The Limited Contribution of Reciprocal Gene Loss to Increased Speciation Rates Following Whole-Genome Duplication

Christopher D. Muir[1,2,*] **and Matthew W. Hahn**[1,3]

1. Department of Biology, Indiana University, Bloomington, Indiana; 2. Biodiversity Research Centre and Botany Department, University of British Columbia, Vancouver, British Columbia, Canada; 3. School of Informatics and Computing, Indiana University, Bloomington, Indiana

Submitted October 25, 2013; Accepted July 18, 2014; Electronically published November 26, 2014

Online enhancements: appendixes, zip file. *Dryad data:* http://dx.doi.org/10.5061/dryad.8k0m3.

ABSTRACT: Hybrid incompatibilities contribute to reproductive isolation between species, allowing them to follow independent evolutionary trajectories. Since hybrid incompatibilities are by definition deleterious, they cannot be selected for directly and must arise as a by-product of evolutionary divergence. Divergent resolution of duplicate genes, a special case of Dobzhansky-Muller incompatibilities, is one mechanism by which hybrid incompatibility can evolve. Following whole-genome duplication, loss of gene copies could possibly increase the opportunity for divergent resolution and, hence, the evolution of hybrid incompatibilities. However, divergent resolution can take place only when populations are isolated in allopatry; genes lost within a species cannot contribute to future speciation. Furthermore, nearly complete allopatry is necessary for passive divergent resolution. Using mathematical models, we demonstrate that these two factors severely impede the ability of divergent resolution alone to increase speciation rates, except under very particular conditions. Instead, we find that the population dynamics of diverging lineages dominate this process, leading to a larger role for ecology relative to genetics in the origin of new species, even by passive mechanisms. Divergent resolution of duplicate genes might increase speciation rates in some clades at some times, but our results indicate that it alone is unlikely to account for the macroevolutionary success of polyploid clades.

Keywords: whole-genome duplication, polyploid, reciprocal gene loss, speciation.

Introduction

Branches on the tree of life vary enormously in their species richness. For example, there are 22,750 described species in the plant family Asteraceae, whereas the much older Ginkgo clade contains only a single extant species, *Ginkgo biloba* (Stevens 2001–). Variation in diversification rates, the difference between speciation and extinction, is thought to be one of several proximate causes generating variation in clade richness (Rabosky 2009). Disparity in diversification rates reflects a wide variety of biological, historical, and geographical factors (Coyne and Orr 2004; Jablonski 2008).

Major features of genome architecture might predispose certain lineages toward higher diversification rates. The observation that some highly successful, species-rich lineages have a history of whole-genome duplication (WGD, also referred to as polyploidy) suggests a possible causal relationship (Taylor et al. 2001; Aury et al. 2006; Scannell et al. 2006; Sémon and Wolfe 2007; Soltis et al. 2009). Although there are several potential explanations for an association between ancient WGD and diversification, reciprocal gene loss is one possible mechanism linking WGD with increased speciation rates (Werth and Windham 1991; Lynch and Force 2000; Edger and Pires 2009; McGrath and Lynch 2012; Fawcett et al. 2013). As we explain below, reciprocal gene loss can result in hybrid incompatibilities, which are important components of total reproductive isolation in many species (Macnair and Christie 1983; Coyne and Orr 2004; Rieseberg and Willis 2007; Phadnis and Orr 2009; Bedinger et al. 2011). Reproductive isolation is central to speciation because it allows populations previously linked by gene flow to diverge along independent evolutionary trajectories (Mayr 1942; Coyne and Orr 2004). Thus, events such as WGD that increase the rate of reciprocal gene loss could plausibly increase speciation as well.

Hybrid incompatibilities that evolve by reciprocal gene loss are a special case of Dobzhansky-Muller incompatibilities (DMIs). Darwin famously recognized that hybrid incompatibilities (sterility or inviability) could not be the direct product of natural selection but instead must have evolved as a by-product of divergence for other reasons

* Corresponding author; e-mail: cdmuir@biodiversity.ubc.ca.

Am. Nat. 2015. Vol. 185, pp. 70–86.
0003-0147/2015/18501-55057$15.00.
DOI: 10.1086/679108

(Darwin 1859). Dobzhansky (1937) and Muller (1942) demonstrated that evolving hybrid incompatibilities does not require passing through valleys of low fitness. Rather, they proposed that hybrid sterility and inviability are caused by deleterious epistatic interactions between loci from both species that have not previously been "tested" in the same background during their evolutionary history. Although Muller (1942), in particular, acknowledged that a variety of evolutionary forces and genetic mechanisms could be responsible for hybrid incompatibilities, it is commonly assumed that either natural selection on "normal" function results in interactions between derived, protein-altering substitutions (Coyne and Orr 2004; Wright et al. 2013) or that genetic conflict drives the evolution of intrinsic incompatibilities (Phadnis and Orr 2009; Presgraves 2010). Reciprocal loss of duplicated genes between two species acts as a DMI (Werth and Windham 1991; Lynch and Force 2000). Reciprocal gene loss occurs when one population loses one gene copy and a second population loses the second copy at a different genomic location (fig. 1). If the gene is required for fertility and/or viability, hybrids with null genotypes will suffer a fitness cost. The same effect can occur via the divergent resolution of individual functions of duplicate genes, without the loss of the entire gene; this is also known as subfunctionalization (Lynch and Force 2000).

Since reciprocal gene loss can cause reproductive isolation, it might be a cause of elevated speciation rates following WGD, when the complement of duplicate genes greatly expands in a single generation. Although the idea was initially controversial (Coyne and Orr 2004; Orr et al. 2007), there is now direct evidence that reciprocal gene loss causes incompatibilities between ecotypes of *Arabidopsis* (Bikard et al. 2009), between species of *Drosophila* (Masly et al. 2006), and between experimental populations of yeast (Maclean and Greig 2011). Gene loss likewise contributes to hybrid incompatibility in *Xiphophorus* fishes (Schartl 2008), albeit not because of divergent resolution. Note, however, that all of these cases arose via standard, small-scale gene duplication and loss rather than WGD. Nevertheless, reciprocal gene loss has been suggested as a cause of speciation in macroevolutionary studies, most notably in a "burst" of speciation following WGD in yeast (Scannell et al. 2006). Sustained periods of elevated speciation in teleost fishes (Taylor et al. 2001; Sémon and Wolfe 2007), ferns (Werth and Windham 1991), and several angiosperm lineages (Soltis et al. 2009) have also been attributed to reciprocal gene loss following WGD. However, newer work has demonstrated that recently polyploid plant lineages actually have lower speciation rates (Mayrose et al. 2011). A theoretical analysis of the reciprocal gene loss hypothesis might help to make sense of the discrepancy between studies.

A heretofore unappreciated complication with the reciprocal gene loss hypothesis is that only a fraction of the thousands of genes duplicated during WGD can actually contribute to reproductive isolation. Following WGD, genomes quickly lose most gene duplicates because of relaxed selection on redundant duplicates, fueling reciprocal gene loss and hybrid incompatibility. However, genes lost within an interbreeding population are no longer available to contribute to future reproductive isolation. As we show below, only allopatric populations can accumulate reciprocal gene loss DMIs that contribute to reproductive isolation and hence that complete speciation. The impact of reciprocal gene loss on speciation rates following WGD therefore depends on how fast duplicate genes are lost and how

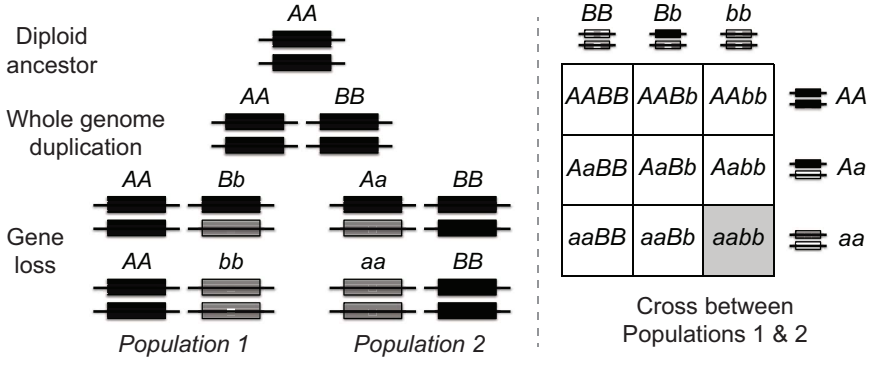

Figure 1: Duplicate-gene loss leads to hybrid incompatibilities after whole-genome duplication (WGD). *Left*, after WGD, there are two functional gene copies (black bars). Because of functional redundancy, one copy is free to accumulate mutations that make it nonfunctional (white bars). In allopatric populations, independent mutations might hit different gene copies. *Right*, upon secondary contact, hybrids between populations will segregate genotypes lacking any functional copies (*aabb* genotypes in bottom right of square), which can lead to hybrid incompatibility.

