Peak and Persistent Excess of Genetic Diversity Following an Abrupt Migration Increase

Nicolas Alcala, Daniela Streit, Jérôme Goudet, and Séverine Vuilleumier[1]
Department of Ecology and Evolution, Biophore, University of Lausanne, CH-1015 Lausanne, Switzerland

ABSTRACT Genetic diversity is essential for population survival and adaptation to changing environments. Demographic processes (*e.g.*, bottleneck and expansion) and spatial structure (*e.g.*, migration, number, and size of populations) are known to shape the patterns of the genetic diversity of populations. However, the impact of temporal changes in migration on genetic diversity has seldom been considered, although such events might be the norm. Indeed, during the millions of years of a species' lifetime, repeated isolation and reconnection of populations occur. Geological and climatic events alternately isolate and reconnect habitats. We analytically document the dynamics of genetic diversity after an abrupt change in migration given the mutation rate and the number and sizes of the populations. We demonstrate that during transient dynamics, genetic diversity can reach unexpectedly high values that can be maintained over thousands of generations. We discuss the consequences of such processes for the evolution of species based on standing genetic variation and how they can affect the reconstruction of a population's demographic and evolutionary history from genetic data. Our results also provide guidelines for the use of genetic data for the conservation of natural populations.

GENETIC diversity in a population of constant size results from the balance between the occurrence of new mutations and the loss of alleles by genetic drift (Fisher 1922; Wright 1931; Kimura and Crow 1964). The expected population genetic diversity can thus be estimated from the effective population size and the mutation rate in the population. In subdivided populations this estimate should further account for the strength of migration (Maruyama 1970; Smith 1970; Nei 1973): limited migration allows for strong differentiation between populations, while strong migration tends to homogenize genetic diversity between populations. Genetic diversity is also known to be affected by population demographic changes; following bottlenecks and founder events, a loss of genetic diversity is expected to occur (Nei *et al.* 1975). Recently, spatial population expansions were shown to lead to increased differentiation between populations and to generate a low level of genetic diversity at the front of the expansion (Excoffier *et al.* 2009).

Although theoretical studies on the dynamics of genetic diversity in subdivided populations started appearing in the 1970s (Nei and Feldman 1972; Latter 1973; Nei 1973; Nagylaki 1974, 1977), the transient dynamics and nonequilibrium states of genetic diversity still do not have a good theoretical basis. Early authors characterized the ultimate rate of change of genetic diversity after a perturbation (either a change in population size or gene flow; Nei and Feldman 1972; Latter 1973; Nei 1973; Nagylaki 1974, 1977). They found that changes in genetic diversity are related to the total effective population size, which results in a slow dynamics of genetic diversity change. They thus first highlighted that nonequilibrium states and transient dynamics are expected to act on very large temporal scales. In particular, they showed that decreases in migration rates (population fragmentation or isolation) have long-term effects on genetic diversity: they reduce the amount of genetic diversity within populations and allow for population differentiation (Latter 1973; Takahata and Nei 1985). Additionally, it has been shown that short timescale random fluctuations in migration increase population differentiation (Nagylaki 1979; Whitlock 1992; Rice and Papadopoulos 2009) while cyclic fluctuations of gene flow (such as seasonal fluctuations) mainly affect genetic diversity within populations (Karlin 1982; Shpak *et al.* 2010). Although the genetic consequences of migration events (admixture) have recently received much attention (*e.g.*, Pritchard *et al.* 2000; Falush *et al.* 2003; Price *et al.* 2009; Gravel 2012), their impact on

doi: 10.1534/genetics.112.147785
Manuscript received August 27, 2012; accepted for publication December 26, 2012
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.147785/-/DC1.
[1]Corresponding author: Department of Ecology and Evolution, University of Lausanne, Biophore, Room 4320, CH-1015 Lausanne, Switzerland. E-mail: severine.vuilleumier@unil.ch

genetic diversity and more particularly the expected induced transient dynamics have not received much attention.

Genetic diversity has a crucial importance in estimating populations at risk of extinction and species' adaptive potential. Current genetic diversity characterizes species at risk of extinction through inbreeding depression, loss of genetic diversity, and accumulation of deleterious mutations (Gilpin and Soule 1986; Jimenez *et al.* 1994; Frankham 1995; Hedrick and Kalinowski 2000). The current level of genetic diversity (or standing genetic variation) is now widely recognized as a determinant for the adaptation of a population to a novel environment (Turner *et al.* 1993; Feder *et al.* 2003; Pelz *et al.* 2005; Colosimo *et al.* 2005; Hermisson and Pennings 2005; Myles *et al.* 2005; Hernandez *et al.* 2011; Jones *et al.* 2012). First, under new selective pressures, the adaptive value of a preexisting allele can switch from neutral or deleterious to beneficial (Gibson and Dworkin 2004; Hermisson and Pennings 2005). Second, alleles from the standing genetic variation are present at higher frequencies in the population than any newly arisen (*de novo*) mutation are; thus, they have higher fixation probabilities and lower times to fixation (Barrett and Schluter 2008). Finally, these alleles have already passed successive selective filters and are consequently more likely to be compatible with the background genome (Orr and Betancourt 2001; Schluter *et al.* 2004; Barrett and Schluter 2008).

Measures of genetic diversity are widely used to understand and infer the demographic and evolutionary history of populations. Indeed, statistical tests using polymorphism data can detect departure from neutrality and infer demographic or selective processes (*e.g.*, Ewens 1972; Watterson 1978; Tajima 1983; Fu and Li 1993; Fay and Wu 2000; see review in Kreitman 2000). Furthermore, due to recent modeling advances in coalescent theory and increased genomic data and computational power, it is now possible to distinguish different demographic scenarios (*e.g.*, population bottleneck and subdivision; Peter *et al.* 2010) and estimate demographic and selective parameters (*e.g.*, populations size and growth rate, proportion of admixture, selection coefficient) using polymorphism data (Beaumont *et al.* 2002; Kim and Stephan 2002; Kim and Nielsen 2004; Nielsen *et al.* 2005; Price *et al.* 2009). Nevertheless, it is often difficult to distinguish between the transient effects of demographic changes and the effects of selection on polymorphism data (Jensen *et al.* 2005; Nielsen 2005; Li and Stephan 2006; Kim and Gulisija 2010; Pavlidis *et al.* 2010). It is also difficult to distinguish between the signatures of different demographic changes such as changes in population size, number, or migration rate (Wakeley 1999). A better understanding of the impact on genetic data of transient dynamics during demographic changes is necessary to disentangle these processes.

Interestingly, although the impact of population subdivision and short timescale population demographic changes on genetic diversity have received a lot of attention, other processes, such as long-term isolation and subsequent population reconnection, have received little attention. Such events have, without a doubt, occurred several times in the past, at long and short timescales. Repeated environmental changes have modified habitats and species distribution and created isolation and reconnection of populations. For example, during the climatic oscillations of the Quaternary period, temperate and tropical species were successively isolated into refugia and experienced habitat and population expansion, allowing for population reconnection (Hewitt 2000, 2004; Zhang *et al.* 2008; Young *et al.* 2009). At the same time, the reduction of sea levels (120 m lower than present; Lambeck *et al.* 2004) allowed the formation of land bridges that connected isolated lands in several parts of the world (Hewitt 2000). Repeated changes in water level resulted in fragmentation and fusion of basins within continents (as in the Great African Lakes; Galis and Metz 1998; Sturmbauer *et al.* 2001). Similarly, geological events such as volcanic eruptions induced periodic isolation and reconnection of islands (Cook 2008), while tectonic processes such as the formation of mountains isolated populations and reconnected others (Hughes and Eastwood 2006; Antonelli *et al.* 2009; Antonelli and Sanmartín 2011). More recently, climatic, environmental, and anthropogenic changes (*e.g.*, global warming, urbanization, and agriculture) have also played important roles in modifying the connectivity pattern between populations (Miller and Hobbs 2002; Delaney *et al.* 2010). Consequently, some species are currently subdivided into poorly connected or completely isolated populations; for examples, ground beetles (Keller *et al.* 2004), salamanders (Noel *et al.* 2007), and crickets (Vandergast *et al.* 2009). In the meantime, other species experience habitat and population expansion (*e.g.*, sparrows, white-tailed deer, zebra mussels; Waples 2010). Isolation and reconnection of populations not only reflect abiotic processes, but they can also represent spatial and temporal interactions of populations (*e.g.*, secondary contacts; Green *et al.* 2010; Domingues *et al.* 2012). Consequently, transient states of genetic diversity are expected to be the norm and deserve much more attention.

In this study, we analytically characterized the dynamics of genetic diversity following a change in migration rate between populations, given any migration rate, mutation rate, population size, and degree of fragmentation. We first analyzed how genetic diversity is affected by an event of isolation of populations and by an event of reconnection of populations. We then generalized our results for situations where the migration rate between populations displays strong variation. We demonstrate that temporal changes of migration generate periods where genetic diversity reaches unexpectedly high values that can be maintained over thousands of generations. We also show that migration changes can produce a signature on summary statistics such as Tajima's D and Ewens–Watterson's statistics that cannot be differentiated from a signature of population size change or from the signature of selection. Finally, we discuss how such processes can affect observed macroevolutionary patterns of

species diversity and how they can affect the reconstruction of populations' demographic and evolutionary history from genetic data.

