of duplicate copies increases, increasingly finer adjustments can be made. This means that if the single-locus fitness surface has a saddle point at the attractor, a population with an infinite number of loci can always climb up the side of the saddle (see Appendix 1 available online). In other words, the fitness function must have a saddle point for any kind of divergence to occur, but in populations with fewer loci the orientation of the saddle must be correct for divergence to occur. This analysis shows that only a small; finely tuned set of parameters will cause divergence to occur only after duplication; we call this set of parameters the ''Ohno zone'' because it reflects Ohno's (1970) belief that duplication was a necessary precursor to genetic divergence. Because members of gene families presumably start out as single copy genes, this implies that genetic divergence will typically occur before or during the spread of a duplication in the population.

Selection for duplication

In addition to explaining the emergence of the genetic divergence and specialization that characterize gene families, this model also provides an explanation for how gene duplicates are fixed by selection. Previous work has shown that the redundancy of gene duplicates can generate, at best, weak selection on the spread of the duplication (Clark 1994; Nowak et al. 1997; Wagner 1999; Walsh 2003), whereas heterozygote advantage can generate strong selection for duplication (Spofford 1969; Otto and Yong 2002; Walsh 2003; Proulx and Phillips 2005). In the system modeled here, heterozygote advantage is created whenever allelic divergence occurs, but is expected to be a transient phenomenon that is relieved by gene duplication. It is important to note that heterozygote advantage is not an assumption of the model, rather our results show that populations can evolve into the region where heterozygote advantage is present. Once heterozygote advantage has evolved, there will be positive selection for the spread of a gene duplication, which we call ''selection for duplication''. This spread can be measured by defining a linear invasion matrix for the duplicate copy of the gene that is initially present at low frequency. Recently, Otto and Yong (2002) used this method to show that, for a specific fitness function where heterozygotes gain the same benefit regardless of the number of copies of each allele, heterozygote advantage generates positive selection for gene duplication. This result can be extended to our model by determining the invasion rate of the duplication into a population with a single-copy gene that is polymorphic for alleles near the attractor. Without loss of generality, we assume that the allele x_1 is initially duplicated. We follow the number of gametes with the duplication that consist of haplotypes of (x_1, x_1) and (x_2, x_1). The linear invasion matrix is

$$\begin{bmatrix} (p \cdot W_H + q \cdot (1 - r) \cdot W_h)/\bar{W} & q \cdot r \cdot W_h/\bar{W} \\ p \cdot r \cdot W_h/\bar{W} & (p \cdot (1 - r) \cdot W_h + q \cdot W_h)/\bar{W} \end{bmatrix}, \tag{12}$$

where $W_H = W(x_1)$, $W_h = W(x_1, x_1, x_2)$, p is the frequency of allele x_1, $q = 1 - p$, and r is the rate of recombination between the two duplicate gene copies (see Appendix 2 available online only at http://dx.doi.org/10.1554/05-507.1.s2). Given our fitness function and the relationships of the fitness derivatives defined in the equations (1–4) and substituting in equilibrium values for p and q, the eigenvalue is always greater than one, and thus the duplication always spreads. This occurs because, under the assumptions of the model, haplotypes with the duplication are more likely to be involved in associations between alternate allelic variants and therefore benefit from heterozygote advantage. Tight linkage is known to play an interesting role in the spread of duplicate genes under heterozygote advantage because chromosomes with alternative allele copies can spread rapidly, but it may take some time to form (Otto and Yong 2002).

Our approach shows that allelic divergence is likely to occur under a similar large subset of the conditions that promote divergence of duplicate genes and that allelic divergence will produce positive selection on the spread of duplicate genes. Taken together, these results imply that allelic divergence will usually occur before duplication and lead to

the rapid spread of gene duplicates, suggesting the possibility that gene families can expand rapidly due to natural selection.

Dynamic Simulation Model

Reaction flux model

To further explore how particular features of gene function influence the creation of gene families and to test our analytical predictions, we have created a stochastic, sexual, individual-based simulation of genetic divergence and gene duplication. Many genes are expressed under a variety of conditions (reviewed in Hughes 1994), including transcription factors during development, metabolic proteins in different tissues, receptors that interact with multiple ligands, structural proteins that interact with multiple partners, and even genes whose function depends on environmental conditions (Fig. 2a). We derived a model loosely based on metabolic flux produced by an enzyme that operates in two tissues. Total organismal fitness is determined by the function of the protein in each tissue and involves no between-tissue epistasis. Tissue i has a different optimum protein (allele) state ζ_i, and the flux produced by a protein (z) is a Gaussian function $G_{[\zeta_i,\sigma]}(x)$ that has a maximum where the protein state matches the optimum protein for that tissue with variance σ^2. This model can be characterized by two main parameters, the difference between the tissue optima ($\hat{\zeta}$) and the specificity of the tissues (σ^2). Total flux within a tissue is determined by a Michaelis-Menton type saturating function:

$$\Phi_i = \frac{\sum_j \frac{z_j}{n}}{1 + \sum_j \frac{z_j}{n}}, \tag{13}$$

where n is the total number of alleles. This approach has been used to model both flux and transcription rates (Hurst and Randerson 2000; von Dassow et al. 2000; Salazar-Ciudad et al. 2001; Johnson and Brookfield 2003). Because we assume that there is no epistasis between tissues, total fitness is simply the product of tissue specific flux, $W = \mathbf{\Phi}_1 \cdot \mathbf{\Phi}_2$. This model is symmetric and an attracting point always exists midway between the tissue-specific optima. The fitness derivatives at this attracting point are