Table 1: Glossary of symbols

Symbol	Description
Dynamics of hybrid incompatibility:	
N	Number of genes in ancestral diploid genome
$N_j(t)$	Number of genes in genome of population j at time t after WGD
d	Rate of duplicate-gene loss per million years
$P(D_{i,j}, t)$	Probability that a duplicate-gene pair i is resolved to a single copy in population j at time t
$I(t)$	Number of hybrid incompatibilities at time t
p	Probability that a divergently resolved duplicate pair causes hybrid incompatibility
$\tau(k)$	Waiting time until the kth hybrid incompatibility
I_{spec}	Number of hybrid incompatibilities needed to complete speciation
φ	Speciation "fuel" (i.e., number of possible speciation events)
Macroevolutionary model:	
T_{sym}	Average time in sympatry after speciation
$T_{\text{allo}-\text{sym}}$	Average time shifting between sympatry and allopatry before speciation is complete
T_{allo}	Average time in allopatry during which enough hybrid incompatibilities evolve to complete speciation
ρ	Transition rate between geographic states (allopatry ↔ sympatry)
L_{spec}	Length of time needed in allopatry to complete speciation by fixing I_{spec} hybrid incompatibilities
T_{spec}	Average total time to speciation ($T_{\text{sym}} + T_{\text{allo}-\text{sym}} + T_{\text{allo}}$)
L	Length of time (random variable) in a particular geographic state
l	Particular realized duration of random variable L
U	Number of rounds (random variable) of allopatry without speciation
u	Particular realized number of U
S	Number of species
λ_t	Instantaneous speciation rate per million years at time t after WGD
λ	Average speciation rate per million years

Note: WGD = whole-genome duplication.

often allopatry—presumably caused by geographic isolation—enables divergence unimpeded by gene flow.

In this article, we analyze a model of gene loss and speciation following WGD. Although there are many effects of WGD on genomes and organisms, here we focus only on the passive loss of genes and gene functions and not on neofunctionalization or the role of selective gene loss in local adaptation. First, we consider the dynamics of gene loss within a single pair of diverging species. Next, we analyze a macroevolutionary model by extending the analysis to multiple rounds of speciation. This model shows that reciprocal gene loss can potentially cause rapid speciation (≈1 Myr^{-1}) following WGD, leading to at most a few hundred species in a clade over ≈10–100 Myr. However, our model indicates that such outcomes are possible only under very particular conditions. More often, population processes that affect how fast allopatry is attained determine realized speciation rates. Even when many potential genetic hybrid incompatibilities are available, population processes can prevent most from actually contributing to speciation. If reciprocal gene loss is the primary cause of speciation following WGD, our model predicts at best a weak and highly erratic relationship between WGD and speciation rates. Finally, using population genetic models, we also demonstrate that the assumption of complete allopatry is quite strict; even low levels of gene flow will thwart reciprocal gene loss.

Material and Methods

Dynamics of Hybrid Incompatibility between Two Populations

Table 1 provides a glossary of symbols used in the models. In this first model, we examine the accumulation of hybrid incompatibilities by reciprocal gene loss between two populations that become allopatric following WGD. Initially, genome size increases from N to $2N$ genes, so that the number of total genes at the start of the process we model is equal to

$$N_1(0) = N_2(0) = 2N. \tag{1}$$

We model the change in gene number over time in two allopatric populations using differential equations:

$$\frac{dN_1}{dt} = -d(N_1 - N), \tag{2a}$$

$$\frac{dN_2}{dt} = -d(N_2 - N). \tag{2b}$$

Under this model, the number of genes in each genome decays exponentially to N, the original diploid complement, depending on the rate of gene loss, d. Exponential decay of gene number is a common feature of post-WGD genome evolution (Lynch and Conery 2000, 2003; Maere et al. 2005; Scannell et al. 2006; Vanneste et al. 2013). The general solution to the number of genes at time t is

$$N_1(t) = N(1 + e^{-dt}), \tag{3a}$$

$$N_2(t) = N(1 + e^{-dt}). \tag{3b}$$

The expected number of incompatibilities at time t is given by (1) the probability that a pair of duplicate genes, i, have been resolved in both populations; (2) the probability that resolution is divergent (0.5); and (3) the probability, p, that divergent resolution leads to a DMI. The probability that a duplicate gene pair i has been resolved at time t in population j is

$$\begin{aligned} P(D_{i,j}, t) &= \frac{N_j(0) - N_j(t)}{N} \\ &= \frac{2N - N(1 + e^{-dt})}{N} \\ &= 1 - e^{-dt}. \end{aligned} \tag{4}$$

The number of incompatibilities $I(t)$ between populations 1 and 2 at time t is

$$I(t) = \int_{i=0}^{N} \frac{p}{2} P(D_{i,1}, t) P(D_{i,2}, t) dN = \frac{Np}{2}(1 - e^{-dt})^2. \tag{5}$$

The number of incompatibilities asymptotically approaches the maximum number of possible incompatibilities, $Np/2$. Although gene loss and divergent resolution are probabilistic phenomena, we treat them throughout as deterministic (see app. A for further details on deterministic vs. stochastic analyses; apps. A–F available online). Another important parameter is the waiting time to the kth incompatibility, which we obtained by setting $I(t) = k$ and solving for t. The waiting time as a function of k increases inversely with the rate of resolution, d:

$$\tau(k) = \frac{-\log(1 - \sqrt{2k/Np})}{d}. \tag{6}$$

Figure 2 depicts the dynamics of incompatibility accumulation between two species.

We assume that, to complete speciation, a minimum number of incompatibilities, I_{spec}, must accumulate in order to prevent homogenization upon secondary contact. The total number of potential hybrid incompatibilities, $Np/2$, divided by I_{spec}, gives the total number of potential rounds of speciation, which we denote φ:

$$\varphi = \frac{Np}{2I_{spec}}. \tag{7}$$

For example, if it takes 10 hybrid incompatibilities to complete speciation and 10% of divergently resolved duplicated genes cause incompatibilities, then for a genome of 20,000 genes, $\varphi = 100$ potential rounds of speciation, leading to an astronomical 10^{30} species in the absence of extinction. Since φ defines the store of potential speciation events, we refer to it metaphorically as speciation "fuel." As we describe below, φ is one of three key parameters determining the speciation rate, but only a small fraction of potential hybrid incompatibilities will actually contribute to speciation.

Macroevolutionary Model

Now that we have a model for how hybrid incompatibilities evolve over time after WGD (eq. [5]), we can use this to ask how a clade's diversity expands through multiple rounds of speciation. Some technical aspects of this macroevolutionary model are complex. Therefore, we first verbally summarize the model in order to clearly lay out its basic structure and underlying assumptions.

First, WGD occurs at $t = 0$, the root of the phylogeny. At this point there is a single species in a single geographic location (sympatry). We assume that the polyploid species is completely isolated from its diploid progenitor, which we ignore in this model.

Second, within a species, duplicate genes are lost through mutation, but divergent resolution is prevented in sympatric populations by gene flow. We do not distinguish between completely sympatric and partially geographically isolated, parapatric populations because their behavior is nearly identical with respect to divergent resolution. The population genetic model presented below shows that divergent resolution is nearly impossible when there is any appreciable migration between populations.

Third, after some time (T_{sym}), a sympatric population becomes instantaneously and completely allopatric (migration = 0). Duplicate genes are lost in both daughter populations independently, and divergent resolution can occur.

Fourth, after some time in allopatry, populations can become sympatric again. At this point, either populations dedifferentiate into a single population or speciation is

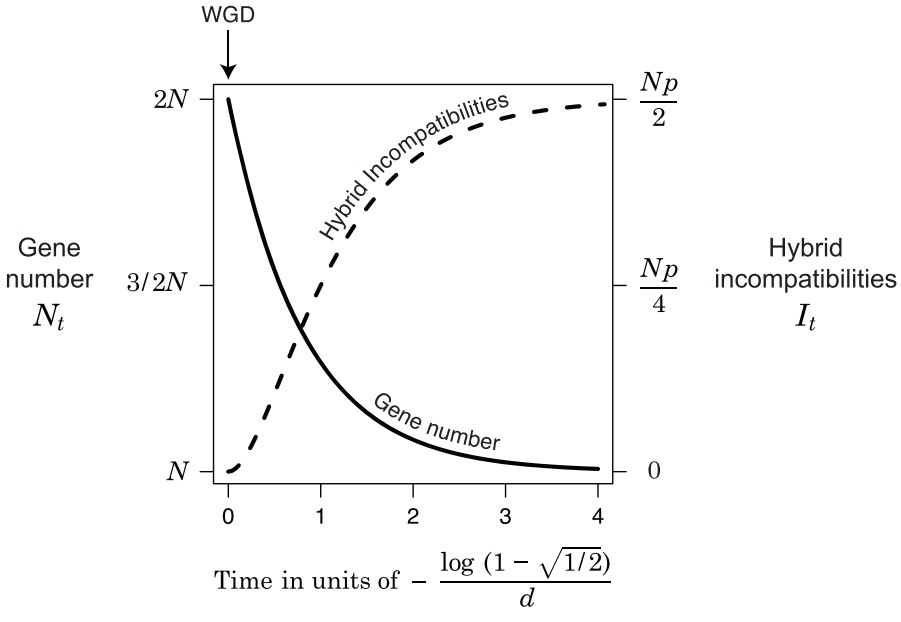

Figure 2: Dynamics of gene loss leads and hybrid incompatibility following whole-genome duplication (WGD). Duplicate genes are lost (solid line) at a constant rate d back to the original genome size ($2N \rightarrow N$), as described by equations (3). Half of all duplicate genes are lost ($N_t = 3/2N$) by $\log 2/d$. As duplicate genes are lost, hybrid incompatibilities accumulate (dashed line), following equation (5). The time to accumulate half of all possible incompatibilities is $-\log(1 - (1/2)^{1/2})/d$, which is the unit of time depicted on the X-axis.

complete if I_{spec} hybrid incompatibilities have accumulated in allopatry.

If speciation is incomplete, all hybrid incompatibilities that had evolved in allopatry are instantaneously purged by selection. Numerical simulations indicate that this last assumption is valid for a large range of parameter space (app. B). The process begins again at the third step until speciation is complete. We denote the total time spent in periods of sympatry and allopatry before speciation is complete as $T_{allo-sym}$.

If speciation is complete, the process begins again at the second step in each new descendent species. The time spent in this last period of allopatry is T_{allo}.