Genetic Diversity of Populations

To study the dynamics of genetic diversity after connectivity changes, we consider diploid individuals in a finite island model composed of n random mating populations of size N, so that the total population size is nN. The populations exchange migrants at a rate m. The mutations follow the infinite allele model (each mutation produces a new allele; Kimura and Crow 1964) and occur at a rate μ. The generations are nonoverlapping (Wright–Fisher model; Fisher 1930; Wright 1931).

Genetic diversity, H, is estimated using the identity-by-descent F between pairs of alleles, through the relationship provided by Nei and Feldman (1972):

$$H = 1 - F. \tag{1}$$

Further, we characterize within-population genetic diversity H_s and between-population genetic diversity H_b, using within- and between-population genetic identities, respectively. Within-population genetic identity, F_s, corresponds to the probability that two genes randomly chosen from the same population are identical by descent. Between-population genetic identity, F_b, corresponds to the probability that two genes randomly chosen from different populations are identical by descent. Considering that within- and between-population genetic identities F_s and F_b at a given time t are, respectively, $F_{s,t}$ and $F_{b,t}$, their values at the next generation (forward in time), respectively, $F_{s,t+1}$ and $F_{b,t+1}$, will follow (Smith 1970; Maruyama 1970; Latter 1973)

$$\begin{aligned} F_{s,t+1} &= \left[a\left(c + (1-c)F_{s,t}\right) + (1-a)F_{b,t}\right](1-\mu)^2 \\ F_{b,t+1} &= \left[b\left(c + (1-c)F_{s,t}\right) + (1-b)F_{b,t}\right](1-\mu)^2, \end{aligned} \tag{2a}$$

where the parameters are

$$a = (1-m)^2 + \frac{m^2}{n-1} \tag{2b}$$

$$b = \frac{1-a}{(n-1)} \tag{2c}$$

$$c = \frac{1}{2N}. \tag{2d}$$

System of Equations 2 can be expressed under matrix notation as

$$\mathbf{F}_{t+1} = \mathbf{A}\mathbf{F}_t + \mathbf{B}, \tag{3a}$$

where

$$\mathbf{A} = (1-\mu)^2 \begin{pmatrix} a(1-c) & 1-a \\ b(1-c) & 1-b \end{pmatrix} \tag{3b}$$

$$\mathbf{B} = (1-\mu)^2 \begin{pmatrix} ac \\ bc \end{pmatrix} \tag{3c}$$

$$\mathbf{F}_t = \begin{pmatrix} F_{s,t} \\ F_{b,t} \end{pmatrix}. \tag{3d}$$

Parameters have the following interpretation: a is the probability that two genes at the same location before migration are still at the same location after migration (either both migrate to the same location or both do not migrate); b is the probability that two genes that were at the same location before migration migrated to different locations; c is the probability that two genes within a population are copies of the same gene; and $(1-\mu)^2$ is the probability that neither of the two randomly chosen genes mutated.

Predicting the Dynamics of Genetic Diversity

To characterize the impact of connectivity changes on genetic diversity, we analyzed the trajectories of within- and between-population genetic identities from any initial genetic identity state. Using Equation 2, Smith (1970) and Maruyama (1970) showed that genetic identities converge toward an equilibrium value $\mathbf{F}^{eq}$ (value given in Supporting Information, File S1). Extending the results obtained by Nei and Feldman (1972) for $n = 2$ populations, we show that the temporal dynamics of genetic identities follow (see Appendix A for more details)

$$\mathbf{F}_t = \mathbf{C}_1\lambda_1^t + \mathbf{C}_2\lambda_2^t + \mathbf{F}^{eq}, \tag{4a}$$

where

$$\mathbf{F}^{eq} = \begin{pmatrix} F_s^{eq} \\ F_b^{eq} \end{pmatrix}. \tag{4b}$$

λ_1 and λ_2 are respectively the largest and smallest eigenvalues of matrix $\mathbf{A}$, and they follow

$$\begin{aligned} \lambda_1 &= \frac{(1-\mu)^2}{2}\left[a(1-c) + 1 - b + \sqrt{(1-a(1-c)+b)^2 - 4bc}\right] \\ \lambda_2 &= \frac{(1-\mu)^2}{2}\left[a(1-c) + 1 - b - \sqrt{(1-a(1-c)+b)^2 - 4bc}\right]. \end{aligned} \tag{5}$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ are column vectors of dimension 2 composed of constant values, which depend on the parameters of the model (m, μ, n, and N) and on the initial genetic identity $\mathbf{F}_0$ (Appendix A).

In the next section, we provide from Equations 4a and 5 the temporal change of genetic diversity and derive the corresponding time to reach genetic diversity equilibrium after a connectivity change.

Time to Reach Genetic Diversity Equilibrium

The change of genetic diversity can be decomposed in two main temporal dynamics: a long-term and a short-term

dynamics. Indeed, the temporal change of genetic diversity depends on two components: $|\mathbf{C}_1\lambda_1^t|$ and $|\mathbf{C}_2\lambda_2^t|$ (Equation 4a). They both follow an exponential decay and their rate of change depends on $r_1 = \ln(\lambda_1)$ and $r_2 = \ln(\lambda_2)$, respectively (Appendix A). As $1 > \lambda_1 > \lambda_2 > 0$, $|\mathbf{C}_2\lambda_2^t|$ decays more rapidly than $|\mathbf{C}_1\lambda_1^t|$ (see Appendix A). Thus r_1 determines the ultimate (or long-term) change of genetic diversity and r_2 determines the transient (or short-term) change of genetic diversity.

When migration and mutation rates are small (*i.e.*, $m \ll 1$ and $\mu \ll 1$) and local population sizes are large (*i.e.*, $N \gg 1$), the decay constants r_1 and r_2 follow

$$\begin{aligned} r_1 &= -2\mu - \frac{1}{2N_e} \\ r_2 &= -2\mu - 2m\frac{n}{n-1} - \frac{1}{2N} + \frac{1}{2N_e} \end{aligned} \tag{6a}$$

with

$$N_e = nN\left(1 + \frac{(n-1)}{nM}\right), \tag{6b}$$

where $M = 4Nm$ is the scaled migration rate and N_e is the effective population size of the total population (inbreeding, eigenvalue, variance, and mutation effective size are equivalent in the finite island model; Whitlock and Barton 1997). As expected from theory (Whitlock and Barton 1997; Wakeley 1999), in the strong migration limit ($M \gg 1$), the effective size is equal to the total population size nN, while in the weak migration limit ($M \ll 1$), the effective size is higher than the total population size.

We can estimate the durations of the ultimate and transient changes of genetic diversity, denoted t_1 and t_2, respectively. Formally, we define t_1 and t_2 as the times (in number of generations) needed for λ_1^t and λ_2^t to be reduced to a number α, where $\alpha \in]0; 1]$:

$$\begin{aligned} \lambda_1^{t_1} &= \alpha \\ \lambda_2^{t_2} &= \alpha. \end{aligned} \tag{7}$$

Assuming that migration and mutation rates are small and population sizes are large, t_1 and t_2 simplify to (Appendix A):

$$\begin{aligned} t_1 &= \frac{-\ln(\alpha)}{2\mu + 1/2N_e} \\ t_2 &= \frac{-\ln(\alpha)}{2\mu + 2m(n/(n-1)) + 1/2N - 1/2N_e} \end{aligned} \tag{8}$$

The genetic diversity changes as follows (Figure 1): (i) a convergence of duration t_2 from the initial genetic diversity value to a transient genetic diversity value and then (ii) a convergence of duration t_1 to the genetic diversity equilibrium $\mathbf{H}^{eq}$. The time to reach genetic diversity equilibrium, t_1, depends only on two terms: the mutation rate (term 2μ) and the genetic drift at the total population level (term $1/2N_e$). The duration of the transient dynamics, t_2, depends on four terms: the mutation rate (term 2μ), the migration rate (term $2m$), the genetic drift in each population (term $1/2N$), and the genetic drift at the total population level (term $1/2N_e$). The convergence to the transient and equilibrium values of genetic diversity can occur on separated timescales (*i.e.*, $t_1 \gg t_2$) depending on the parameter values. The timescales t_1 and t_2 can differ from several orders of magnitude. When $n > 14$, differences are the highest ($t_1 \gg t_2$), in the domain where $\mu \ll 1/2N$ and also when $\mu \gg 1/2N$ and $m > \mu$. When $n \leq 14$, the same conditions apply for $t_1 \gg t_2$ except in a restricted domain where $m \simeq 1/2N$ (see Appendix A). For example, the duration of the transient dynamics is $t_2 \simeq 134$ and the time to reach equilibrium is $t_1 \simeq 1.5 \times 10^5$ generations (with $\alpha = 5\%$), when 10 populations of size 2500 with a mutation rate of 10^{-6} are connected with a migration rate of 0.01.