$$W_{\langle 0,2\rangle} = \frac{\hat{\zeta}^2 \exp\left(\frac{\hat{\zeta}^2}{4\sigma^2}\right) - 8\sigma^2\left[\exp\left(\frac{\hat{\zeta}^2}{8\sigma^2}\right) + \exp\left(\frac{\hat{\zeta}^2}{4\sigma^2}\right)\right]}{8\sigma^4\left[1 + \exp\left(\frac{\hat{\zeta}^2}{8\sigma^2}\right)\right]^4}$$

and (14)

$$W_{\langle 1,1\rangle} = -\frac{\hat{\zeta}^2 \exp\left(\frac{\hat{\zeta}^2}{8\sigma^2}\right)\left[2 + \exp\left(\frac{\hat{\zeta}^2}{8\sigma^2}\right)\right]}{8\sigma^4\left[1 + \exp\left(\frac{\hat{\zeta}^2}{8\sigma^2}\right)\right]^4}, \tag{15}$$

where $\hat{\zeta} = \zeta_2 - \zeta_1$. These expressions can be substituted into expressions (9–11) to show that divergence will only occur when the tissues are more different than a critical amount that depends on the strength of selection on enzymes within each tissue (σ^2; Fig. 2b).

Simulation of the flux model

We conducted individual-based simulations using the flux model. Haplotypes were defined by two allele values at two loci. Each allele could either be a null allele or have a value between the two tissue optima. A fixed number of alleles partition the allele space between the two endpoints. Adults were formed by assuming random mating, with adult fitness determined by the full complement of alleles (2, 3, or 4). The next generation of gametes was produced in frequencies determined by the parental fitness with recombination occurring between the two parental haplotypes. Because mutation was assumed to be rare, at most one mutational event could occur in each haplotype. Mutation could increase or decrease the allele value by one allele class, cause gene duplication in single-copy haplotypes, or cause the loss of a gene copy in duplicate-gene haplotypes. The population size was held constant by selecting the haplotypes for the next generation from a multinomial distribution with the probability for each haplotype set to the relative frequency of that haplotype.

In some simulations, populations were initialized with only single-copy genes present. Duplication was not allowed until generation 2500 so that the single-locus dynamics could equilibrate. At generation 2500 the duplication rate was set to 10^{-5}, and the rate of gene loss set to 10^{-3}. This scenario assumes that genes typically evolve as single-copy genes before gene duplication has an opportunity to occur. We also simulated populations that were initialized with the duplication already present in every haplotype. This scenario more closely matches traditional models of the evolution of gene duplicates where random drift is assumed to fix duplications.

We show results from simulations where there were 20 alleles, $\sigma^2 = 5$, the mutation rate between alleles is 0.0005, recombination occurs at rate 0.1, population size is 10,000, and the simulation was run for 10,000 generations. The value of $\hat{\zeta}$ varied through 4.5, 5.5, 7.0, 8.0 (Fig. 2b). The allelic mutation parameters were chosen to ensure that there were on average 10 new mutants per generation. We use these same parameter values to explore the relationship between population size and the relative timing of the allelic divergence process and a neutral gene duplication process, except that in these simulations $\hat{\zeta}$ is fixed at 7 and the allelic mutation rate is set to 10^{-5}.

Simulation results

The simulations show divergence under the conditions predicted by our analytical approach (Fig. 2c). When the tissues differ by a small amount, the population evolves to the attracting point where a single allele encoding a protein that functions well in both tissues is predominant in the population (Fig. 2c, row 1). When duplications arise by mutation, they remain at low frequency. Even if a duplication is already fixed in the population, selection keeps the population in a monomorphic state. For parameters in the Ohno zone, the single-locus population converges on the attractor and does not diverge even when duplicates are introduced via mutation (Fig. 2c, row 2). However, if the population is already fixed