For each speciation event, one initial period of sympatry (T_{sym}) and one period of allopatry (T_{allo}) are necessary. In between, there can be 0 or an infinite number of transitions between allopatry and sympatry without speciation ($T_{allo-sym}$). Hence, the total waiting time for speciation is

$$T_{spec} = T_{sym} + T_{allo-sym} + T_{allo}. \quad (8)$$

Next, we analytically derive the expected waiting times at any point t after WGD, using a Markov chain model of shifting between geographic states (allopatry ↔ sympatry). From this, we can calculate average speciation rates over macroevolutionary time if divergent resolution of duplicate genes were the sole cause of reproductive isolation. In addition to calculating average speciation rates, we examine the variance, using stochastic simulations. The key insight from these models is that not all genes lost during either T_{sym} or $T_{allo-sym}$ actually contribute to realized reproductive isolation and hence speciation. Furthermore, many hybrid incompatibilities that evolve during T_{allo} are superfluous in the sense that, once I_{spec} incompatibilities have evolved, any beyond that do not increase the speciation rate.

Modeling the Waiting Time to Speciation

Equation (8) shows that the waiting time to speciation is determined by the amount of time would-be species spend in various geographic modes (allopatry or sympatry). We treat the geographic mode of diverging populations at any time probabilistically as a time-homogeneous Markov chain with two discrete states, sympatric and allopatric. In this section, we derive the expected value of parameters in equation (8); in appendix A, we analyze parameter stochasticity, using simulations. The geography of gene flow between populations and its evolution through time in nature are undoubtedly complex, but for analytical tractability we use this simple model of geography. The model captures the central point that many potential hybrid incompatibilities do not actually contribute to speciation. We discuss the effect of relaxing geographic assumptions in appendix C.

In sympatry, there is no divergent resolution of duplicate genes. In allopatry, duplicate-gene pairs can be resolved divergently, and incompatibilities evolve following equation (5). The transition probability from sympatry to allopatry is ρ, and therefore the average waiting time to allopatry after a previous speciation event is simply the inverse of the rate,

$$T_{\text{sym}} = \rho^{-1}. \tag{9}$$

Once incipient species are allopatric, speciation is not guaranteed, since populations must evolve at least I_{spec} hybrid incompatibilities to prevent fusion upon secondary contact. Secondary contact occurs at a rate determined by the transition probability of going from allopatry back to sympatry, which we assume is also ρ. In other words, the probability of populations going from sympatry to allopatry is equal to that of going from allopatry to sympatry. Once in sympatry, populations can become allopatric once more with probability ρ, and so on until they are allopatric long enough to complete speciation.

Determining the time this process takes ($T_{\text{allo}-\text{sym}}$) is complex, but we start by specifying that there is some length of time (L_{spec}) of continuous allopatry required to fix I_{spec} incompatibilities and complete speciation. If allopatry is any less than L_{spec}, species fuse; if allopatry is equal to or longer than L_{spec}, speciation is complete. We derive L_{spec} in terms of model parameters below. Note that L_{spec} will typically be shorter than the actual time spent in allopatry (T_{allo}), because populations can remain allopatric even after there are enough incompatibilities to complete speciation. Given a constant transition probability ρ, waiting times in one geographic mode or the other are an exponentially distributed random variable that we denote L. The probability density of a particular waiting time l is therefore $f(l) = \rho e^{-\rho l}$.

The time $T_{\text{allo}-\text{sym}}$ is the sum of u waiting times in sympatry and allopatry, where each waiting time in sympatry can last any length of time with probability $f(l)$, but the waiting time in allopatry must be less than L_{spec}; otherwise, speciation would be complete. Hence, u is the number of rounds of sympatry and allopatry without speciation. For example, imagine that allopatric populations become sympatric before speciation is complete, then complete speciation during a second allopatric period. In this case, the number of allopatric and sympatric waiting times, u, is 1, and the string of geographic states is $\text{Allo}_1 \rightarrow \text{Sym}_1 \rightarrow \text{Allo}_{\text{spec}}$. The average duration of the first allopatric period can be determined from $f(l)$, conditional on l being too brief to complete speciation ($l < L_{\text{spec}}$). Since waiting times are exponentially distributed with rate ρ, the average duration of allopatry without speciation is

$$\mathbb{E}(L|l < L_{\text{spec}}) = \frac{\int_0^{L_{\text{spec}}} lf(l)dl}{\int_0^{L_{\text{spec}}} f(l)dl} = \frac{1 - e^{\rho L_{\text{spec}}} + \rho L_{\text{spec}}}{\rho(1 - e^{\rho L_{\text{spec}}})}. \tag{10}$$

The average duration of a sympatric period between allopatric periods (e.g., Sym_1 above) is simply $\mathbb{E}(L) = \rho^{-1}$, the inverse of the transition rate. If $u = 1$, then on average $T_{\text{allo}-\text{sym}} = \mathbb{E}(L|l < L_{\text{spec}}) + \mathbb{E}(L)$. For any given $u = U$, $T_{\text{allo}-\text{sym}} = u(\mathbb{E}(L|l < L_{\text{spec}}) + \mathbb{E}(L))$. However, u is not predetermined and could potentially vary between 0 and ∞. If ρ is low enough, such that speciation is nearly always complete in the first period of allopatry, then u will usually be 0. However, if ρ is relatively high, then it will take many trials (high u) before populations are allopatric long enough to complete speciation.

Revisiting the simple case where $u = 1$, the probability of that particular string of geographic states ($\text{Allo}_1 \rightarrow \text{Sym}_1 \rightarrow \text{Allo}_{\text{spec}}$) is therefore the probability that the first allopatric period lasted less than L_{spec} and that the second was greater than or equal to L_{spec}. Formally, this can be expressed as $P(l < L_{\text{spec}})P(l \geq L_{\text{spec}})$. If, instead, $u = 2$, then the string of geographic states is $\text{Allo}_1 \rightarrow \text{Sym}_1 \rightarrow \text{Allo}_2 \rightarrow \text{Sym}_2 \rightarrow \text{Allo}_{\text{spec}}$, and the probability is $P(l < L_{\text{spec}})^2 P(l \geq L_{\text{spec}})$. In general, the probability that there are $u = U$ periods of allopatry before speciation, $\text{Allo}_1 \rightarrow \text{Sym}_1 \rightarrow \text{Allo}_2 \rightarrow \text{Sym}_2 \rightarrow \ldots \rightarrow \text{Allo}_u \rightarrow \text{Sym}_u \rightarrow \text{Allo}_{\text{spec}}$, is $P(l < L_{\text{spec}})^u \times P(l \geq L_{\text{spec}})$. These probabilities can be derived by integrating over portions of $f(l)$:

$$P(l < L_{\text{spec}}) = \int_0^{L_{\text{spec}}} f(l)dl = 1 - e^{-\rho L_{\text{spec}}}, \tag{11}$$

$$P(l \geq L_{\text{spec}}) = \int_{L_{\text{spec}}}^{\infty} f(l)dl = e^{\rho L_{\text{spec}}}. \tag{12}$$

The total average waiting time ($T_{\text{allo}-\text{sym}}$) until a period of allopatry long enough for speciation to occur is the sum of time in sympatry and allopatry, $u(\mathbb{E}(L|l < L_{\text{spec}}) + \mathbb{E}(L))$, weighted by the probability that $u = U$ integrated over all u is

$$\begin{aligned} T_{\text{allo}-\text{sym}} &= \int_0^{\infty} u[\mathbb{E}(L|l < L_{\text{spec}}) + \mathbb{E}(L)] \\ &\quad \times P(l < L_{\text{spec}})^u P(l \geq L_{\text{spec}})du \\ &= \frac{2e^{pL_{\text{spec}}} - \rho L_{\text{spec}} - 2}{\rho e^{\rho L_{\text{spec}}}(e^{\rho L_{\text{spec}}} - 1)\log(1 - e^{-\rho L_{\text{spec}}})^2}. \end{aligned} \tag{13}$$

This unwieldy expression can be closely approximated because $e^{\rho L_{\text{spec}}}(e^{\rho L_{\text{spec}}} - 1)\log(1 - e^{-\rho L_{\text{spec}}})^2$ quickly approaches unity as L_{spec} increases from 0. Therefore,

$$T_{\text{allo}-\text{sym}} \approx \frac{2e^{\rho L_{\text{spec}}} - \rho L_{\text{spec}} - 2}{\rho}. \quad (14)$$

Now that we have derived analytical expressions for T_{sym} (eq. [9]) and $T_{\text{allo}-\text{sym}}$ (eq. [14]), to finish calculating T_{spec} we need only T_{allo}. This is simply the average duration of allopatry long enough to complete speciation ($l \geq L_{\text{spec}}$):

$$T_{\text{allo}} = \mathbb{E}(L|l \geq L_{\text{spec}}) = \frac{\int_{L_{\text{spec}}}^{\infty} lf(l)dl}{\int_{L_{\text{spec}}}^{\infty} f(l)dl} = L_{\text{spec}} + \rho^{-1}. \quad (15)$$

This last result follows from the fact that the minimum waiting time is by assumption L_{spec}, and to that must be added ρ^{-1}, since the exponential is a memoryless distribution. In other words, the waiting time for the next event does not depend on how long one has already waited. Here, we are assuming that the next round of speciation cannot begin until secondary sympatry. If, however, speciation can proceed anew as soon as a bout of allopatry has exceeded L_{spec}, then $T_{\text{allo}} = L_{\text{spec}}$, decreasing T_{spec} by ρ^{-1}. This doubles the speciation rate when ρ is very small but becomes negligible as ρ increases. Putting together the time for initial sympatry following earlier speciation events (T_{sym}), unsuccessful bouts of allopatry ($T_{\text{allo}-\text{sym}}$), and the time in allopatry in which speciation is completed (T_{allo}) yields the total time for speciation:

$$T_{\text{spec}} = T_{\text{sym}} + T_{\text{allo}-\text{sym}} + T_{\text{allo}} \approx \frac{2e^{\rho L_{\text{spec}}}}{\rho}. \quad (16)$$