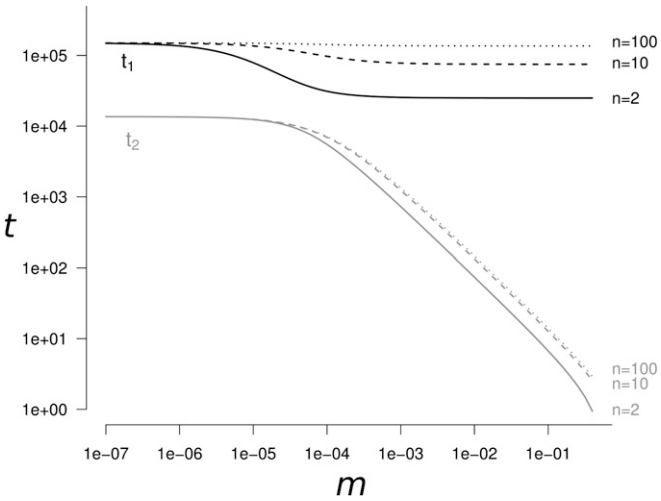

Figure 1 The time (in number of generations) t_1 to reach genetic diversity equilibrium and the length of the transient dynamics period t_2 as a function of the migration rate m. The solid line corresponds to n = 2 populations, the dashed line to n = 10, and the dotted line to n = 100. t_1 is always at least one order of magnitude higher than t_2. This separation of the two periods becomes even greater when $m > 1/4N = 10^{-4}$. Parameter values are N = 2500, $\mu = 10^{-5}$, α = 5%.

Dynamics of Genetic Diversity After an Isolation Event

We analyzed the dynamics of genetic diversity after an isolation event, starting with a situation in which populations are connected and at their equilibrium value; *i.e.*, within- and between-population genetic diversity H_s and H_b are at the expected connection equilibrium values $H_{s,con}^{eq}$ and $H_{b,con}^{eq}$ (see Maruyama 1970; Smith 1970; and File S1).

We observe (Figure 2) that immediately after an isolation event, within-population genetic diversity decreases due to genetic drift to the point where it reaches the mutation-drift equilibrium of an isolated population $H_{s,iso}^{eq}$ (see File S1 and Kimura and Crow 1964), at a rate determined by r_2 (from

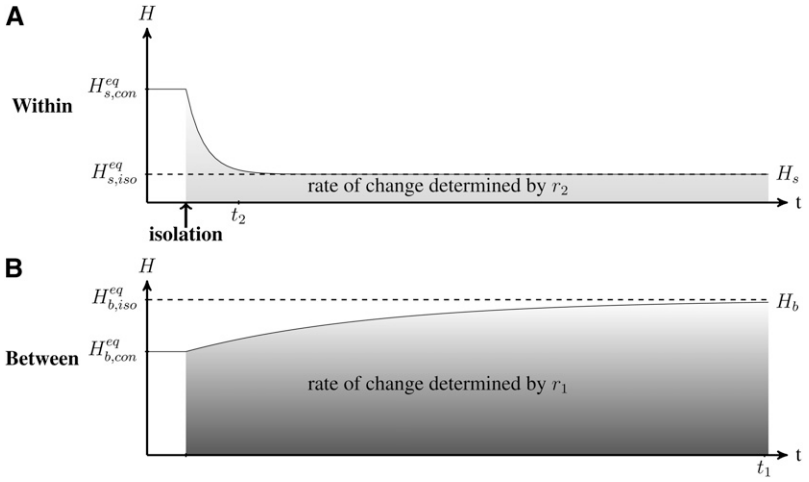

Figure 2 Dynamics of (A) within-population H_s and (B) between-population H_b genetic diversity after an isolation event. Within- and between-population diversity (solid lines) were previously at their respective connection equilibrium $H^{eq}_{s,con}$ and $H^{eq}_{b,con}$. After the isolation event, within- and between-population diversities reach their isolation equilibrium $H^{eq}_{s,iso}$ and $H^{eq}_{b,iso}$ (dashed lines) at rates determined by r_2 and r_1 (Equation 6). t_2 and t_1 estimate the time to reach the within- and between-population genetic diversity equilibrium, respectively (Equation 8). Under the effect of genetic drift, within-population diversity reaches its equilibrium value faster than between-population genetic diversity. Parameters are $n = 10$, $N = 2500$, $m = 10^{-4}$ before isolation and $m = 0$ afterward, and $\mu = 10^{-5}$. $t_1 \simeq$ 149,000 generations and $t_2 \simeq$ 13,600 generations (for $\alpha =$ 5%).

Equation 6). Meanwhile, between-population genetic diversity slowly increases due to the differentiation of populations induced by mutations (at a rate determined by r_1 from Equation 6). Populations ultimately reach complete differentiation (equilibrium value of $H^{eq}_{b,iso} = 1$). We can show from Equation 2 that following an isolation event, within- and between-population diversities change independently and reach their equilibrium value in t_2 and t_1 generations, respectively (File S2).

The decrease of population genetic diversity (within) can occur quickly relative to population differentiation (between-population genetic diversity; see Figure 1). After an isolation event, within-population genetic diversity (H_s) remains above its expected equilibrium $H^{eq}_{s,iso}$, while between-population genetic diversity (H_b) remains below its expected equilibrium $H^{eq}_{b,iso}$. However, the timescales of these nonequilibrium periods differ. When populations are isolated ($m = 0$), H_s reaches a value close to its equilibrium value in $t_2 \sim 1/(2\mu + [1/2N])$ generations, while H_b reaches a value close to its equilibrium value in $t_1 \sim 1/2\mu$ generations (Equation 8). Therefore, when $2\mu \ll 1/2N$, H_s converges much more quickly than H_b, and when $2\mu \gg 1/2N$, both converge in approximately the same amount of time. For example, assuming $\mu = 10^{-5}$ and $N = 1{,}000$, H_b is significantly lower than the equilibrium value for approximately $t_1 \simeq 150{,}000$ generations while H_s is significantly higher than the equilibrium value for approximately $t_2 \simeq$ 6000 generations (given $\alpha = 5\%$).

Dynamics of Genetic Diversity After a Connection Event

We analyzed the dynamics of genetic diversity after a connection event, starting with a situation in which populations are isolated and at their equilibrium value $H^{eq}_{s,iso}$ and $H^{eq}_{b,iso}$ (see Smith 1970; Maruyama 1970 and File S1). After a connection event (Figure 3), the genetic diversity accumulated in each population during the isolation period is quickly spread to all populations (Figure 3, fast dynamics in light shading, at a rate determined by r_2 from Equation 6). Consequently, within-population genetic diversity quickly increases and reaches a high value that is above its expected connected equilibrium value ($H^{eq}_{s,con}$, Figure 3; see File S1). This process creates a peak of within-population genetic diversity ΔH_s and a transient excess of genetic diversity between populations ΔH_b (see Figure 3). Then, due to genetic drift, both the within- and the between-population diversities decrease (slow dynamics in dark shading in Figure 3, at a rate determined by r_1 from Equation 6), to the point where the diversities reach the expected value of mutation–migration–drift equilibrium ($H^{eq}_{s,con}$ and $H^{eq}_{b,con}$ from Equation 4b; Kimura and Crow 1964).

Within- and between-population diversities change successively according to two timescales: first, a fast transient dynamics, followed by a slow asymptotic dynamics (separation of timescales is derived in Appendix A and illustrated in Figure 3). Because the transient dynamics can be shorter than the asymptotic dynamics, the excess of genetic diversity (ΔH_s and ΔH_b) can be maintained for a very long period (from Figure 1, t_1 is longer than 10,000 generations).

Peak of Genetic Diversity Generated by a Connection Event

In this section, we characterize the peak of within-population genetic diversity, ΔH_s, and the excess of between-population genetic diversity, ΔH_b, observed after a connection event as a function of the mutation rate, the genetic drift, the number of populations, and the migration rate after connection. The exact value of the within-population genetic diversity peak is represented in Figure 4. Assuming that migration and mutation rates are small, we can show that good approximations of the values of ΔH_s and ΔH_b are (see derivations in Appendix B)