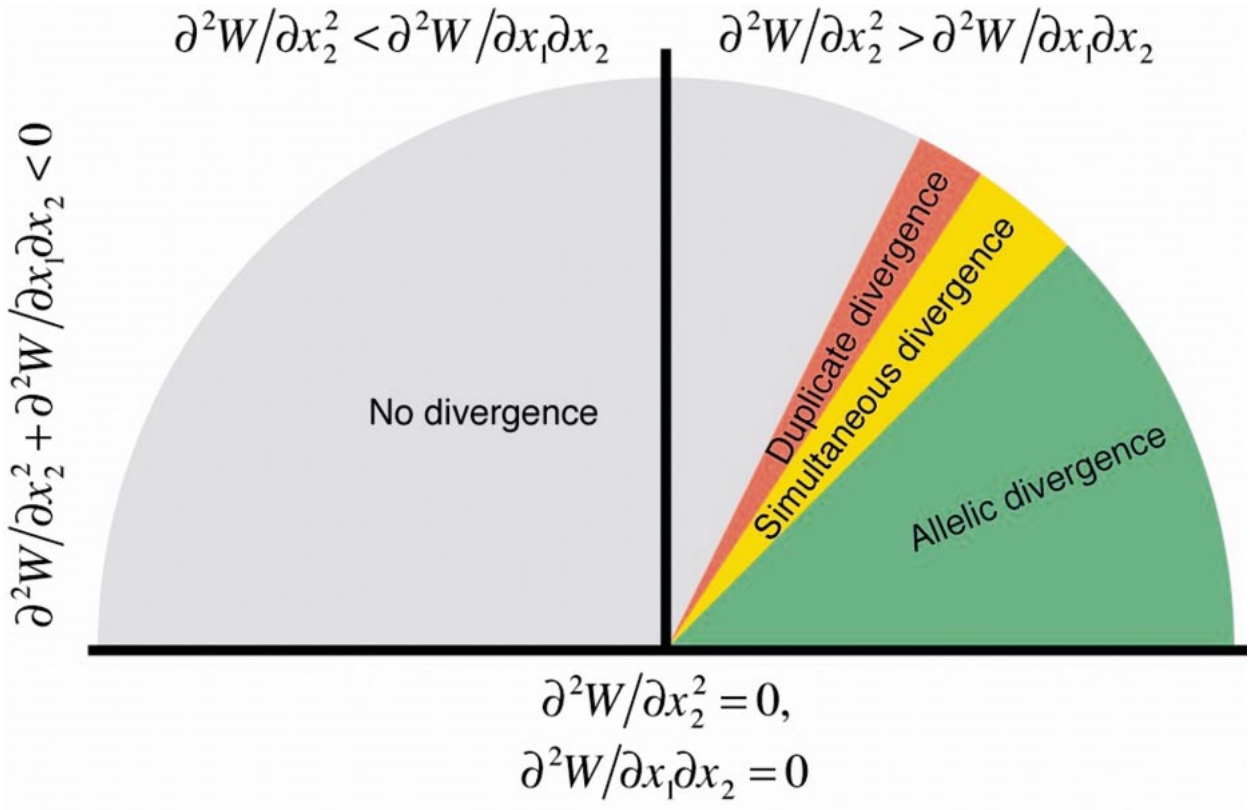

FIG. 1. Allelic divergence is the most likely form of divergence. Inequalities (2–4) are plotted to show the regions of parameter space that lead to the three kinds of divergence. These axes are transformed onto the vectors $[\partial^2W/\partial x_2^2, \partial^2W/(\partial x_1\partial x_2)]$ and $[\partial^2W/\partial x_2^2, \partial^2W/(\partial x_1\partial x_2)]$. The conditions for an evolutionary attractor restrict the parameters to the area in the plot. In the gray area on the left genetic divergence will not occur, regardless of the number of gene copies. In the red region labeled ''Duplicate divergence,'' divergence can only occur if the gene duplicate has already been fixed in the population. In the gold region labeled ''Simultaneous divergence,'' genetic divergence can occur if the duplication has already been fixed or while it is segregating. In the green region labeled ''Allelic divergence,'' divergence can occur at a single locus, while a duplication is segregating, or at a pair of duplicate loci. Because gene families must begin as single-copy genes, populations in the green region are likely to undergo genetic divergence before duplication occurs. Likewise, in the gold region divergence is likely to begin before a duplication is fixed in the population. Only in the red region is divergence likely to follow duplication.

for the duplication, then divergence does occur and each individual expresses two distinct proteins. When the tissues differ by a larger amount, the single-locus population still evolves to the evolutionary attractor, but when duplications arise they are able to spread into the population while genetic divergence accumulates (Fig. 2c, row 3). Because this part of the process depends on the availability of specific combinations of mutants, the duplication must achieve a moderate frequency before divergence can occur. For all larger values of the difference between the tissues, allelic divergence occurs in populations with one gene copy (Fig. 2c, row 4). Individuals in these populations will express either one protein specialized for only one tissue or two proteins, one specialized for each tissue. This is a rapid process and is not mutation limited. Following allelic divergence the gene duplication spreads through the population to fixation.

These simulations show how the stability conditions from equations (2–4) are realized dynamically. For populations in the Ohno zone, divergence of duplicate pairs proceeds much more slowly and stops at less extreme values than populations that experience simultaneous or allelic divergence (right column of Fig. 2c). This implies that even when duplications become fixed by chance in the Ohno zone, the amount of time required for divergence may be large enough to allow pseudogenes to form before a significant fitness advantage accumulates (Fig. 2c, row 3; see Walsh 2003). In contrast, populations in the allelic divergence zone undergo rapid genetic divergence that can be triggered by a small change in the parameters. This implies that changes in environmental conditions that alter the internal parameters could trigger a burst of allelic divergence and the sudden creation of many gene families.

While these simulations verify our analytic results by showing that allelic divergence can occur, we can also use the simulations to determine the relative time required for allelic divergence in populations subject to drift. The allelic divergence process can be divided into two phases, the initial phase of divergence at a single locus and the time required for a duplication event to occur and spread to fixation. We conducted a series of simulations in populations of different census size to quantify the effects of drift and selection on the time required for this process to occur. In this set of simulations, we set $\sigma^2 = 5$ and $\hat{\zeta} = 7$. These values were chosen because they are just outside of the Ohno zone and