Equation (16) reveals that, despite a somewhat involved derivation, the waiting time to speciation is a relatively simple function of the time required to fix I_{spec} incompatibilities (L_{spec}) and the transition rate between allopatry and sympatry, ρ. All else being equal, greater L_{spec} means longer waiting times between speciation. In the next section, we show how L_{spec} depends on the amount of fuel for speciation, φ, and the rate of gene loss, d. However, we can already appreciate an important result from equation (16): most duplicate genes are lost in sympatry and do not contribute to speciation, even under the best-case scenario. From equation (16), we can see that the speciation rate (T_{spec}^{-1}) is hump shaped with respect to ρ, meaning that an intermediate value of ρ maximizes the rate of speciation. In fact, it is straightforward to show, by setting the first derivative to 0 and solving, that the maximum speciation rate occurs where $\rho_{\text{opt}} = L_{\text{spec}}^{-1}$. Substituting ρ_{opt} into equation (16) yields $T_{\text{spec}} = 2eL_{\text{spec}}$. In the best-case scenario, when the transition rate ρ is optimal for speciation, the realized waiting time to speciation is much greater ($2e$ is ≈5.5 times greater) than the actual time needed to fix I_{spec} hybrid incompatibilities. This means that under ideal circumstances, fewer than 20% of potential incompatibilities will actually contribute to speciation, and under most circumstances it will be far less than this. As we show in the next section, the significance of these insights will become more apparent when we complete the macroevolutionary model by iterating the accumulation of hybrid incompatibilities over many rounds of speciation.

Modeling Speciation Rates

To obtain the speciation rate of a clade following WGD, we need to know the waiting time to speciation (T_{spec}; eq. [16]) at any time point after WGD, after one or more rounds of speciation have taken place. For this, we extend the first model of hybrid incompatibility accumulation between two populations over multiple rounds of speciation. As the genome shrinks through gene loss, the rates of gene loss and accumulation of incompatibilities necessarily decrease. Consequently, after multiple rounds of speciation, the rate of accumulation of hybrid incompatibility due to reciprocal gene loss from genes duplicated at the WGD slows and eventually reaches 0. Letting s be the time since the most recent speciation event at time t, the genome size of a population j is

$$N_j(s|t) = N(1 + e^{-d(t+s)}). \quad (17)$$

Not only does the rate of gene loss slow through time, but two sister populations will initially have no divergently resolved duplicate genes, because they originate from the same parent population. After each round of speciation, the accumulation of hybrid incompatibilities must begin totally anew, even though many duplicate genes will already have been lost. Following the derivation of equation (4), the probability that one copy of gene i has been lost in population j at time s is

$$\begin{aligned} P(D_{i,j}, s|t) &= \frac{N_j(t) - N_j(t+s)}{N_j(t) - N} \\ &= \frac{N(1 + e^{-dt}) - N(1 + e^{-d(t+s)})}{N(1 + e^{-dt}) - N} \\ &= 1 - e^{-ds}. \end{aligned} \quad (18)$$

Following the derivation of equation (5), hybrid incompatibilities accumulate between populations j and k as

$$\begin{aligned} I(s|t) &= \int_{i=0}^{N(t)-N} \frac{p}{2} P(D_{i,j},s|t)P(D_{i,k},s|t)dN \\ &= \frac{p}{2}(N(t) - N)(1 - e^{-ds})^2 \\ &= \frac{Np}{2} e^{-dt}(1 - e^{-ds})^2. \end{aligned} \quad (19)$$

From equation (19), we can now derive L_{spec}, the time needed to accumulate I_{spec} incompatibilities and complete speciation, following an earlier round of speciation at time t by solving for s:

$$\begin{aligned} L_{spec}(t) &= -\frac{\log(1-\sqrt{2I_{spec}e^{dt}/Np})}{d} \\ &= -\frac{\log(1-\sqrt{e^{dt}\varphi^{-1}})}{d} \end{aligned} \tag{20}$$

Substituting equation (20) into equation (16), the waiting time to speciation as a function of t is

$$\begin{aligned} T_{spec}(t) &= \frac{2e^{\rho L_{spec}(t)}}{\rho} \\ &= \frac{2(1-\sqrt{e^{dt}\varphi^{-1}})^{-\rho d}}{\rho}. \end{aligned} \tag{21}$$

In a bifurcating tree, clade species richness expands through time depending on the speciation rate,

$$S_{t+1} = S_t e^{\log(2)T_{spec}^{-1}} = S_t 2^{T_{spec}^{-1}}. \tag{22}$$

Expressed in this way, the reciprocal of the waiting time to speciation (T_{spec}^{-1}) is the instantaneous rate at which clade diversity (S) doubles (i.e., the speciation rate in a strictly bifurcating tree). We denote the instantaneous speciation rate as a function of time as λ_t. Although we are modeling a phylogeny that is bifurcating in a discrete fashion, we use the instantaneous speciation rate to treat diversification as if it were a continuous process where

$$\frac{dS}{dt} = S_t \log(2)\lambda_t. \tag{23}$$

We were unable to obtain a general analytical solution to this model and instead numerically simulated it, using an ordinary differential equation solver implemented in the R packages deSolve (Soetaert et al. 2010) and rootSolve (Soetaert and Herman 2009). We ran simulations until the steady state, which occurs when L_{spec} approaches ∞ and all speciation fuel is spent:

$$t_{max} = \frac{\log(Np/2I_{spec})}{d}. \tag{24}$$

Simulations were performed over realistic values of d (0.01–1 Myr^{-1}) and φ (10–1,000). Lynch and Conery (2000, 2003) calculated that newly arisen duplicate genes have a half-life of ≈4 Myr, equivalent to an exponential decay rate of $d \approx 0.17$ Myr^{-1}. Gene loss rates after WGD might be slower than single gene duplications. For example, in yeast the half-life of duplicates after WGD is probably more than 10 Myr (Wolfe and Shields 1997; Scannell et al. 2006). Hence, 0.01–1 Myr^{-1} should encompass realistic values of d. We made several educated guesses to arrive at an approximate range of φ. Diploid genome sizes are on the order of 10^4. Presumably, most deficiencies resulting from hybridization after divergent resolution will be deleterious ($0.5 < p < 1$), although redundancy could mask some of these. The number of DMIs necessary to complete speciation, I_{spec}, is unknown, but using Lynch and Force's (2000) equation for hybrid fertility—$(15/16)^I$, where I is the number of DMIs—a 50% reduction in fertility requires ≈10 incompatibilities. Plugging these estimates into equation (7), we estimate that $\varphi = Np/2I_{spec} = (10^4 \times 0.75)/(2 \times 10) \approx 375$. This suggests that 10^2–10^3 is a reasonable range, but we included φ as low as 10, since there is interest in microorganisms (e.g., yeast) that have considerably smaller genomes. For each combination of d and φ, we simulated over a large number of values of ρ to ensure that we obtained the optimum ρ for a given area of parameter space. Realistic values of ρ, based on estimates of time to secondary sympatry in birds, are on the order of 0.1–1 Myr^{-1}, on average (Weir and Price 2011; Pigot and Tobias 2013). However, ρ must be considerably higher in plants and other taxa with higher speciation rates than birds, but we are not aware of additional taxon-specific estimates. Numerical simulations calculate speciation rates, given the expected value of parameters, and hence do not convey the variation arising from stochasticity in the rate of geographic transitions. For a subset of parameter space, we used stochastic simulations to estimate variation in speciation rates (app. A).

In all simulations, we kept track of the clade diversity through time and used this to calculate the average speciation rate, λ, from WGD ($t = 0$) until t_{max}, assuming no extinction. Thus, λ can be calculated for each simulation run from t_{max} and clade diversity at that time $S_{t_{max}}$:

$$\lambda = \frac{\log_2 S_{t_{max}}}{t_{max}}. \tag{25}$$

This is the standard definition of the speciation rate in a Yule pure-birth model. The $\log_2$ arises from the fact that we are assuming a purely bifurcating tree. Although we calculate both total species number and speciation rate, we focus on the latter because most empirical studies estimate speciation rates following WGD (Santini et al. 2009; Mayrose et al. 2011). Furthermore, for WGD that occurred long in the past, speciation rates and species number might become decoupled through subsequent variation in speciation and extinction rates. We classified calculated speciation rates higher than 0.1 "high" and those higher than 1 as "very high." This classification is based on empirical estimates showing that speciation rates higher than 0.2 Myr^{-1} are exceptional among mammals (Stadler 2011) and

birds (Jetz et al. 2012), whereas many angiosperm radiations have rates higher than 1 Myr^{-1} (Linder 2008; Valente et al. 2010, 2014; Bell et al. 2012; Drummond et al. 2012; Hoffmann et al. 2013).

Population Genetic Model of Reciprocal Gene Loss with Migration and Weak Selection against Extraneous Gene Copies

In the macroevolutionary model, we assumed that geographic state was discrete (allopatric or sympatric), with very different evolutionary behaviors in each. In allopatry, populations evolved independently; in sympatry, populations were completely homogenized by gene flow. In reality, populations can diverge even in the face of gene flow, through genetic drift or countervailing deterministic forces such as natural selection. Therefore, two arguments could be made against our assumption that complete allopatry is required to fix hybrid incompatibility under this model. First, divergent resolution of duplicate genes can occur with gene flow because of genetic drift. Second, direct selection against redundant gene copies can facilitate divergent resolution even with gene flow between incipient species.