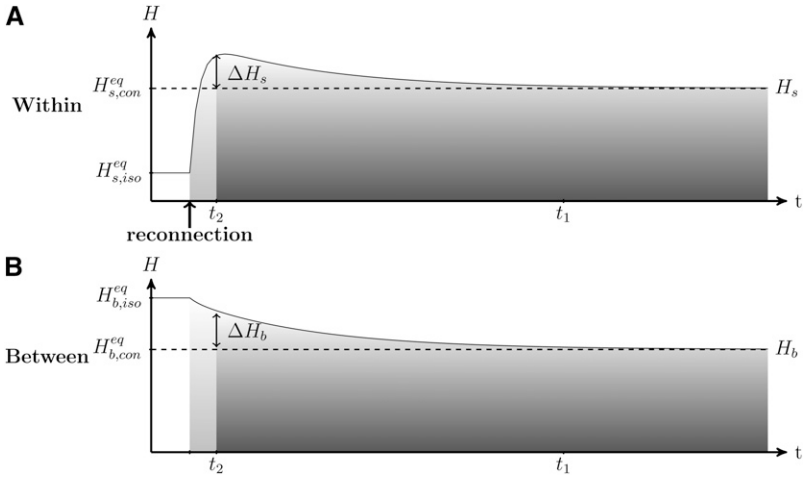

Figure 3 Dynamics of (A) within-population genetic diversity H_s and (B) between-population genetic diversity H_b after a reconnection event. Within- and between-population diversities were previously at their respective isolation equilibriums $H^{eq}_{s,iso}$ and $H^{eq}_{b,iso}$ (Equation 4b). After the reconnection event, within- and between-population diversities reach their respective connection equilibriums $H^{eq}_{s,con}$ and $H^{eq}_{b,con}$ (dashed lines). As shown in Equation 8, the time to reach genetic diversity equilibrium t_1 and the length of the transient period t_2 are well separated. The two periods are: (1) fast convergence at a rate determined by r_2 (Equation 6) that is driven by the spread of diversity that had accumulated within populations during isolation, which creates the peak of within-population diversity (ΔH_s) and the excess of between-population diversity (ΔH_b) and (2) slow dynamics at a rate determined by r_1 (Equation 6) that is caused by the gradual loss of genetic diversity. A large number of generations is needed to reach equilibrium. When $n = 10$, $N = 2500$, $m = 10^{-4}$ after reconnection, and $\mu = 10^{-5}$, $t_1 \simeq 97{,}000$ generations, $t_2 \simeq 6{,}900$ generations (for $\alpha = 5\%$) and $\Delta H_s \simeq \Delta H_b \simeq 0.11$.

$$\begin{aligned}\Delta H_s &= \left(1-\left(1-F^{\text{eq}}_{\text{s,iso}}\right)\frac{M(n/[n-1])}{1+M(n/[n-1])}\right)\\ &\quad\times F^{\text{eq}}_{\text{b,con}}\,\frac{M}{1+M(n/[n-1])-2N/N_{\text{e}}}0.05^{t_2/t_1}\\ \Delta H_b &= \left(1-\left(1-F^{\text{eq}}_{\text{s,iso}}\right)\frac{M(n/[n-1])}{1+M(n/[n-1])}\right)\\ &\quad\times F^{\text{eq}}_{\text{b,con}}\,\frac{1+M-N/N_{\text{e}}}{1+M(n/[n-1])-2N/N_{\text{e}}}0.05^{t_2/t_1},\end{aligned} \tag{9a}$$

where

$$F^{\text{eq}}_{\text{s,iso}} = \frac{1}{1+\theta} \tag{9b}$$

is the expected equilibrium identity within an isolated population (Kimura and Crow 1964), and

$$F^{\text{eq}}_{\text{b,con}} = \frac{M}{M+(n-1)\theta(1+\theta+(n/[n-1])M)} \tag{9c}$$

is the expected equilibrium identity between connected populations with the scaled migration rate M, the number of populations n, and the scaled mutation rate $\theta = 4N\mu$ (Maruyama 1970; Smith 1970). These approximations lead to the largest absolute error when a small number of populations ($n = 2$) is combined with weak mutation ($\theta < 1$) and intermediate migration ($M \simeq 5$). Nevertheless, this error is small (error <0.025 for ΔH_s and <0.08 for ΔH_b), so Equation 9 provides a good approximation of ΔH_s and ΔH_b for all n, M, and θ values (see Appendix B for more details about the validity of approximation 9).

The peak of genetic diversity increases with the difference between the two timescales (ΔH_s and ΔH_b increase with t_2/t_1) as genetic diversity increases during the transient phase and decreases during the asymptotic phase. Indeed, when those two phases are separated there is no loss of genetic diversity caused by the asymptotic decay during the transient phase (terms $0.05^{t_2/t_1} \simeq 1$ in Equation 9).

In the domain where the peak is the largest ($M \gg 1$ and $\theta \ll 1$), ΔH_s and ΔH_b reach the same value:

$$\Delta H^{\max} = \frac{n-1}{n}\frac{1}{1+n\theta}. \tag{10}$$

In this domain, $\Delta H^{\max}$ is maximized when the number of populations is (dashed line in Figure 4B):

$$n^* = 1+1/\sqrt{\theta} \tag{11}$$

The corresponding peak of genetic diversity, reached at n^*, is

$$\Delta H^{\max}|_{n^*} = \frac{1}{1+2\sqrt{\theta}}.$$

Interestingly, the number of populations and the peak of diversity have a nonmonotonous relationship. The peak of genetic diversity decreases when the number of populations approaches 2 and when it tends to infinity, while an intermediate number of populations n^* maximizes the peak of genetic diversity. This can be easily explained by the following processes. During isolation, a small number of populations accumulates less between-population genetic diversity; thus, once reconnected, they share a smaller amount of diversity. In contrast, a large number of populations accumulates a higher level of genetic diversity but also has a higher connection equilibrium value; thus, once reconnected, diversity reaches its expected equilibrium and no peak of diversity is observed.

In summary, high peaks of genetic diversity ($\Delta H_s > 0.25$ in Figure 4) can occur for a large range of the parameter space: when mutation is weak ($\theta < 0.05$) and migration is moderate to strong ($M > 0.5$). Under these conditions,

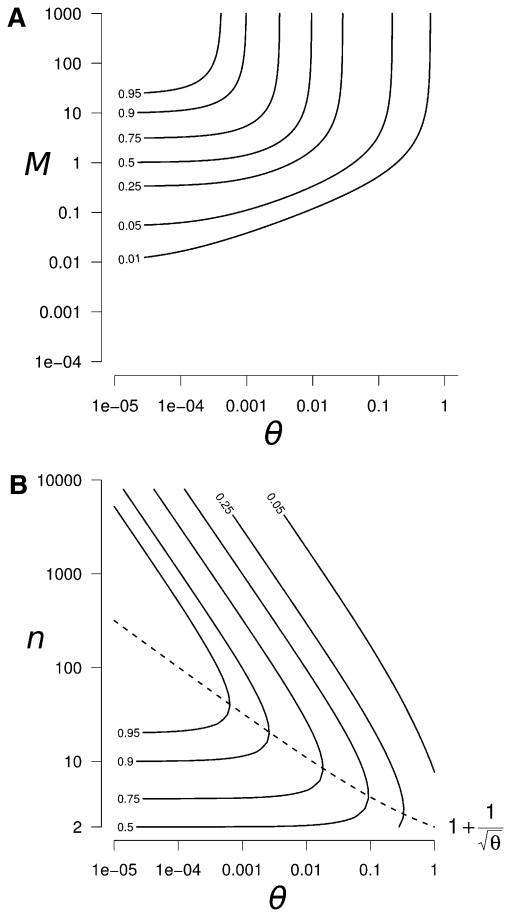

Figure 4 Peak of within-population genetic diversity ΔH_s generated by a reconnection event. (A) Contour plot of ΔH_s as a function of θ and M, for $n = 100$. We can clearly see the highest peak of diversity in the high M and low θ region. (B) Contour plot of the peak of genetic diversity after a reconnection event as a function of θ and n, for $M \gg 1$ (high M region identified in A). In the high M region, the within-population diversity peak ΔH_s and the between-population diversity excess ΔH_b are equal. The dashed line represents the number of populations which maximizes the peak of diversity $n^* = 1 + 1/\sqrt{\theta}$.

drastic genetic diversity changes can be observed (ΔH_s values >0.95; Figure 4B for $M \geq 50$ and $\theta < 5 \times 10^{-4}$). The number of populations that maximize the peak of diversity, n^*, ranges from a few populations when $\theta \simeq 1$, up to a few hundred populations when $\theta = 10^{-6}$ (values of $\theta < 10^{-6}$ are expected to be very rare, and they would require a mutation rate lower than 2.5×10^{-12}/bp for a 1-kb gene and a population size of 100). Interestingly, a significant peak of genetic diversity is also observed when only two populations reconnect ($\Delta H^{\max}|_{n=2} = 0.5$; Figure 4B).

Peak of Genetic Diversity Resulting from a Migration Rate Increase

Complete isolation of populations is not required to generate peaks of genetic diversity. Indeed, an abrupt increase of migration can generate the peak of genetic diversity characterized in the previous sections. In File S3, we determined that if migration crosses a threshold value M_T, peaks of genetic diversity can occur. The value of the threshold M_T, assuming that $m \ll 1$ and $\mu \ll 1$, is

$$M_T = (n-1)\theta \frac{1+\theta}{1+n\theta}. \quad (12)$$

Consequently, an increase of migration from M_0 to M crossing the threshold value M_T (*i.e.*, $M_0 \ll M_T$) generates a peak of genetic diversity that can be approximated by Equation 9 (see File S3). For example, in a subdivided population of $n = 10$ and $\theta = 0.1$, an increase in migration from $M_0 = 0.01$ to $M = 10$ (which crosses the migration threshold $M_T = 0.495$; Equation 12), generates a peak of within-population diversity of 0.350, while a reconnection event in a similar situation would generate a peak of similar intensity (0.358).