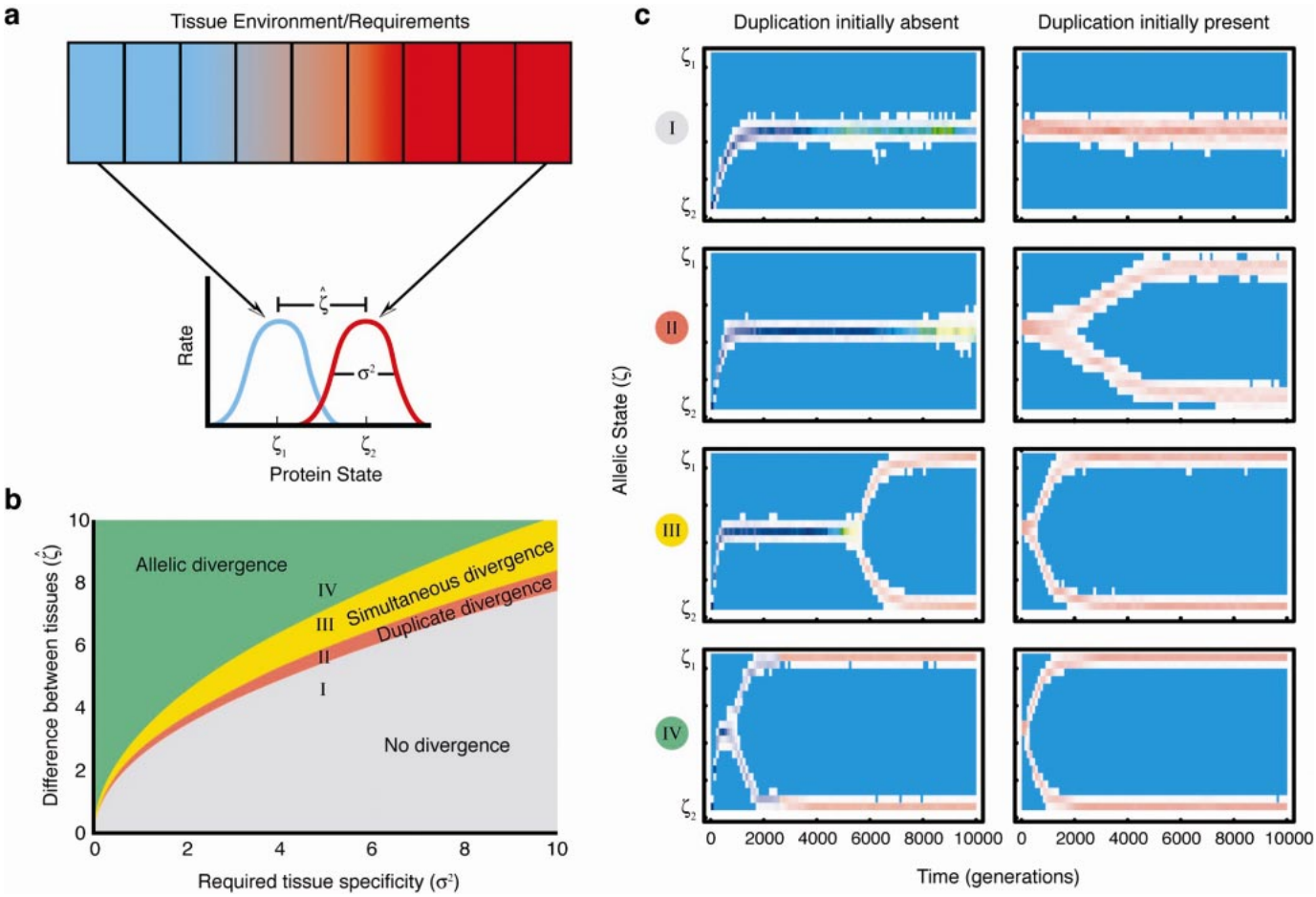

FIG. 2. Selection to perform in multiple tissues or cell types can lead to the evolutionary divergence among genes. (a) In our flux model, an organism is thought of as containing two distinct tissues that have different internal conditions. These distinct tissues can be due to a gradient in a transcription factor, differences in cell types, or even exposure to different environmental conditions. Differences in the tissues leads to differential performance of alleles expressing proteins within these tissues. The overall difference is given by $\hat{\zeta}$ and the specificity of a given tissue is given by σ^2. (b) Genetic divergence will evolve if the difference between the tissues exceeds a critical value. When the difference is below this threshold, divergence will not occur regardless of the duplication state of the gene (gray region). Increasing the difference above the threshold causes divergence to occur only if a duplication has already been fixed in the population (red region). At higher levels of difference, duplications can invade a population of single-copy individuals simultaneous with the divergence of alleles (gold region). Finally, for higher levels allelic divergence can occur before any duplication arises (green region). Only in the small red region is duplication expected to precede evolutionary divergence. (c) Individual based simulations of the dynamics of duplication and allelic divergence for the parameter values indicated by the roman numerals in b ($\hat{\zeta}$ = 4.5, 5.5, 7.0, 8.0). Each figure shows the results from a representative simulation run for a single population. In the panels on the left, the simulation begins with only one copy of the gene, with the allele value away from the attracting point, and duplication is not allowed to occur until generation 2500. The frequency of each allele is represented by the color intensity, while the hue represents the frequency of duplicate genes carrying each allele (blue represents single-copy genes and red represents duplicate genes). The population first moves toward the attracting point before divergence can occur. In the panels on the right, the simulation begins with the duplication already fixed. As the difference between the tissues increases, divergence occurs under a broader range of circumstances.

in the range where allelic divergence is predicted. Using these values gives a conservative estimate for the time required for allelic divergence, as larger values of $\hat{\zeta}$ will create stronger selection for allelic divergence and speed up the process.

We measured the time until polymorphism was achieved by running simulations that start with all individuals fixed for a single allele, the attractor allele (allele 10), at a single locus. The duplication rate was set to zero and the simulation was run until a polymorphism was fixed in the population. Polymorphism was defined as a population where the attractor allele (allele 10) and the neighboring alleles (alleles 9 and 11) had gone extinct and the frequency of alleles on either side of the attractor was greater than 0.4. Figure 3a shows results for 20 simulations for each population size shown.

The time until duplication following polymorphism was measured by initializing populations with a polymorphism at a single locus. The allele values were set at 8 and 12 at equal frequencies to reflect the state of the population at the end of the simulations of divergence time. The population could evolve through allelic substitutions at either locus and through duplication of the original locus. We compared these duplication times to the predicted waiting time for a duplication to fix under a neutral model (Fig. 3b).