To address these questions, we analyzed a two-population, two-locus population genetic model. The main result of this model is that divergent resolution is highly unlikely when there is gene flow, justifying our treatment of geography as a discrete state. Table 2 provides a glossary of symbols used in this model. We denote the two functional paralogues as A and B, the null alleles at those loci as a and b, and the frequency of the null alleles as x and y, respectively. We assume that a large "mainland" population has lost one paralogue of a duplicate pair but that a small "island" population still retains functional copies of both paralogues. By convention, we assume that b (the null allele) is fixed in the mainland population, such that $\bar{x} = 0$ and $\bar{y} = 1$. Allele frequencies in the mainland population do not change, since selection prevents loss of the second functional paralogue and migration from island to mainland is assumed to be inconsequential, as in Wright's continent-island model (Wright 1931). In the island population, both paralogues are initially fixed for functional copies, so $x_0 = y_0 = 0$. Null copies of either paralogue can fix in the island population by either mutation, migration, or selection against extra (superfluous) functional copies. These directional forces are denoted μ, m, and s, respectively. Loss of both functional copies is prevented by "epistatic" selection, denoted ε following the notation of Crow and Kimura (1970). Since ε is also the strength of reproductive isolation caused by the incompatibility, we restrict our analyses to the applicable case of $s \ll \varepsilon$. That is, any fitness cost of retaining genes in duplicate is much less than the cost of having no functional copies.

Table 2: Glossary of symbols in population genetic model

Symbol	Description
A, B, a, b	Functional (A, B) and null (a, b) alleles at a duplicated locus
x	Frequency of a
y	Frequency of b
μ	Mutation rate from functional to null allele; there is no back mutation in the model
m	Migration rate
s	Selection coefficient against redundant gene copies
ε	Epistatic selection coefficient; this is also the strength of reproductive isolation caused by an incompatibility

We assessed the possibility of divergent resolution in the island population under the following four scenarios:

Scenario 1. *No migration, no selection against extra copies.*

Scenario 2. *Migration, no selection against extra copies.*

Scenario 3. *No migration, selection against extra copies.*

Scenario 4. *Migration and selection against extra copies.*

We derive the equilibria and basins of attraction for the deterministic model analytically and numerically, assuming infinite population size. Equilibria were determined by finding the intersection between isoclines that fell within the feasible parameter space (allele frequencies between 0 and 1). We used Routh-Hurwitz criteria to analyze the stability of equilibria. Basins of attraction were determined under representative parameter space by numerically simulating models with many different starting conditions, using deSolve. This code is available upon request from the authors.

Scenario 1: No Migration, No Selection against Extra Copies. The fitnesses of all genotypes analyzed in scenarios without selection against extra copies are given in table 3*A*. In a standard, discrete-time population genetic model, the changes in x and y over one generation due to epistatic selection are

$$\Delta x = -(1 - x)\left(\frac{\varepsilon x^2 y^2}{1 - \varepsilon x^2 y^2}\right), \tag{26a}$$

$$\Delta y = -(1 - y)\left(\frac{\varepsilon x^2 y^2}{1 - \varepsilon x^2 y^2}\right). \tag{26b}$$

Here, we use a continuous time model and approximate

Table 3: Frequency and fitness in population genetic model

Genotype	Frequency	*AA*	*Aa*	*aa*
A. Without selection against extra gene copies:				
Frequency		$(1 - x)^2$	$2x(1 - x)$	x^2
BB	$(1 - y)^2$	1	1	1
Bb	$2y(1 - y)$	1	1	1
bb	y^2	1	1	$1 - \varepsilon$
B. With selection against extra gene copies:				
Frequency		$(1 - x)^2$	$2x(1 - x)$	x^2
BB	$(1 - y)^2$	$1 - 4s$	$1 - 3s$	$1 - 2s$
Bb	$2y(1 - y)$	$1 - 3s$	$1 - 2s$	$1 - s$
bb	y^2	$1 - 2s$	$1 - s$	$1 - \varepsilon$

the change in allele frequencies due to selection, following Christiansen and Frydenberg (1977), by

$$\frac{\varepsilon x^2 y^2}{1 - \varepsilon x^2 y^2} \approx \varepsilon x^2 y^2. \tag{27}$$

This approximation is valid so long as the frequency of the double homozygote *aabb* is low ($x \ll 1$ and/or $y \ll 1$), which should be the case, since selection prevents the incompatibility from reaching high frequency in a population. With that approximation, we can derive a system of ordinary differential equations:

$$\frac{dx}{dt} = -x(1 - x)\varepsilon x y^2 + \mu(1 - x), \tag{28a}$$

$$\frac{dy}{dt} = -y(1 - y)\varepsilon x^2 y + \mu(1 - y). \tag{28b}$$

We include mutation in this first scenario only; otherwise, there is no directional force causing one of the duplicate genes to be lost. We do not include it in the other scenarios because other directional forces (selection and/or migration) are assumed to overwhelm the effects of mutation. Since the per-locus mutation rate is on the order of 10^{-6}, migration would have to be on the same order for mutation to matter. Such low migration rates are effectively allopatric. Furthermore, unlike migration, the net effect of mutation at both loci does not bias the outcome toward divergent or parallel resolution. For analytical tractability, we also assumed free recombination between paralogues ($r = 1$), which is greater than the maximum possible recombination rate ($r = 0.5$). Since the paralogues we are considering arose by WGD, they should be on different chromosomes and therefore not physically linked. Numerical simulations using a realistic recombination rate ($r = 0.5$) were nearly identical to those ignoring recombination (app. D).

Scenario 2: Migration, No Selection against Extra Copies. The change in allele frequencies due to migration is

$$\begin{aligned} \Delta x &= m(\bar{x} - x), \\ \Delta y &= m(\bar{y} - y), \end{aligned} \tag{29}$$

where $\bar{x}$ and $\bar{y}$ are the allele frequencies in the mainland population. By assumption, $\bar{x} = 0$ and $\bar{y} = 1$. Adding migration to the previous system of equations (eq. [28]) and taking out mutation yields the new system of equations for this scenario:

$$\frac{dx}{dt} = -x(1 - x)\varepsilon x y^2 - mx, \tag{30a}$$

$$\frac{dy}{dt} = -y(1 - y)\varepsilon x^2 y + m(1 - y). \tag{30b}$$

We do not consider more-complex models incorporating multiple segregating incompatibilities that would further reduce immigrant fitness, causing the effective migration rate at a given locus to decrease as the number of incompatibilities increased.

Scenario 3: No Migration, Selection against Extra Copies. The fitnesses of all genotypes analyzed in scenarios with selection against extra copies are given in table 3*B*. From this, we derived a new system of equations:

$$\frac{dx}{dt} = x(1 - x)(s - \varepsilon x y^2), \tag{31a}$$

$$\frac{dy}{dt} = y(1 - y)(s - \varepsilon x^2 y). \tag{31b}$$

This system of equations does not exactly follow from the fitness values given in table 3*B* because the effect of s on x has some dependency on y, and vice versa. In the earlier equations, s does not have this property. However, this simpler system of equations captures the key feature of the model, which is that there are two equal fitness peaks corresponding to parallel and divergent resolution. Consequently, unlike equations (28), where null copies are always selected against

because of epistatic selection, in this scenario x and y can increase in frequency if $s > \varepsilon xy^2$ or $s > \varepsilon x^2 y$.

Scenario 4: Migration and Selection against Extra Copies. Finally, the fourth scenario combines migration and selection as derived in equations (28) and (30):

$$\frac{dx}{dt} = x(1 - x)(s - \varepsilon xy^2) - mx, \tag{32a}$$

$$\frac{dy}{dt} = y(1 - y)(s - \varepsilon x^2 y) + m(1 - y). \tag{32b}$$

This is the most important model, since it addresses whether selection can counteract the homogenizing force of gene flow and favor divergent resolution in parapatry.

Results

Accumulation of Hybrid Incompatibilities following WGD

Following WGD, redundant gene copies are lost because there is no selection against nonfunctional alleles. We have assumed that gene number declines from $2N$ to N (fig. 2), but in nature, many WGD-derived duplicates are retained indefinitely (Lynch 2007; Otto 2007). Selection to retain duplicates reduces the number available for incompatibilities but does not qualitatively change the model results; it simply means that there is less "fuel" for speciation. Conversely, if we include the possibility of the subfunctionalization of genes without complete loss, we need to add the number of possible functions to $2N$. As duplicate genes are lost, divergent resolution occurs between isolated populations (incipient species), leading to hybrid incompatibilities (fig. 2). The key result here is that the rate of gene loss (or gene function loss), d, dramatically affects the rate at which hybrid incompatibilities accumulate, even though the total number of potential incompatibilities is set by $Np/2$ (fig. 2). Hence, lower rates of duplicate-gene loss lead to slower accumulation of hybrid incompatibilities and a longer waiting time to speciation but to a greater duration of increased post-WGD speciation rates for the clade as a whole.

Macroevolutionary Model

Our macroevolutionary model demonstrates that most potential hybrid incompatibilities following WGD do not actually contribute to speciation. Rather, most duplicate copies are lost within interbreeding populations and hence cannot be divergently resolved. The realized waiting time to speciation after WGD is thus much longer than the time required to simply fix enough potential incompatibilities to complete speciation. We have identified three key parameters that affect the speciation rate in this model: the speciation "fuel" (φ), which is proportional to the number of remaining duplicates, the rate of gene loss (d), and the transition rate between allopatry and sympatry (ρ). The parameters φ and d affect speciation rates in a similar fashion that is qualitatively different from that for ρ, and therefore we present results from the first two terms together. Saved simulation outputs on which these analyses are based are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.8k0m3 (Muir and Hahn 2014).