Implications for the Inference of Demography and Selection

To describe the impact of migration changes on the inference of demography and selection from genetic data, we described the dynamics of two broadly used summary statistics: the Ewens–Watterson statistics (Watterson 1978) and Tajima's D (Tajima 1989). Both the Ewens–Watterson statistics and Tajima's D are known to detect an excess (resp. deficit) of rare alleles, which induces negative (resp. positive) values of the statistics, compared with the expected neutral equilibrium (constant size population without selection). Usually, an excess of rare alleles is interpreted either as the signature of balancing selection or population expansion, and a deficit of rare alleles is interpreted as the signature of directional selection or as a population bottleneck. We used the Ewens–Watterson statistics, which we denote H_{EW} and follows (Watterson 1978)

$$H_{EW} = H_s - H_A, \quad (13)$$

where H_s is the genetic diversity, and H_A is the expected genetic diversity given the observed number of alleles K. We also used Tajima's D, which we denote D_T and follows (Tajima 1989)

$$D_T = \frac{\pi - S/a_1}{\sqrt{Var\,(\pi - S/a_1)}}, \quad (14)$$

where $a_1 = \sum_{i=1}^{m} 1/i$, l is the sample size, π is the average number of pairwise nucleotide differences, and S is the number of segregating sites.

We simulated samples of 50 sequences of 1 kb, with a per-nucleotide mutation rate of 2×10^{-8}, in four populations of size 2500, and ran 5000 replicate simulations. We simulated an isolation event, where the migration rate changed from 0.002 to 0 and a reconnection event in which the migration rate changed from 0 to 0.002. The simulations

were performed with the software *fastsimcoal* (Excoffier and Foll 2011), and the data analysis was performed with *Arlequin* (Excoffier and Lischer 2010). We simulated samples from the same population (with the parameter values that were used, the sampling scheme had a very weak impact; see Chikhi *et al.* 2010 for a discussion of how the sampling scheme affects the values of Tajima's *D*). To allow for convergence of the coalescent algorithm, we always assumed that the populations were connected prior to the isolation phase. In the reconnection event simulations, we set the duration of the isolated phase to 10*N*, which allowed the genetic diversity values to reach their equilibrium value.

We followed the dynamics of the statistics and estimated their distribution as a function of time. The results in Figure 5 show that an isolation event produces the same signature on the Ewens–Watterson statistics (Figure 5A) and Tajima's *D* (Figure 5D), as expected from a bottleneck event and from directional selection. Indeed, following an isolation event, genetic drift first causes the elimination of rare alleles and then eliminates more common alleles. Consequently, the number of alleles K decreases more quickly than genetic diversity H_s (Figure 5, B and C). Similarly, the number of segregating sites S decreases faster than the number of pairwise differences π (Figure 5, E and F). Therefore, D_T and H_{EW} are skewed toward positive values, as expected after a bottleneck or under the effect of directional selection. Moreover, the statistics remain skewed for a long period of time (<10,000 generations in our simulations, see Figure 5).

Results in Figure 6 show that a reconnection event can successively produce the same signature as expected from a population expansion or from a bottleneck event on H_{EW} (Figure 6A) and D_T (Figure 6, B and C). Indeed, following a reconnection event, migrants first create an excess of rare variants. The number of alleles K increases more quickly than the genetic diversity H_s (Figure 6, B and C), and the number of segregating sites S increases faster than the number of pairwise differences π (Figure 6, E and F), which skews H_{EW} and D_T toward negative values. Second, new alleles brought by migrants increase in frequency, creating an excess of common variants. Consequently H_s increases more than K, and π increases more than S, which skews D_T and H_{EW} toward positive values.

Interestingly, the observed duration of the periods in which both statistics are skewed are similar to the expected duration of the dynamics of genetic diversity (from Equation 8). After an isolation event, we observe, in Figure 5, that all statistics reach their equilibrium value within ~10,000 generations (4*N* generations). This duration corresponds to the value of the time required to reach within-population genetic diversity equilibrium after an isolation event, $t_2 \simeq$ 12,000 generations (4.8*N* generations, estimated from Equation 8 with $\alpha = 5\%$). In this example, genetic drift is stronger than mutation ($2\mu \ll 1/2N$) and thus $t_2 \sim 2N$. t_2 corresponds to the duration of the period in which the deficit of rare alleles skews the distribution of H_{EW} and D_T. After a reconnection event, we observe (Figure 6) that H_{EW} and D_T reach a "peak" within ~600 generations (0.24*N* generations). This duration corresponds to the value of the duration of the transient dynamics following a reconnection event, $t_2 \simeq$ 540 generations (0.216*N* generations, estimated from Equation 8 with $\alpha = 5\%$). In this example, migration is stronger than genetic drift and mutation ($m \gg 1/2N$ and $m \gg \mu$), and thus, $t_2 \sim 1/2m$. t_2 corresponds to the period during which the distribution of H_{EW} and D_T is skewed. Subsequently, H_{EW} and D_T reach their equilibrium value in ~80,000 generations (32*N* generations); this duration corresponds to the time required to reach the genetic diversity equilibrium value after a reconnection event, $t_1 \simeq$ 75,000 generations (30*N* generations, estimated from Equation 8 with $\alpha = 5\%$). t_1 corresponds to the period during which the deficit of rare alleles is eliminated.

In conclusion, both an isolation and a reconnection event induce changes in the proportion of rare alleles, which skews the values of H_{EW} and D_T, thus producing a signature that cannot be differentiated from the signature of past demographic events or of selection.

Discussion

We documented a simple neutral mechanism, which creates long-term peaks of genetic diversity. This peak of genetic diversity appears shortly after an abrupt increase in migration and is conserved for a long time. We also demonstrated that such genetic diversity peaks can occur for a large and plausible range of population sizes, migration rates, mutation rates, and numbers of populations. Subsequent to the genetic diversity peak, the rate of decay of genetic diversity was slow. Consequently, the mechanisms described here leave a strong and long-term footprint on genetic diversity that affects the Ewens–Watterson statistics and Tajima's *D* that are commonly used to infer the history of populations from genetic data.

The peak of genetic diversity is due to the spread of the genetic diversity accumulated during (partial) isolation. Therefore, the migration model that is assumed (island model of migration) is a leading factor in determining the strength of the observed genetic diversity peak. Assuming isolation by distance, the within-population genetic diversity is expected to have locally lower peaks. At the same time, under this assumption, the between-population genetic diversity is expected to be higher. Indeed, once populations are connected, each population shares with its neighboring populations alleles accumulated during isolation; thus, differentiation between distant populations will be maintained. Additionally, the amount of genetic diversity accumulated during isolation determines the size of the peak of genetic diversity. The maximum value is reached when populations are completely differentiated, *i.e.*, when no alleles are shared between populations. Our results are robust to the relaxation of the complete isolation and complete differentiation assumptions: when isolation is not complete