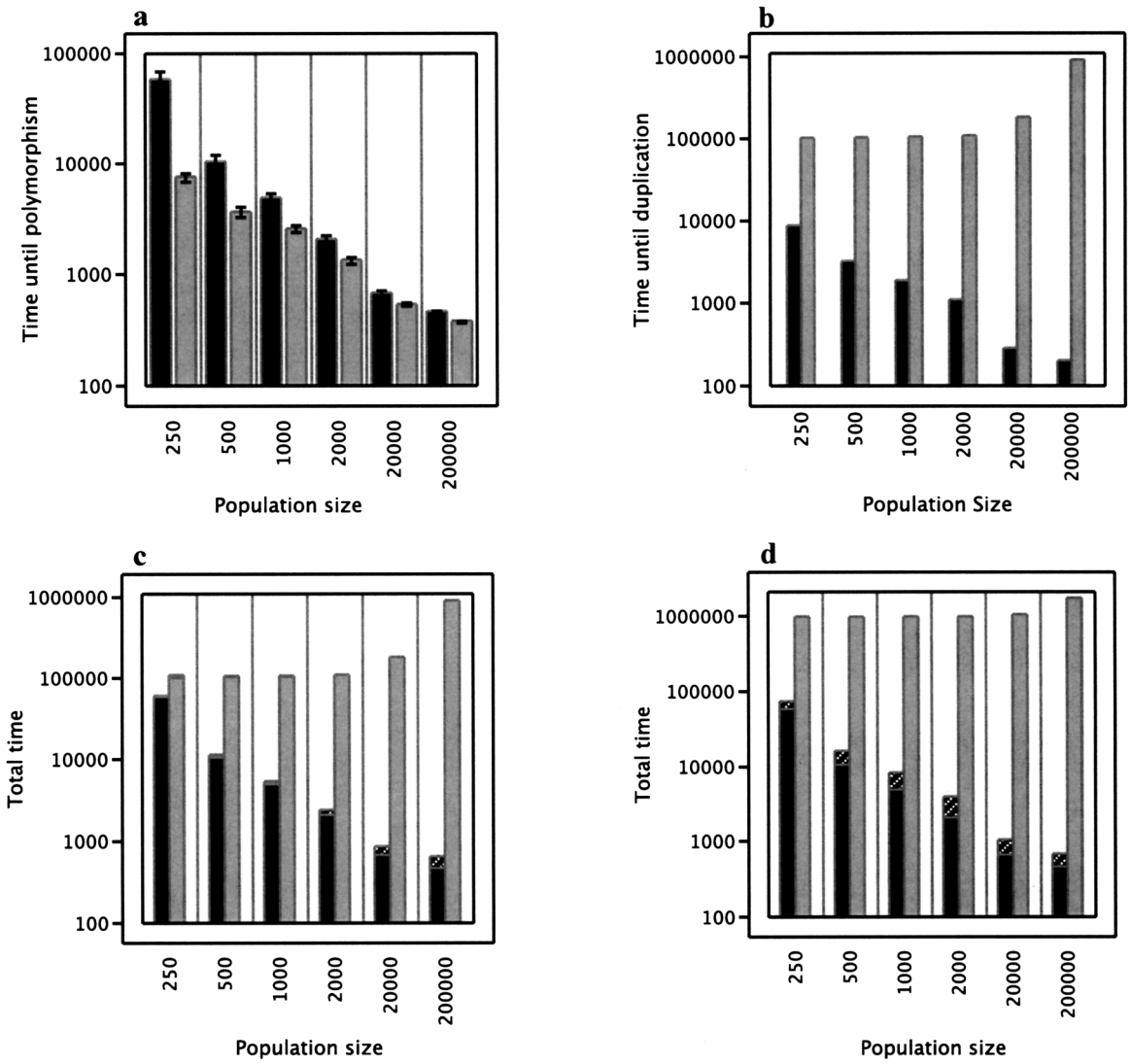

FIG. 3. Amount of time taken by components of the allelic divergence and gene duplication process. Panels compare mean time for the two phases of the process under the allelic divergence model proposed here with a model of neutral gene duplication followed by adaptive divergence. For all simulations shown there were 20 alleles, so that the attractor allele has the value 10, $\sigma^2 = 5$, $\hat{\zeta} = 7$, and recombination occurs at rate 0.1. (a) Time until polymorphism in populations with only a single copy gene (black bars) or a previously duplicated gene (light bars) shown on a log scale. Bars show mean time until the attractor allele and the two neighboring alleles were lost in 20 simulations with standard error bars. The allelic mutation rate was set to 10^{-5} and gene duplication was prohibited. (b) Time until duplication in populations that can undergo allelic divergence (black bars) and under a neutral model (light bars) shown on a log scale. Black bars show the mean time until all individuals carry two copies of the focal gene with the allelic mutation rate and duplication rate set to 10^{-5}. Light bars show the expected time until fixation of a gene duplicate. (c, d) Total time for the allelic divergence process and for a process of neutral duplication followed by divergence, shown on a log scale. The black bars show the sum of the time until polymorphism at a single-copy locus (solid) and the time until duplication of a polymorphic locus (hatched). Thus, the total height of the black bar represents the total time for the allelic divergence process to reach completion by first evolving polymorphism and then fixing a duplication. The light bars show the sum of expected time for duplication under the neutral model (solid) and the time until polymorphism at a previously duplicated locus (hatched). Thus, the light bars represent total time for a neutral duplication process followed by adaptive divergence (note that the hatched portion is too small to see). (c) Results for simulations with equal allelic mutation rate and duplication rate set to 10^{-5}. If the duplication rate is an order of magnitude lower at 10^{-6} (d), then time until duplication via neutral processes is orders of magnitude greater than the amount of time required for the entire allelic divergence process.