All else being equal, more fuel and faster rates of gene loss lead to greater maximum speciation rates (λ_{max}; fig. 3). It is critical to emphasize that λ_{max} is the maximum rate possible for a given φ and d but that the rate is generally much lower, depending on ρ (see below). The reason that d and φ increase maximum possible speciation rates is simple: both decrease the waiting time between incompatibilities (eq. [21]). The parameter φ sets the maximum possible number of species during a radiation, while d affects speciation rates by setting the duration, t_{max} (see eq. [24]). The parameter d has a proportionally greater effect than φ (fig. 3) because, once populations are allopatric, there is a lag (reflected in the squared term of eq. [5]) before enough genes are lost in each population to ensure that paralogues of the same gene have been lost. Faster d decreases both the total duration over which λ is calculated and the lag time before substantial numbers of incompatibilities accumulate.

However, ρ must be properly tuned to a given d and φ in order to take advantage of incompatibilities. This can most intuitively be understood by considering the fact that a population that was always sympatric would never accumulate incompatibilities, whereas one that was always allopatric would accumulate only a single speciation event. The optimal ρ that maximizes speciation rates in the macroevolutionary model is conceptually similar to ρ_{opt} described above for equation (16), but it is somewhat different, since L_{spec} changes through time (eq. [20]). As the rate of gene loss increases, the optimal ρ likewise becomes higher (fig. 4). Similarly, at a given d, increasing φ increases the rate at which incompatibilities evolve and the optimal ρ for speciation. When potential hybrid incompatibilities are evolving rapidly, L_{spec} is low, and the rate-limiting step in speciation is the initial period of allopatry (T_{allo}). Consequently, faster transition rates allow more rapid speciation as the rate of hybrid incompatibility formation increases.

High Speciation Rates Are Possible Only in a Limited Area of Parameter Space

High speciation rates are possible in our model, but only in a restricted area of parameter space. For example, when d is sufficiently low, speciation rates will always be low

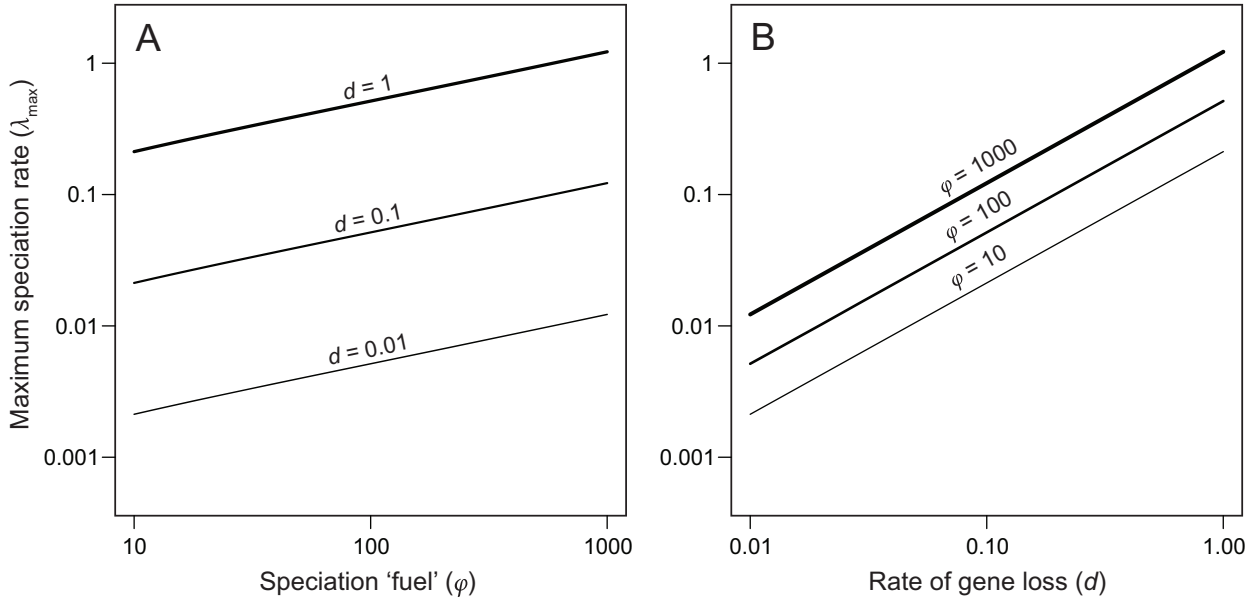

Figure 3: Gene loss and speciation "fuel" determinants of the maximum speciation rate (λ_{max}). Faster rates of duplicate-gene loss (d; *A*) and more speciation "fuel" (φ, *B*) lead to higher λ_{max}.

(fig. 4*A*). In our model, which assumes that incompatibilities arise only from gene loss, when $d = 0.01$ genes Myr^{-1}, the speciation rate was rarely over a slow 0.01 Myr^{-1}. This rate of gene loss is slow, and it would imply a waiting time of 69.3 Myr after WGD for half the gene duplicates to be lost. Even when $d = 0.1$ genes Myr^{-1}, in which half of the duplicate genes would be lost in 6.93 Myr, "high" speciation rates ($\lambda > 0.1$) are possible only if φ and ρ are also very high (fig. 4*B*). "Very high" speciation ($\lambda > 1$) is possible only when φ is high and gene loss is very rapid (fig. 4*C*). However, when gene loss is so rapid, a long, sustained increase in speciation rates is not possible, since all the fuel will be used rapidly. These results are taken from the expected values in our model, but stochastic simulations indicate that while substantial variance in speciation rates exists for any given parameter combination, they are generally within the same order of magnitude as the average (fig. A1; figs. A1, F1–F6 available online).

Divergent Resolution with Gene Flow Is Improbable

The results from analysis of Scenarios 1–3 in the population genetic model are straightforward. Without migration, neither drift nor selection for reduced copy number affects the probability of divergent versus parallel resolution (provided that paralogues are indistinguishable). In contrast, any migration between populations greatly reduces the probability of divergent resolution. Detailed analyses of Scenarios 1–3 are given in appendix E. Here, we focus on Scenario 4, where we combined migration between populations with selection against extra copies. This scenario addresses whether selection toward one of the fitness peaks can deterministically overcome migration and permit divergent resolution when there is gene flow. One might reasonably expect that sufficiently strong selection for loss of either paralogue would overcome migration, but instead we find that selection reinforces the homogenizing force of migration, because the system begins in the basin of attraction for the parallel resolution equilibrium.

In this system, there are trivial isoclines where $x = 1$ and $y = 1$, but there are also four other isoclines. Internal, stable equilibria, corresponding approximately to parallel ($\hat{x} \approx 0, \hat{y} \approx 1$) and divergent ($\hat{x} \approx 1, \hat{y} \approx 0$) resolution of duplicate genes, exist where two of the isoclines intersect each other:

$$\frac{dx}{dt} = 0$$

when

$$y = \sqrt{\frac{(m - s + sx)}{(-\varepsilon x + \varepsilon x^2)}},$$

and

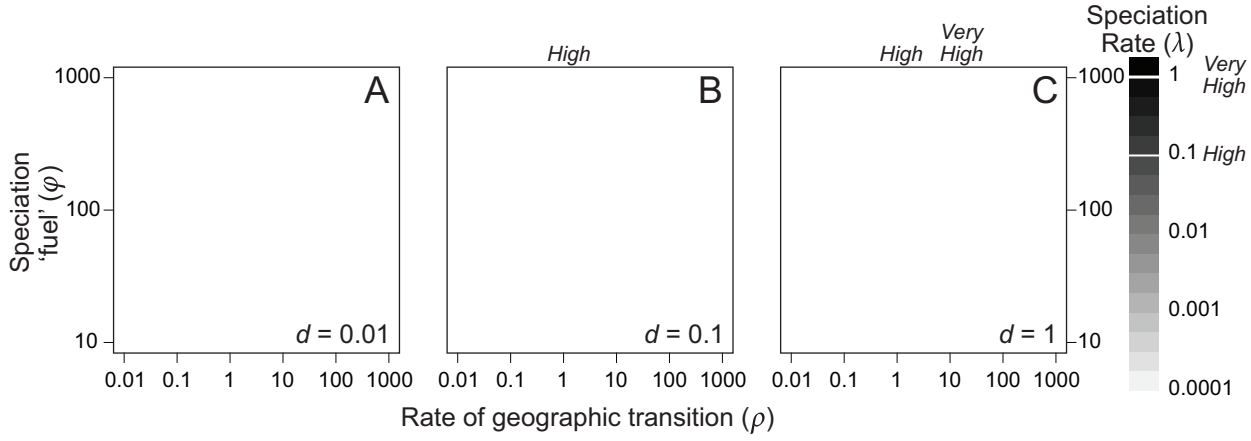

Figure 4: Elevated speciation rates occur only in some areas of parameter space. In all panels, the *X*-axis is ρ, the transition rate between allopatry and sympatry, on a logarithmic scale. Gray shading between contours indicates the modeled speciation rate λ, with darker shades corresponding to higher rates (see log-scale legend to the right). Contours demarcating "high" ($0.1 < b < 1$) and "very high" ($\lambda > 1$) speciation rates are in white. More speciation "fuel" (φ, *Y*-axis) and greater rates of gene loss (d, increasing from 0.01 to 1 from *A* to *C*) lead to higher potential speciation rates. However, there is always an optimal ρ that maximizes the speciation rate λ, depending on the duplicate-gene loss rate (d) and the speciation "fuel" (φ).

$$\frac{dy}{dt} = 0$$

when

$$y = \frac{s + \sqrt{s^2 + 4\varepsilon m x^2}}{2\varepsilon x},$$

but only when gene flow is sufficiently weak relative to selection (approximately $m < s$; see app. E for detailed analysis). We could not solve for these equilibria analytically but did so numerically under relevant parameter space (app. E). Figure 5 illustrates these equilibria for representative parameters. When gene flow is sufficiently strong (approximately $m > s$), the internal equilibria collapse and the system always proceeds to the parallel resolution equilibrium:

$$\hat{x} = \frac{s}{m - \varepsilon},$$

$$\hat{y} = 1.$$

At this equilibrium, y is fixed and x is maintained at low frequency when $s \ll \varepsilon$. Even if there is a feasible divergent resolution equilibrium when $m \ll s$, it will be nearly impossible to reach because the island population starts near the origin ($x_0 \approx 0$, $y_0 \approx 0$), in the basin of attraction of the parallel resolution equilibrium (fig. 5). Only when $4N_e s < 1$ will drift be able to overcome selection and allow it to reach the divergent resolution equilibrium. Perhaps counterintuitively, selection against extra copies either does not change the probability of divergent resolution (Scenario 2) or makes it more difficult (Scenario 4), because selection and migration work in the same direction over much of the allele frequency space.