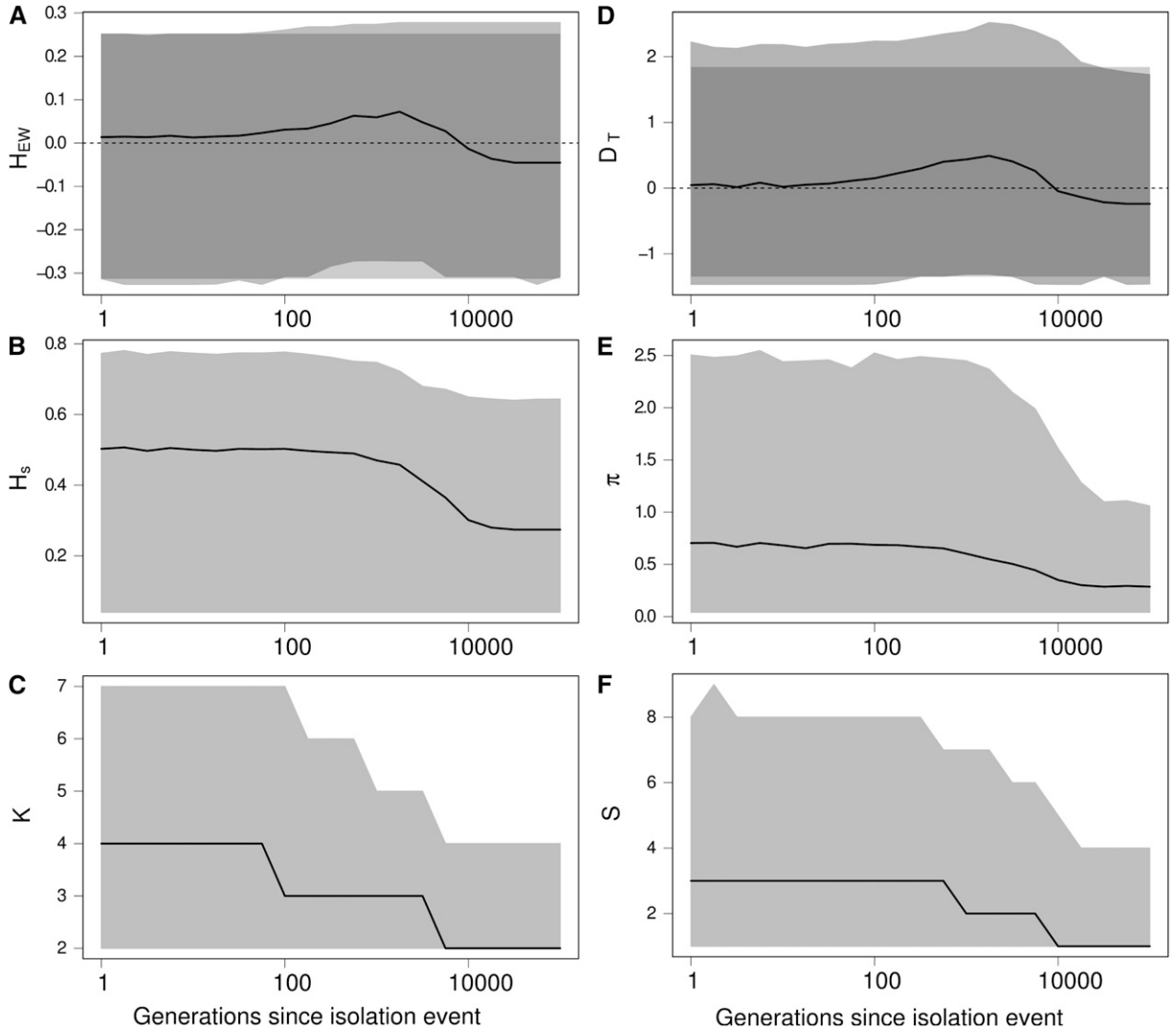

Figure 5 Effect of an isolation event on Ewens–Watterson and Tajima's *D* neutrality tests and on related summary statistics. (A) Ewens–Watterson statistics (H_{EW}), (B) genetic diversity (H_s), (C) number of alleles (K), (D) Tajima's *D* (D_T), (E) number of pairwise differences (π), and (F) number of segregating sites (S). For each statistics, the solid line represents the median of the distribution and the light shading represents the 97.5 and 2.5% quantiles of the distribution as a function of the number of generations t after the isolation event. Dark shading in A and D represent the expected distribution of the statistics in an isolated population at equilibrium. Values of H_{EW} and *D* after an isolation event are skewed toward positive values (signature of a bottleneck or directional selection), while there was no change in the size of the population. K and S decrease more quickly than H_s and π, because rare alleles are eliminated by genetic drift more quickly than common alleles. Coalescence simulations of a 1-kb locus with a mutation rate of 2×10^{-8}/bp, where four populations of size 2500 are isolated; 5000 replicates.

(because of small migration or nonequilibrium genetic diversity), we show that a genetic diversity peak is still observed (see File S3).

A connection event that occurs after an isolation period might play an important role on species diversification. Indeed, we have demonstrated that such events create an excess of genetic diversity. A high level of genetic diversity has often been hypothesized as being a key factor for species diversification. First, evolution from standing genetic variation might be stronger than from *de novo* mutation (Gibson and Dworkin 2004; Hermisson and Pennings 2005; Myles *et al.* 2005; Barrett and Schluter 2008). Second, both theoretical (Gavrilets 2003; Gavrilets and Losos 2009) and empirical work suggest that a high level of preexisting genetic diversity in a population increases its rate of diversification (Harmon *et al.* 2003; Seehausen 2004; Barrett and Schluter 2008). Interestingly, in several cases of adaptive radiation, a high genetic diversity of founder populations has been documented (*e.g.*, Barrier *et al.* 1999; Bezault *et al.* 2011). Several authors argued that the connection of populations after a period of isolation might have played an important role in many adaptive radiations (Hughes and Eastwood 2006; Antonelli and Sanmartín 2011; Bezault *et al.* 2011; Joyce *et al.* 2011). Therefore, species that experienced population isolation followed by reconnection events could have benefited from a temporary genetic diversity peak, which

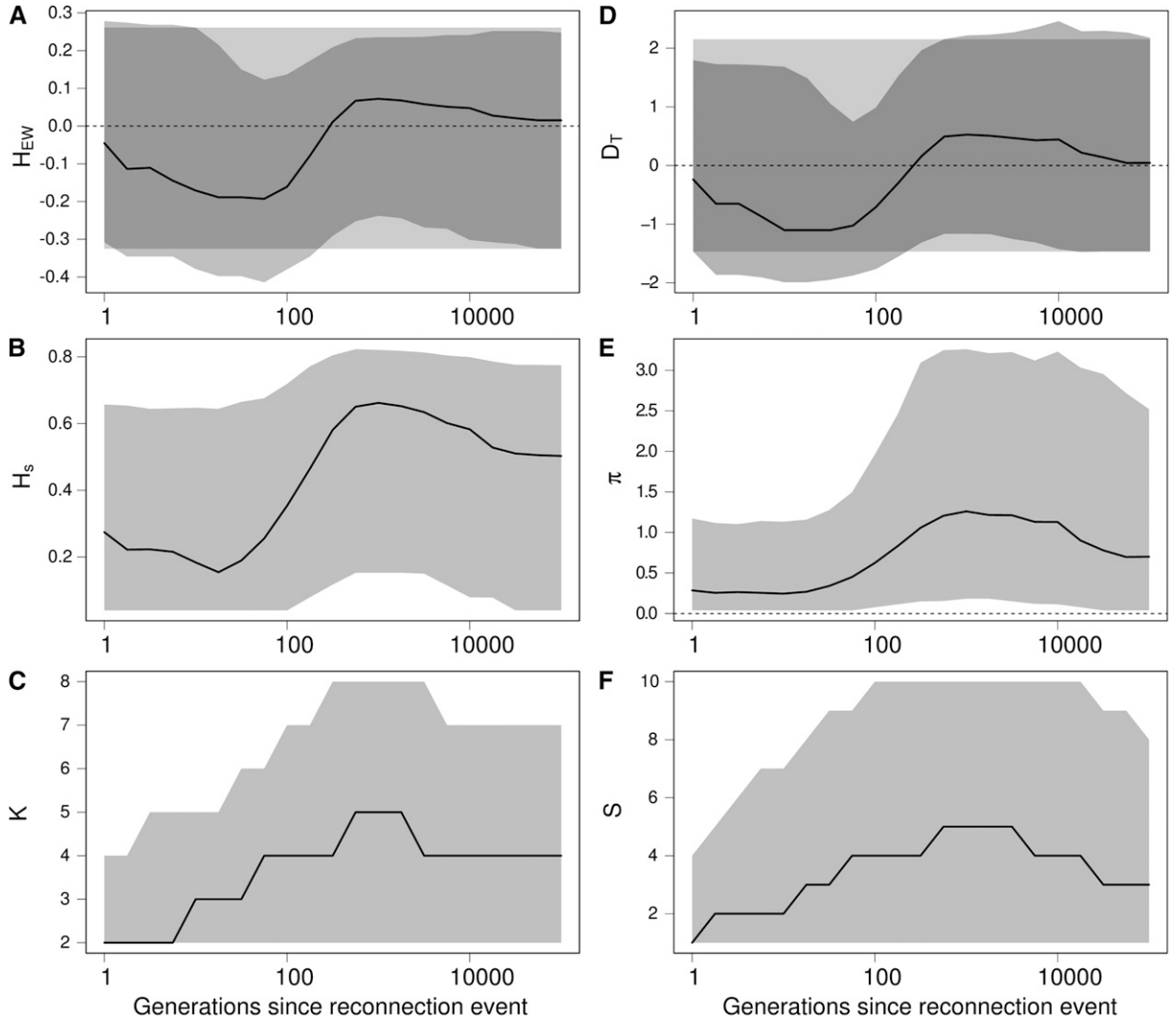

Figure 6 Effect of a reconnection event on Ewens–Watterson and Tajima's D neutrality tests and on related summary statistics. (A) Ewens–Watterson statistics (H_{EW}), (B) genetic diversity (H_s), (C) number of alleles (K), (D) Tajima's D (D_T), (E) number of pairwise differences (π), and (F) number of segregating sites (S). For each statistics, the solid line represents the median of the distribution, and the light shading represents the 97.5% and 2.5% quantiles of the distribution, as a function of the number of generations t after the isolation event. Dark shading in A and D represent the expected distribution of the statistics in an isolated equilibrium population. Values of H_{EW} and D after a reconnection event are first skewed toward negative values (signature of a population expansion or balancing selection) and then toward positive values (signature of a bottleneck or directional selection), while there was no change in the size of the population. K and S first increase more quickly than H_s and π because immigrants bring rare alleles, and then H_s and π reach a higher value because immigrant alleles increase in frequency. Finally, alleles are eliminated by genetic drift until the statistics reach their expected equilibrium value when populations are connected. Coalescence simulations of a 1 kb locus with a mutation rate of 2×10^{-8} per bp, where 4 populations of size 2500 isolated during 25,000 generations are reconnected with a migration rate $m = 0.002$; 5000 replicates.

has promoted the diversification of that species. Numerous species are known to have experienced such connectivity changes in the past and show remarkable levels of genetic and species diversity (Arnegard *et al.* 1999). For example, cichlid fishes in the great African lakes experienced periods of habitat fragmentation and reconnection due to lake water level fluctuations (Arnegard *et al.* 1999); there is some evidence that these processes might have played a role in the explosive radiation of the species (Owen *et al.* 1990; Young *et al.* 2009). Additionally, a high rate of speciation is correlated with the timeframe surrounding the uplift of the Northern Andes (Sedano and Burns 2010). The mechanisms described here are thus expected to considerably affect the ability of species to adapt to novel environmental conditions and to diversify over a very long period of time.