The total time required for each process was compared by summing the duplication time and the divergence time for each process. We calculated the total time of the allelic diveregence (AD) process by adding the mean waiting time until a polymorphism formed to the mean waiting time for duplication in a polymorphic population. This was compared to the total time of a process involving duplication by neutral processes followed by adaptive divergence of the two loci (NA). This is an underestimate of the NA process because the rate of gene loss is set to zero. The expected time until a gene duplicate is fixed under the NA process is given by $(1/v) + 4N$, where v is the duplication rate. Small populations have relatively low waiting times for neutral duplication but would also have similarly low waiting times for loss of the

duplicate locus through mutation and drift. A more accurate analysis would include the expected proportion of time the population spends in the duplicated state, which would depend on the ratio of the rate of duplication to the rate mutational inactivation of the gene.

When the duplication rate is the same as the rate of allelic mutation, the total time for the AD process is several orders of magnitude faster than that of the NA process in all but the smallest populations. In simulations where the duplication rate is lower than the allelic mutation rate (as is likely to be the case), the AD process is always much quicker than the NA process. These results are driven by the relaxation of selection in small populations, which allows drift to play a larger role. In very small populations the waiting time for a neutral duplication is relatively low, but the allelic divergence process becomes mutation limited and is subject to stochastic variation that prevents the appropriate combination of alleles on either side of the attractor from becoming common enough to drive further adaptation. Large population sizes work against the duplication phase of the NA process to such an extent that the difference between the AD and NA processes becomes even greater.

Discussion

Our model suggests a complete picture for the evolutionary formation and diversification of gene families. Because all genes begin as single copies, evolutionary forces have time to act on them before duplication. If conditions promote genetic divergence, then multiple different alleles will evolve and be maintained in the population. During this phase the population will exhibit heterozygote advantage and experience a segregation load. The presence of heterozygote advantage then creates strong positive selection on the spread of gene duplications as they arise, so that once a chance duplication event occurs it is expected to rapidly become fixed in the population. As the duplication spreads through the population, the allelic frequencies at each locus will diverge as each locus settles on a single allele. At this point a two-gene family has been created with natural selection acting at every time point to speed the process along. Thus, adaptive natural selection can play a substantial role in the formation of gene families in particular and genome evolution in general.

In our model we examined the evolution of gene families when there was no direct selection for or against gene duplication per se. This was accomplished through our central assumption that selection acts only on the relative dosage of alternative alleles, and not directly on copy number (eq. 1). In future work we will fully explore the effect that direct selection on dosage has on altering the process of allelic divergence. However, while selection for increased copy number could certainly explain the maintenance of duplicate genes, it cannot explain the evolution of functionally diverse gene families because strong stabilizing selection on the allele copy number would persist following duplication. Thus, relaxation of our fitness assumptions does not change the conclusion that divergence in gene families begins before duplication and in either case natural selection, rather than drift, dominates the fixation process.

Our mathematical analysis assumed that genotypes that are separated by a small number of mutations are phenotypically similar. This sort of assumption is necessary for the calculus of adaptive dynamics. However, the logical relationship between genotype fitnesses and the maintenance of polymorphism holds in the more general case of an arbitrary mapping between genotype and phenotype. As long as the phenotypes created by multiple-allele genotypes depend on the allelic values in the same way as for two-allele genotypes, evolutionary divergence of alleles will occur under similar conditions before and after duplication.

Comparison with other processes

This process differs from all recent conceptualizations of duplicate gene evolution, in that divergence between alleles can accumulate before duplication. Previous work has noted that multifunctionality can precede duplication (Hughes 1994, 1999) and that heterozygote advantage can promote duplication (Otto and Yong 2002; Walsh 2003), but studies have not investigated the relationship between the conditions that allow divergence before and after duplication. The surprising similarity between these conditions substantially alters our view of how gene families are created.

Our model allows for all divergence to occur following duplication, as it does in the traditional models, but our analysis shows that this is unlikely to occur. In general, the same conditions that promote divergence after a duplication has already been fixed promote divergence before duplication. Only in the small set of parameters comprising the Ohno zone is divergence unlikely at a single locus but selected for after duplication. Thus, genes that are capable of divergent selection as gene duplicates are likely to begin the process of divergence when they are present as only single-copy genes. The process of divergence, in turn, creates positive selection for duplication, promoting the process of gene family expansion.

Although heterozygote advantage is generated in our model and creates selection for the spread of gene duplicates, it is important to note that heterozygote advantage is not an assumption of our model and may not persist for long periods of time. Previous work has examined the duplication process when two alleles with fixed fitness effects are assumed to be present. For example, Walsh (2003) considered several models of subfunctionalization and neofunctionalization. In each scenario he considered, two alleles were segregating in the population either before duplication or while the population was polymorphic for a duplication. His model focused on an extreme fitness relationship where individuals carrying only the neofunctional allele had fitness of zero, but individuals heterozygous for the ancestral and neofunctional allele had a modest fitness increase. Our approach allows an arbitrary fitness function without specifying a fixed pair of segregating alleles. Because our model follows evolution in the adaptive landscape preceding duplication, it sheds light on the conditions that lead to the evolution of heterozygote advantage on the one hand and multifunctional genes on the other.