Discussion

Speciation rates vary enormously across clades, but the cause of this variation is not well understood. Whole-genome duplication (WGD) has been posited to increase diversification rates (Otto and Whitton 2000; Zhou et al. 2001; Donoghue and Purnell 2005; Scannell et al. 2006; Sémon and Wolfe 2007; Soltis et al. 2009; Van de Peer et al. 2009), but recent evidence suggests that speciation is actually slower in neopolyploid than in diploid plants (Mayrose et al. 2011) and that any effect of WGD on diversification takes millions of years to appear (Santini et al. 2009; Schranz et al. 2012). We consider these data further below. One mechanism by which WGD has been hypothesized to increase diversification rates is by increasing the rate at which hybrid incompatibilities and, hence, reproductive isolation evolve. However, we have shown that there is no simple relationship between the rate at which incompatibilities evolve and speciation rate. Population processes, specifically the rate at which a population splits into multiple allopatric populations, have a profound effect on the realized speciation rate. In light of our results, below we evaluate the hypothesis that WGD increases diversification rates by increasing the rate at which hybrid incompatibilities accumulate via divergent resolution of duplicate genes.

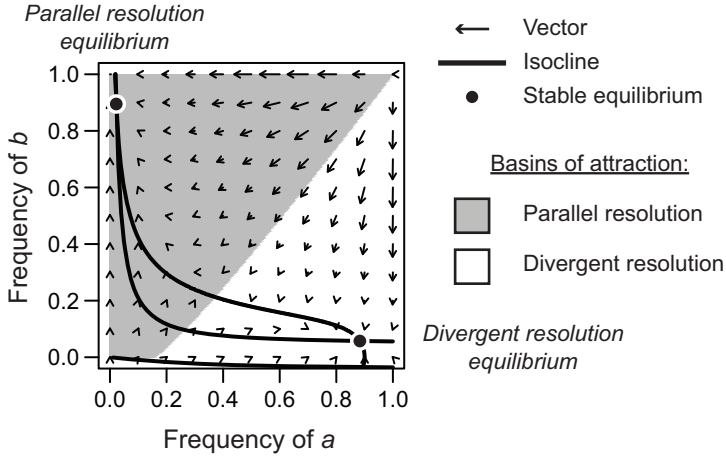

Figure 5: Divergent resolution with gene flow and selection against superfluous gene copies is unlikely: vector field, nontrivial isoclines, relevant equilibria, and basins of attractions in the population genetic model with migration and selection against superfluous paralogues (Scenario 4; see eq. [32]). Vector field: the direction and length of arrows correspond, respectively, to the direction and magnitude of allele frequency change in the system at that point. Isoclines: thick, black lines indicate nontrivial isoclines ($dx/dt = 0$ or $dy/dt = 0$). Equilibria: where isoclines intersect are the stable equilibria (filled circles) of the system. The stable equilibria correspond approximately to either parallel (top left vertex) or divergent (bottom right vertex) resolution of duplicate genes. Basins of attraction: the basins of attraction for parallel and divergent resolution are in gray and white, respectively. A population that starts in the basin of attraction of one equilibrium will deterministically reach that equilibrium and not another. Parameters: $s = 0.01$, $\varepsilon = 0.5$, $m = 0.001$. The value of s is probably unrealistically large but was chosen for visual clarity.

Divergent Resolution Predicts an Erratic Relationship between WGD and Speciation Rates

Some polyploid clades are more diverse than closely related diploid clades, suggesting higher speciation rates or lower extinction rates (Crow and Wagner 2006). However, our model predicts that if divergent resolution of duplicate genes is the primary cause of speciation, then the relationship between WGD and speciation rates across clades and through time should be highly erratic. Although the number of potential incompatibilities and the rate of gene loss set the maximum possible speciation rate (fig. 3), the realized speciation rate depends heavily on ρ, the rate at which species transition from a single interbreeding population to at least two non-interbreeding populations (fig. 4). For any given combination of φ and d, very high ($\lambda > 1$ Myr^{-1}) and very low ($\lambda < 0.1$ Myr^{-1}) speciation rates are possible. There is no a priori reason to expect that factors such as genome size, the rate of duplicate-gene loss, and the rate at which populations become allopatric should be closely coupled across extremely different groups of organisms (e.g., ferns and teleost fishes) or over long epochs of time. Divergent resolution of duplicate genes following WGD might increase speciation rates in some lineages at some times, but it is difficult to see how this mechanism could lead to a consistent pattern of prolonged speciation following WGD.

Population Processes, Not the Rate of Divergent Resolution, Often Determine Speciation Rates

Our model indicates that population processes, rather than the genome-dependent rate at which potential hybrid incompatibilities evolve, frequently determine the speciation rate. In our model, for populations to become reproductively isolated, several duplicate genes must be divergently resolved and lead to hybrid incompatibilities. However, if a duplicate gene is resolved within a population or between populations sharing migrants, our population genetic model shows that parallel, rather than divergent, resolution is the dominant outcome, except in very small populations or with very low gene flow. Thus, all else being equal, the more quickly populations split and become allopatric, the sooner divergent resolution can begin. However, rapid transitions between allopatry and sympatry also hamper speciation. If populations come back together before speciation is complete, accumulated incompatibilities will be purged by selection. Hence, slower transition rates that allow more time in allopatry should (up to a point) allow speciation to proceed more rapidly. Consequently, an intermediate rate of transitioning between allopatry and sympatry, tuned to the rate at which incompatibilities are evolving, maximizes the speciation rate. More-rapid speciation would be possible if transition rates were asymmetric (Allo $\rightarrow$ Sym $\neq$ Sym $\rightarrow$ Allo), but asymmetric rates could just as likely slow speciation, depending on the par-

ticular values of φ and d. In the absence of a particular spatial model to infer how transition rates may differ, for analytical clarity we have assumed symmetric rates. This does not change the primary conclusion that rapid divergent resolution must be coupled with specific population processes in order to achieve rapid realized speciation.

Even Small Amounts of Gene Flow Thwart Divergent Resolution

The primary purpose of the population genetic model was to address whether it was reasonable to assume in the macroevolutionary model that geographic state was a discrete character, being either completely sympatric or allopatric. As with any neutral mutation, pseudogenes spread easily across populations through gene flow because there are no countervailing deterministic forces to stop them. This means that parallel, rather than divergent, resolution of duplicate genes will dominate whenever there is appreciable gene flow. Practically, this means that sympatric and parapatric populations behave similarly in terms of our model, and it justifies our simplifying assumption. It also suggests that divergent resolution should be extremely difficult unless geographic barriers to gene flow are strong or reproductive isolation is already complete. In the latter case, divergent resolution following WGD might cause hybrid incompatibility but would not actually affect the speciation rate. This is a problem for all nonselective models of speciation. Consequently, the likelihood that divergent resolution plays a major role in speciation depends on how often there is gene flow during speciation, in which cases selection must be acting.

Reevaluating the Relationship between WGD and Speciation Rates

To reiterate, our models have identified two critical problems with the hypothesis that an increased opportunity for divergent resolution leads to rapid post-WGD diversification: (1) High speciation rates occur only in a limited area of parameter space. (2) Even small amounts of gene flow thwart divergent resolution, suggesting that speciation might already be complete before divergent resolution can contribute to speciation.

Given these challenges, we must be skeptical that divergent resolution following WGD by itself leads to increased speciation rates. Nevertheless, our model also indicates that high speciation rates and large clades (see supplementary data set of numerical simulations for exact clade size) after WGD are possible under some parameter space. For example, in a moderately sized eukaryotic genome (20,000 genes) where 10 incompatibilities are required to complete speciation, $\varphi = 1{,}000$. With a high rate of gene loss ($d = 1$) and a transition rate $\rho \approx 10$–$100\ \text{Myr}^{-1}$ (100,000–10,000 yr to become completely allopatric), very high speciation rates ($\lambda > 1\ \text{Myr}^{-1}$) are possible (fig. 4*C*) and could lead to reasonably high clade diversity (300–400 species) if extinction rates are low.

It is also possible that WGD does not actually increase diversification rates or that there is a substantial lag time between WGD and increased diversification. Indeed, a recent analysis in plants found that polyploids did not diversify faster soon after WGD than did their diploid relatives (Mayrose et al. 2011). The apparent success of polyploid lineages, as indicated by the fact that many organisms have a history of WGD, could result from the asymmetry between genome doubling, which occurs frequently (Ramsey and Schemske 1998), and the reduction in genome size back to the ancestral state, which is a slower process. After many millions of years, most lineages can have some history of WGD even if it confers no macroevolutionary advantage (Meyers and Levin 2006). Alternatively, some have postulated that polyploids are more successful in rare but seminal historical epochs (Van de Peer et al. 2009; Arrigo and Barker 2012), such as mass-extinction events (Fawcett et al. 2009). The success of polyploids after major extinctions would plausibly be consistent with the reciprocal gene loss mechanism only if such events affected the transition rate to allopatry in some way that made hybrid incompatibilities accumulate faster. For example, Schranz et al. (2012) showed that there is usually a substantial time lag between WGD and increased diversification in several major plant clades, with greater diversification usually being associated with a major biogeographic shift or spread. Perhaps the novel ecological opportunity afforded by mass extinction and/or colonization of new land areas shifts the transition rate from allopatry to sympatry in such a way that allows reciprocal gene loss to contribute more to speciation.

It is also possible that other consequences of WGD, such as the greater opportunity for the adaptive evolution of new gene duplicates ("neofunctionalization"), contribute to increased rates of speciation. However, all such hypotheses implicitly assume that the pre-WGD speciation rate is mutation limited (i.e., speciation would be faster if hybrid incompatibilities evolved faster). Instead, our model shows that the speciation rate is most often limited by population processes rather than by the available mutations. While we have demonstrated only the importance of population processes in driving speciation relative to passive gene loss following WGD, we believe that it may generally be the limiting factor, consistent with recent data demonstrating a lack of correlation between speciation rate and the rate at which hybrid incompatibilities evolve in *Drosophila* and birds (Rabosky and Matute 2013).

In conclusion, our model shows that reciprocal gene

loss can plausibly increase speciation rates after WGD, bu[t] only under special conditions. Some radiations might therefore have been spurred by reciprocal gene loss. However, the emerging empirical pattern is that WGD doe[s] not immediately elevate speciation rates, perhaps in part because population processes and/or gene flow thwart the evolution of hybrid incompatibilities via reciprocal gen[e] loss.

Acknowledgments

Members of the Moyle, Otto, and Whitlock labs provided feedback. The research was supported by a National Sci[-] ence Foundation (NSF) Graduate Research Fellowship to C.D.M. and NSF grant DBI-0845494 to M.W.H.

Literature Cited

Arrigo, N., and M. S. Barker. 2012. Rarely successful polyploids an[d] their legacy in plant genomes. Current Opinion in Plant Biology 15:140–146.

Aury, J.-M., O. Jaillon, L. Duret, B. Noel, C. Jubin, B. M. Porcel, B. Ségurens, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171–178.

Bedinger, P. A., R. T. Chetelat, B. McClure, L. C. Moyle, J. K. Rose, S. M. Stack, E. van der Knaap, et al. 2011. Interspecific reproductive barriers in the tomato clade: opportunities to decipher mechanisms of reproductive isolation. Sexual Plant Reproduction 24:171–187.

Bell, C. D., E. Mavrodiev, P. Soltis, A. Calaminus, D. Albach, N. Cellinese, N. Garcia-Jacas, and D. Soltis. 2012. Rapid diversification of *Tragopogon* and ecological associates in Eurasia. Journal of Evolutionary Biology 25:2470–2480.

Bikard, D., D. Patel, C. Le Metté, V. Giorgi, C. Camilleri, M. J. Bennett, and O. Loudet. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science 323:623–626.

Christiansen, F. B., and O. Frydenberg. 1977. Selection-mutation balance for two nonallelic recessives producing an inferior double homozygote. American Journal of Human Genetics 29:195–207.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Crow, K. D., and G. P. Wagner. 2006. What is the role of genome duplication in the evolution of complexity and diversity? Molecular Biology and Evolution 23:887–892.

Darwin, C. R. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. J. Murray, London.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York.

Donoghue, P. C., and M. A. Purnell. 2005. Genome duplication, extinction and vertebrate evolution. Trends in Ecology and Evolution 20:312–319.

Drummond, C. S., R. J. Eastwood, S. T. Miotto, and C. E. Hughes. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. Systematic Biology 61:443–460.

Edger, P. P., and J. C. Pires. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Research 17:699–717.

Fawcett, J. A., S. Maere, and Y. Van de Peer. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proceedings of the National Academy of Sciences of the USA 106:5737–5742.

Fawcett, J. A., Y. Van de Peer, and S. Maere. 2013. Significance and biological consequences of polyploidization in land plant evolution. Pages 277–293 *in* I. J. Leitch, J. Greilhuber, and J. Dolezel, eds. Plant genome diversity. Volume 2. Springer, Vienna.

Hoffmann, M. H., J. Schneider, P. Hase, and M. Röser. 2013. Rapid and recent world-wide diversification of bluegrasses (*Poa*, Poaceae) and related genera. PLoS ONE 8:e60061.

Jablonski, D. 2008. Species selection: theory and data. Annual Review of Ecology, Evolution, and Systematics 39:501–524.

Jetz, W., G. Thomas, J. Joy, K. Hartmann, and A. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Linder, H. P. 2008. Plant species radiations: where, when, why? Philosophical Transactions of the Royal Society B: Biological Sciences 363:3097–3105.

Lynch, M. 2007. The origins of genome architecture. Sinauer, Sunderland, MA.

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

———. 2003. The evolutionary demography of duplicate genes. Journal of Structural and Functional Genomics 3:35–44.

Lynch, M., and A. G. Force. 2000. The origin of interspecific genomic incompatibility via gene duplication. American Naturalist 156:590–605.

Maclean, C. J., and D. Greig. 2011. Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. Evolution 65:932–945.

Macnair, M., and P. Christie. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*? Heredity 50:295–302.

Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer. 2005. Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Sciences of the USA 102:5454–5459.

Masly, J. P., C. D. Jones, M. A. Noor, J. Locke, and H. A. Orr. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. Science 313:1448–1450.

Mayr, E. 1942. Systematics and the origin of species: from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:1257.

McGrath, C. L., and M. Lynch. 2012. Evolutionary significance of whole-genome duplication. Pages 1–20 *in* P. S. Soltis and D. E. Soltis, eds. Polyploidy and genome evolution. Springer, Berlin.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. Evolution 60:1198–1206.

Muir, C. D., and M. W. Hahn. 2014. Data from: The limited contribution of reciprocal gene loss to increased speciation rates fol-

lowing whole-genome duplication. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.8k0m3.

Muller, H. 1942. Isolating mechanisms, evolution and temperature. Biological Symposia 6:71–125.

Orr, H. A., J. Masly, and N. Phadnis. 2007. Speciation in *Drosophila*: from phenotypes to molecules. Journal of Heredity 98:103–110.

Otto, S. P. 2007. The evolutionary consequences of polyploidy. Cel 131:452–462.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34:401–437.

Phadnis, N., and H. A. Orr. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323:376–379.

Pigot, A. L., and J. A. Tobias. 2013. Species interactions constrain geographic range expansion over evolutionary time. Ecology Letters 16:330–338.

Presgraves, D. C. 2010. The molecular evolutionary basis of species formation. Nature Reviews Genetics 11:175–180.

Rabosky, D. L. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecology Letters 12:735–743.

Rabosky, D. L., and D. R. Matute. 2013. Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. Proceedings of the National Academy of Sciences of the USA 110:15354–15359.

Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29:467–501.

Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. Science 317 910–914.

Santini, F., L. J. Harmon, G. Carnevale, and M. E. Alfaro. 2009. Did genome duplication drive the origin of teleosts? a comparativ study of diversification in ray-finned fishes. BMC Evolutionary Biology 9:194.

Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gen loss in polyploid yeasts. Nature 440:341–345.

Schartl, M. 2008. Evolution of *Xmrk*: an oncogene, but also a speciation gene? BioEssays 30:822–832.

Schranz, E. M., S. Mohammadin, and P. P. Edger. 2012. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. Current Opinion in Plant Biology 15 147–153.

Sémon, M., and K. H. Wolfe. 2007. Reciprocal gene loss betwee *Tetraodon* and zebrafish after whole genome duplication in their ancestor. Trends in Genetics 23:108–112.

Soetaert, K., and P. M. Herman. 2009. A practical guide to ecological modelling: using R as a simulation platform. Springer, Dordrecht.

Soetaert, K., T. Petzoldt, and R. W. Setzer. 2010. Solving differential equations in R: Package deSolve. Journal of Statistical Software 33: 1–25.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, P. K. Wall, and P. S. Soltis. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96: 336–348.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences of the USA 108:6187–6192.

Stevens, P. F. 2001–. Angiosperm phylogeny website, version 12. http://www.mobot.org/MOBOT/research/APweb/, accessed August 29, 2013.

Taylor, J. S., Y. Van de Peer, and A. Meyer. 2001. Genome duplication, divergent resolution and speciation. Trends in Genetics 17:299–301.

Valente, L. M., A. W. Britton, M. P. Powell, A. S. Papadopulos, P. M. Burgoyne, and V. Savolainen. 2014. Correlates of hyperdiversity in southern African ice plants (Aizoaceae). Botanical Journal of the Linnean Society 174:110–129.

Valente, L. M., V. Savolainen, and P. Vargas. 2010. Unparalleled rates of species diversification in Europe. Proceedings of the Royal Society B: Biological Sciences 277:1489–1496.

Van de Peer, Y., S. Maere, and A. Meyer. 2009. The evolutionary significance of ancient genome duplications. Nature Reviews Genetics 10:725–732.

Vanneste, K., Y. Van de Peer, and S. Maere. 2013. Inference of genome duplications from age distributions revisited. Molecular Biology and Evolution 30:177–190.

Weir, J. T., and T. D. Price. 2011. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. American Naturalist 177:462–469.

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

Wolfe, K. H., and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–712.

Wright, K. M., D. Lloyd, D. B. Lowry, M. R. Macnair, and J. H. Willis. 2013. Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. PLoS Biology 11:e1001497.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

Zhou, R., H. Cheng, and T. R. Tiersch. 2001. Differential genome duplication and fish diversity. Reviews in Fish Biology and Fisheries 11:331–337.

Associate Editor: Daniel I. Bolnick
Editor: Susan Kalisz