In some sense, the difference between this model and the DDC model is in the shape of the fitness landscape. In the DDC model, mutations are assumed to remove or reduce

functionality of one subfunction while preserving others. This is a mathematically degenerate fitness set because the derivative of fitness with respect to change at one allele at the locus (or pair of loci) has no effect on fitness, but if all alleles lose the ability to perform a single subfunction, then fitness is decreased (Force et al. 1999). This leads to a kind of degenerate saddle point in the fitness landscape at the attracting allele. Mutations that lose one subfunction have no fitness cost when rare, whether or not the population has one or two copies of the gene. In single-copy populations, however, as a subfunctionalized allele spreads it will sometimes be in the homozygous state and suffer a fitness cost. These subfunctional alleles can only be maintained at low frequency via mutation-selection balance in a similar manner to that of neofunctional alleles (Walsh 2003). In populations that already have duplicate copies of the gene, this degenerate saddle point allows the accumulation of complementary degenerate mutations and can lead to the maintenance of duplicate gene copies (Force et al. 1999; Lynch and Force 2000). This represents the extreme edge of the Ohno zone in our framework, where the criterion of inequality (11) is exactly zero and divergence is initiated through drift. If mutations can occur that decrease the ability of the gene to perform one subfunction but increase its ability to perform the other subfunction, then the potential for allelic divergence is opened up because the saddle point becomes nondegenerate, and the spread of alternative alleles will be driven by selection instead of drift.

Because allelic divergence will create selection to maintain both copies of a gene after duplication, duplicate genes that are maintained for long periods of time will be near or in the zone of allelic divergence, regardless of the mechanism of duplication. In particular, while we have focussed primarily on tandem duplication in this article, the same arguments apply to genes involved in whole-genome duplications. While many forces may lead to the maintenance of increased chromosome number after a whole-genome duplication, the fate of each duplicate gene pair will depend on selection at that pair of loci. If allelic divergence has occurred before the whole-genome duplication, then that pair of loci are expected to be maintained as true genes and diverge in function. It is even possible that the aggregate effect of allelic divergence at many individual loci could facilitate the fixation of a whole-genome duplication, although this effect will be ameliorated by negative selection at other loci where copy number is under stabilizing selection.

An alternative approach to the evolution of gene duplicates has focussed on the role of gene conversion in inhibiting divergence between recent gene duplicates (Ohta and Dover 1983; Walsh 1987). Gene conversion can cause one copy of a tandem gene duplicate to be converted to the other copy. This acts like a kind of biased mutational reversion to the most common allele, which will be the ancestral allele in the early stages of divergence. While we have not modeled this process explicitly, we believe that it will push the balance further in favor of allelic divergence. Gene conversion acts only in already duplicated loci, so it will have no effect on allelic divergence at single loci. In addition, because allelic divergence creates a polymorphism at the single-locus gene, the segregating alleles will not be at low frequency when the duplication event occurs. Gene conversion will have two effects on allele frequency dynamics at the duplicate locus. First, it will tend to push allele frequencies at the rare duplicate locus toward those in the ancestral locus through conversion from the copy on the homologous chromosome. Second, it will increase linkage disequilibrium by creating pairs of identical alleles on the same chromosome. The first effect will accelerate the initial spread of a duplication following allelic divergence, whereas the second will have a retarding effect. However, this retarding effect is at most as strong as the gene conversion rate and will act like a form of biased mutation. This retarding effect will be even stronger, however, when it affects genes that have duplicated without diverging because all of the alleles will initially be identical. Thus, gene conversion is expected to have more of a retardant effect on neofunctionalization and subfunctionalization processes than on the allelic divergence process.

Empirical evidence

Our model makes several predictions that are consistent with recently uncovered genomic patterns as well as novel predictions that have yet to be tested. Because allelic divergence is expected to precede duplication, paralogs will appear older than other evidence would suggest, which can explain the apparent burst of evolution after duplication (Dermitzakis and Clark 2001; Lynch and Conery 2003b). Several recent studies have revealed patterns of divergence that are inconsistent with both the DDC and the neofunctionality model (Notebaart et al. 2005; Crow et al. 2006). Crow et al. (2006) observed symmetric positive selection in recently duplicated hox genes but detect few synonymous substitutions. They argued that this pattern is inconsistent with both neofunctionalization and the DDC model. These data are consistent with the allelic divergence model because it predicts that both alleles will evolve adaptations to specific tissues and that the alleles will divergence rapidly and thus show short branch lengths.

The allelic divergence model predicts that alleles will diverge before duplicating, which would produce a signal of evolutionary divergence near the time of duplication. For example, when duplicates arise by whole-genome duplication, we expect genes that are maintained as duplicates to have orthologs in closely related species that are also in or near the zone of allelic divergence. These orthologs should have more genetic variation than orthologs of genes that are not maintained as duplicates, either because they have already diverged or because they are in the Ohno zone and experience relaxed selection even in the single-copy state (Fig. 2c).

Duplication and organismal complexity

In the DDC framework, duplicate gene pairs can diverge under neutral evolutionary processes through the specialization of each gene on some subfunction of the original single-copy locus. This process alone, however, cannot lead to an increase in organismal complexity unless there is selection for the subfunctions to diverge further or for them to be modified in some way from their state in the ancestral single-copy gene. While subfunctionalization may sometimes be the mechanism of duplicate gene divergence, selection for mul-

tifunctionality is the ultimate cause. The allelic divergence model shows how multifunctionality can arise at a single locus and how this can lead to the adaptive formation of gene families.

This theory points to a new role for sex, ploidy, and organismal complexity in the diversification of gene families. This mode of genetic diversification requires that multiple alleles be present in a single individual during the phase of the life cycle that they perform in. Organisms that are predominately haploid during their multicellular phase will therefore only be able to evolve gene families after duplications have been fixed by chance. Allelic divergence before duplication points to an adaptive explanation for the observation that gene duplicates persist for longer periods of time in more complex organisms (Lynch and Conery 2003b, 2004; Daubin and Moran 2004). Complexity within organisms produces conditions that favor specialization even before duplication and defines the processes that lead to the formation and diversification of gene families.

Acknowledgments

We thank J. Thornton, B. Cresko, B. Walsh, G. Wagner, and two anonymous reviewers for discussion and comments on the manuscript. This work was supported by an National Institutes of Health fellowship (GM06382) to SRP and by an National Science Foundation grant (DEB-0236180) to PCP.

Literature Cited

Abrams, P. A., H. Matsuda, and Y. Harada. 1993. Evolutionary unstable fitness maxima and stable fitness minima of continuous traits. Evol. Ecol. 7:465–487.

Chervitz, S. A., L. Aravind, G. Sherlock, C. A. Ball, E. V. Koonin, S. S. Dwight, M. A. Harris, K. Dolinski, S. Mohr, T. Smith, and many others. 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science 282: 2022–2028.

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

Crow, K. D., P. F. Stadler, V. J. Lynch, C. Amemiya, and G. P. Wagner. 2006. The ''fish-specific'' hox cluster duplication is coincident with the origin of teleosts. Mol. Biol. Evol. 23: 121–136.

Daubin, V., and N. A. Moran. 2004. Comment on ''the origins of genome complexity.'' Science 306:978.

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

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34:579–612.

Doebeli, M. 1996. A quantitative genetic competition model for sympatric speciation. J. Evol. Biol. 9:893–909.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Geritz, S. A. H., and E. Kisdi. 2000. Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. Proc. R. Soc. Lond. B 267:1671–1678.

Geritz, S. A. H., E. Kisdi, G. Meszena, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12:35–57.

Hammerstein, P. 1996. Darwinian adaptation, population genetics and the streetcar theory of evolution. J. Math. Biol. 34:511–532.

Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab, D. R. Nusskern, P. Wincker, A. G. Clark, J. M. C. Ribeiro, R. Wides, and many others. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298: 129–149.

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

———. 1999. Adaptive evolution of genes and genomes. Oxford Univ. Press, New York.

Hurst, L. D., and J. P. Randerson. 2000. Dosage, deletions and dominance: simple models of the evolution of gene expression. J. Theor. Biol. 205:641–647.

Johnson, L. J., and J. F. Brookfield. 2003. Evolution of spatial expression pattern. Evol. Dev. 5:593–599.

Kisdi, E. 2001. Long-term adaptive diversity in Levene-type models. Evol. Ecol. Res. 3:721–727.

Kisdi, E., and S. A. H. Geritz. 1999. Adaptive dynamics in allele space: evolution of genetic polymorphism by small mutations in a heterogeneous environment. Evolution 53:993–1008.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, and many others. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

———. 2003a. The evolutionary demography of duplicate genes. J. Struct. Funct. Genomics 3:35–44.

———. 2003b. The origins of genome complexity. Science 302: 1401–1404.

———. 2004. Response to comment on ''the origins of genome complexity.'' Science 306:978.

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

Lynch, M., M. O'Hely, B. Walsh, and A. Force. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804.

Metz, J. A. J., S. A. H. Geritz, G. Meszena, F. J. A. Jacobs, and J. S. van Heerwaarden. 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. Pp. 183–231 *in* S. J. van Strien and S. M. Verduyn Lunel, eds. Stochastic and spatial structures of dynamical systems. North-Holland, Amsterdam.

Notebaart, R. A., M. A. Huynen, B. Teusink, R. J. Siezen, and B. Snel. 2005. Correlation between sequence conservation and the genomic context after gene duplication. Nucleic Acids Res. 33: 6164–6171.

Nowak, M. A., M. C. Boerlijst, J. Cooke, and J. M. Smith. 1997. Evolution of genetic redundancy. Nature 388:167–171.

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin.

Ohta, T., and G. A. Dover. 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. Proc. Natl. Acad. Sci. USA 80:4079–4083.

Otto, S. P., and P. Yong. 2002. The evolution of gene duplicates. Adv. Genet. 46:451–483.

Phillips, P. C., and N. A. Johnson. 1998. The population genetics of synthetic lethals. Genetics 150:449–458.

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

Proulx, S. R., and P. C. Phillips. 2005. The opportunity for canalization and the evolution of genetic networks. Am. Nat. 165: 147–162.

Salazar-Ciudad, I., S. A. Newman, and R. V. Sole. 2001. Phenotypic and dynamical transitions in model genetic networks. I. Emergence of patterns and genotype-phenotype relationships. Evol. Dev. 3:84–94.

Spaethe, J., and A. D. Briscoe. 2004. Early duplication and functional diversification of the opsin gene family in insects. Mol. Biol. Evol. 21:1583–1594.

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

Van Dooren, T. J. 2000. The evolutionary dynamics of direct phenotypic overdominance: emergence possible, loss probable. Evolution 54:1899–1914.

von Dassow, G., E. Meir, E. M. Munro, and G. M. Odell. 2000. The segment polarity network is a robust developmental module. Nature 406:188–192.
Wagner, A. 1999. Redundant gene functions and natural selection. J. Evol. Biol. 12:1–16.
Walsh, B. 1987. Sequence-dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion? Genetics 117:543–557.
———. 2003. Population-genetic models of the fates of duplicate genes. Genetica 118:279–294.
Walsh, J. B. 1995. How often do duplicated genes evolve new functions? Genetics 139:421–428.
Young, J. M., B. M. Shykind, R. P. Lane, L. Tonnes-Priddy, J. A. Ross, M. Walker, E. M. Williams, and B. J. Trask. 2003. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. Genome Biol. 4:R71.

Corresponding Editor: L. Katz