Statistics on allelic frequencies such as the Ewens–Watterson statistics (Watterson 1978) and Tajima's D (Tajima 1989) allow the inference of either selection or population demographic changes. Here, we demonstrate that migration changes can lead to signatures that cannot be differentiated from a selection process or a population size change when using the Ewens–Watterson test and Tajima's D. Therefore, past migration changes must be considered more carefully and should be viewed as an alternative explanation of bias

in neutrality tests and bottleneck or expansion signals. Recently, authors have shown that population structure can bias neutrality tests and produce false bottleneck signals (Leblois *et al.* 2006; Städler *et al.* 2009; Chikhi *et al.* 2010) and that shortly after an isolation event departure from the neutrality can be incorrectly inferred (as shown with simulations by Broquet *et al.* 2010 and discussed in Waples 2010). The proper interpretation of genetic signatures is crucial for the understanding of the evolutionary history of populations. An interesting extension of this work would be to analyze in more detail the molecular signature of the mechanisms described here and to provide methods that allow the differentiation of such events from selection or demographic changes. Moreover, our results are also relevant for the study of genealogies. Indeed, genetic identities as considered here are commonly used to describe coalescence time distributions (Slatkin 1991; Rousset 1996; Wakeley 1999). Future investigations should also investigate the consequences of isolation and connection events on phylogenetic tree reconstruction. Statistical tools that are available to estimate demographic parameters classically focus on *a priori* specific scenarios (*e.g.*, population bottleneck, expansion, population with constant migration, population split with subsequent migration; see review in Kuhner 2009). Given the strong impact of migration changes on genetic diversity, accounting for such scenarios is necessary. Recent methods allowing a larger range of population demographic scenarios, such as approximate Bayesian computation (Beaumont *et al.* 2002; Beaumont 2010), may be powerful tools with which to disentangle the signature of demographic processes from the observed genetic diversity.

One of the major goals of conservation genetics is to maintain genetic diversity, decrease extinction risks, avoid inbreeding depression, maintain species evolutionary potential, and decrease species vulnerability to environmental change (Gilpin and Soule 1986; Newman and Pilson 1997; Jump *et al.* 2009). In this context, conservationists need to estimate the genetic diversity of a population and its effective size. Such measures are commonly obtained from genetic data and are estimated with standard statistics (Wright 1950; Jorde and Ryman 2007). Although new approaches that consider populations at a nonequilibrium state are emerging, to estimate population size changes and instantaneous migration rates (*e.g.*, Hey and Nielsen 2004), the expected level of genetic diversity is still commonly estimated assuming that populations are at an equilibrium. As shown here, genetic diversity is more likely to be in a transient state. We have demonstrated that reconnecting isolated populations increases genetic diversity above the expected equilibrium value, while isolating populations induces a slow decrease of genetic diversity. Consequently, any estimate inferred from data collected from a population that underwent strong migration changes will not reflect the demographic situation of the population (*e.g.*, census size, genetic diversity). This can have drastic consequences on the selection of conservation strategies and for the management of species (Pearse and Crandall 2004; Caballero *et al.* 2010).

Acknowledgments

The authors thank two anonymous reviewers and Noah Rosenberg for their valuable comments and suggestions, which increased the clarity and the scope of the article. This project was funded by the Swiss National Research Foundation (SNRF) grant nos. PZ00P3–121702, PZ00P3–139421/1, and 31003A–130065.

Literature Cited

Antonelli, A., and I. Sanmartín, 2011 Why are there so many plant species in the neotropics? Taxon 60: 403–414.

Antonelli, A., J. A. A. Nylander, C. Persson, and I. Sanmartín, 2009 Tracing the impact of the andean uplift on neotropical plant evolution. Proc. Natl. Acad. Sci. USA 106: 9749–9754.

Arnegard, M. E., J. A. Markert, P. D. Danley, J. R. Stauffer, A. J. Ambali *et al.*, 1999 Population structure and colour variation of the cichlid fishes labeotropheus fuelleborni ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. Proc. R. Soc. Lond. B Biol. Sci. 266: 119–130.

Barrett, R. D. H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends Ecol. Evol. 23: 38–44.

Barrier, M., B. G. Baldwin, R. H. Robichaux, and M. D. Purugganan, 1999 Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. Mol. Biol. Evol. 16: 1105–1113.

Beaumont, M. A., 2010 Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41: 379–406.

Beaumont, M. A., W. Zhang, and D. J. Balding, 2002 Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035.

Bezault, E., S. Mwaiko, and O. Seehausen, 2011 Population genomic tests of models of adaptive radiation in Lake Victoria region Cichlid fish. Evolution 65: 3381–3397.

Broquet, T., S. Angelone, J. Jaquiery, P. Joly, J.-P. Lena *et al.*, 2010 Genetic bottlenecks driven by population disconnection. Conserv. Biol. 24: 1596–1605.

Caballero, A., S. T. Rodriguez-Ramilo, V. Avila, and J. Fernandez, 2010 Management of genetic diversity of subdivided populations in conservation programmes. Conserv. Genet. 11: 409–419.

Chikhi, L., V. C. Sousa, P. Luisi, B. Goossens, and M. A. Beaumont, 2010 The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. Genetics 186: 983–995.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson *et al.*, 2005 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307: 1928–1933.

Cook, L. M., 2008 Species richness in Madeiran land snails, and its causes. J. Biogeogr. 35: 647–653.

Delaney, K. S., S. P. D. Riley, and R. N. Fisher, 2010 A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. PLoS ONE 5: e12767.

Domingues, V. S., Y.-P. Poh, B. K. Peterson, P. S. Pennings, J. D. Jensen *et al.* 2012 Evidence of adaptation from ancestral variation in young populations of beach mice. Evolution 66: 3209–3223.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3: 87–112.

Excoffier, L., and M. Foll, 2011 fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27: 1332–1334.

Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. Mol. Ecol. Res. 10: 564–567.
Excoffier, L., M. Foll, and R. J. Petit, 2009 Genetic consequences of range expansions. Annu. Rev. Ecol. Evol. Syst. 40: 481–501.
Falush, D., M. Stephens, and J. K. Pritchard, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.
Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive darwinian selection. Genetics 155: 1405–1413.
Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith *et al.*, 2003 Allopatric genetic origins for sympatric host-plant shifts and race formation in rhagoletis. Proc. Natl. Acad. Sci. USA 100: 10314–10319.
Fisher, R. A., 1922 Darwinian evolution of mutations. Eugen. Rev. 14: 31–34.
Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
Frankham, R., 1995 Conservation genetics. Annu. Rev. Genet. 29: 305–327.
Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics 133: 693–709.
Galis, F., and J. A. Metz, 1998 Why are there so many cichlid species? Trends Ecol. Evol. 13: 1–2.
Gavrilets, S., 2003 Perspective: models of speciation: What have we learned in 40 years? Evolution 57: 2197–2215.
Gavrilets, S., and J. B. Losos, 2009 Adaptive radiation: contrasting theory with data. Science 323: 732–737.
Gibson, G., and I. Dworkin, 2004 Uncovering cryptic genetic variation. Nat. Rev. Genet. 5: 681–690.
Gilpin, M., and M. Soule, 1986 *Minimum viable populations: processes of species extinction*, pp. 19–34 in *Conservation Biology: The Science of Scarcity and Diversity*. Sinauer, Sunderland, MA.
Gravel, S., 2012 Population genetics models of local ancestry. Genetics 191: 607–619.
Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel *et al.*, 2010 A draft sequence of the neandertal genome. Science 328: 710–722.
Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos, 2003 Tempo and mode of evolutionary radiation in iguanian lizards. Science 301: 961–964.
Hedrick, P. W., and S. T. Kalinowski, 2000 Inbreeding depression in conservation biology. Annu. Rev. Ecol. Syst. 31: 139–162.
Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169: 2335–2352.
Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton *et al.*, 2011 Classic selective sweeps were rare in recent human evolution. Science 331: 920–924.
Hewitt, G., 2000 The genetic legacy of the quaternary ice ages. Nature 405: 907–913.
Hewitt, G. M., 2004 Genetic consequences of climatic oscillations in the quaternary. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359: 183–195, discussion 195.
Hey, J., and R. Nielsen, 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167: 747–760.
Hughes, C., and R. Eastwood, 2006 Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the andes. Proc. Natl. Acad. Sci. USA 103: 10334–10339.
Jensen, J. D., Y. Kim, V. B. DuMont, C. F. Aquadro, and C. D. Bustamante, 2005 Distinguishing between selective sweeps and demography using dna polymorphism data. Genetics 170: 1401–1410.
Jimenez, J. A., K. A. Hughes, G. Alaks, L. Graham, and R. C. Lacy, 1994 An experimental study of inbreeding depression in a natural habitat. Science 266: 271–273.
Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55–61.
Jorde, P. E., and N. Ryman, 2007 Unbiased estimator for genetic drift and effective population size. Genetics 177: 927–935.
Joyce, D. A., D. H. Lunt, M. J. Genner, G. F. Turner, R. Bills *et al.*, 2011 Repeated colonization and hybridization in Lake Malawi cichlids. Curr. Biol. 21: R108–R109.
Jump, A. S., R. Marchant, and J. Peñuelas, 2009 Environmental change and the option value of genetic diversity. Trends Plant Sci. 14: 51–58.
Karlin, S., 1982 Classifications of selection migration structures and conditions for a protected polymorphism. Evol. Biol. 14: 61–204.
Keller, I., W. Nentwig, and C. R. Largiader, 2004 Recent habitat fragmentation due to roads can lead to significant genetic differentiation in an abundant flightless ground beetle. Mol. Ecol. 13: 2983–2994.
Kim, Y., and D. Gulisija, 2010 Signatures of recent directional selection under different models of population expansion during colonization of new selective environments. Genetics 184: 571–585.
Kim, Y., and R. Nielsen, 2004 Linkage disequilibrium as a signature of selective sweeps. Genetics 167: 1513–1524.
Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777.
Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.
Kreitman, M., 2000 Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. 1: 539–559.
Kuhner, M. K., 2009 Coalescent genealogy samplers: windows into population history. Trends Ecol. Evol. 24: 86–93.
Lambeck, K., F. Antonioli, A. Purcell, and S. Silenzi, 2004 Sea-level change along the Italian coast for the past 10,000 yr. Quat. Sci. Rev. 23: 1567–1598.
Latter, B. D., 1973 The island model of population differentiation: a general solution. Genetics 73: 147–157.
Leblois, R., A. Estoup, and R. Streiff, 2006 Genetics of recent habitat contraction and reduction in population size: Does isolation by distance matter? Mol. Ecol. 15: 3601–3615.
Li, H., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2: e166.
Maruyama, T., 1970 Effective number of alleles in a subdivided population. Theor. Popul. Biol. 1: 273–306.
Miller, J., and R. Hobbs, 2002 Conservation where people live and work. Conserv. Biol. 16: 330–337.
Myles, S., N. Bouzekri, E. Haverfield, M. Cherkaoui, J.-M. Dugoujon *et al.*, 2005 Genetic evidence in support of a shared Eurasian–North African dairying origin. Hum. Genet. 117: 34–42.
Nagylaki, T., 1974 The decay of genetic variability in geographically structured populations. Proc. Natl. Acad. Sci. USA 71: 2932–2936.
Nagylaki, T., 1977 Decay of genetic variability in geographically structured populations. Proc. Natl. Acad. Sci. USA 74: 2523–2525.
Nagylaki, T., 1979 The island model with stochastic migration. Genetics 91: 163–176.
Nei, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70: 3321–3323.
Nei, M., and M. W. Feldman, 1972 Identity of genes by descent within and between populations under mutation and migration pressures. Theor. Popul. Biol. 3: 460–465.
Nei, M., T. Maruyama, and R. Chakraborty, 1975 The bottleneck effect and genetic variability in populations. Evolution 29: 1–10.

Newman, D., and D. Pilson, 1997 Increased probability of extinction due to decreased genetic effective population size: experimental populations of Clarkia pulchella. Evolution 51: 354–362.
Nielsen, R., 2005 Molecular signatures of natural selection. Annu. Rev. Genet. 39: 197–218.
Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton *et al.*, 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3: e170.
Noel, S., M. Ouellet, P. Galois, and F.-J. Lapointe, 2007 Impact of urban fragmentation on the genetic structure of the eastern red-backed salamander. Conserv. Genet. 8: 599–606.
Orr, H. A., and A. J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875–884.
Owen, R. B., R. Crossley, T. C. Johnson, D. Tweddle, I. Kornfield *et al.*, 1990 Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. Proc. R. Soc. Lond. B Biol. Sci. 240: 519–553.
Pavlidis, P., J. D. Jensen, and W. Stephan, 2010 Searching for footprints of positive selection in whole-genome snp data from nonequilibrium populations. Genetics 185: 907–922.
Pearse, D. E., and K. A. Crandall, 2004 Beyond fst: analysis of population genetic data for conservation. Conserv. Genet. 5: 585–602 .10.1007/s10592–004–1863-z.
Pelz, H.-J., S. Rost, M. Hünerberg, A. Fregin, A.-C. Heiberg *et al.*, 2005 The genetic basis of resistance to anticoagulants in rodents. Genetics 170: 1839–1847.
Peter, B. M., D. Wegmann, and L. Excoffier, 2010 Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. Mol. Ecol. 19: 4648–4660.
Price, A. L., A. Tandon, N. Patterson, K. C. Barnes, N. Rafaels *et al.*, 2009 Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. 5: e1000519.
Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
Rice, S. H., and A. Papadopoulos, 2009 Evolution with stochastic fitness and stochastic migration. PLoS ONE 4: e7130.
Rousset, F., 1996 Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142: 1357–1362.
Schluter, D., E. A. Clifford, M. Nemethy, and J. S. McKinnon, 2004 Parallel evolution and inheritance of quantitative traits. Am. Nat. 163: 809–822.
Sedano, R. E., and K. J. Burns, 2010 Are the Northern Andes a species pump for neotropical birds?: Phylogenetics and biogeography of a clade of neotropical tanagers (aves: Thraupini). J. Biogeogr. 37: 325–343.
Seehausen, O., 2004 Hybridization and adaptive radiation. Trends Ecol. Evol. 19: 198–207.
Shpak, M., J. Wakeley, D. Garrigan, and R. C. Lewontin, 2010 A structured coalescent process for seasonally fluctuating populations. Evolution 64: 1395–1409.
Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. 58: 167–175.
Smith, J. M., 1970 Population size, polymorphism, and the rate of non-Darwinian evolution. Am. Nat. 104: 231–237.
Städler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber, 2009 The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics 182: 205–216.
Sturmbauer, C., S. Baric, W. Salzburger, L. Rüber, and E. Verheyen, 2001 Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. Mol. Biol. Evol. 18: 144–154.
Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437–460.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
Takahata, N., and M. Nei, 1985 Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110: 325–344.
Turner, R. C., J. C. Levy, and A. Clark, 1993 Complex genetics of type 2 diabetes: thrifty genes and previously neutral polymorphisms. Q. J. Med. 86: 413–417.
Vandergast, A. G., E. A. Lewallen, J. Deas, A. J. Bohonak, D. B. Weissman *et al.*, 2009 Loss of genetic connectivity and diversity in urban microreserves in a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus n. sp "santa monica"). J. Insect Conserv. 13: 329–345.
Wakeley, J., 1999 Nonequilibrium migration in human history. Genetics 153: 1863–1871.
Waples, R. S., 2010 Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Mol Ecol Resour 10: 785–796.
Watterson, G. A., 1978 The homozygosity test of neutrality. Genetics 88: 405–417.
Whitlock, M. C., 1992 Temporal fluctuations in demographic parameters and the genetic variance among populations. Evolution 46: 608–615.
Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics 146: 427–441.
Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.
Wright, S., 1950 Genetical structure of populations. Nature 166: 247–249.
Young, K. A., J. M. Whitman, and G. F. Turner, 2009 Secondary contact during adaptive radiation: a community matrix for Lake Malawi cichlids. J. Evol. Biol. 22: 882–889.
Zhang, H., J. Yan, G. Zhang, and K. Zhou, 2008 Phylogeography and demographic history of Chinese black-spotted frog populations (Pelophylax nigromaculata): evidence for independent refugia expansion and secondary contact. BMC Evol. Biol. 8: 21.

Communicating editor: N. A. Rosenberg

Appendix A: Dynamics of Genetic Diversity

In this appendix, we describe the temporal change of genetic diversity (derivation of Equations 4a and 8 and the separation of the dynamics of genetic diversity into two timescales).

Temporal change of genetic diversity

The solution to Equation 3 is

$$\mathbf{F}_t = \mathbf{A}^t(\mathbf{F}_0 - \mathbf{F}^{\mathrm{eq}}) + \mathbf{F}^{\mathrm{eq}}. \tag{A1}$$

Denoting $\mathbf{P}$ the transformation matrix with eigenvector $\mathbf{U}_1$ (associated with λ_1) as first column and eigenvector $\mathbf{U}_2$ (associated with λ_2) as second column,

$$\mathbf{P} = (\,\mathbf{U}_1 \quad \mathbf{U}_2\,) \tag{A2}$$

and denoting

$$\begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix} = \mathbf{P}^{-1}(\mathbf{F}_0 - \mathbf{F}^{\mathrm{eq}}), \tag{A3}$$

Equation A1 becomes

$$\begin{aligned} \mathbf{F}_t &= \mathbf{P}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^t \mathbf{P}^{-1}(\mathbf{F}_0 - \mathbf{F}^{\mathrm{eq}}) + \mathbf{F}^{\mathrm{eq}} \\ &= \mathbf{P}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^t \begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix} + \mathbf{F}^{\mathrm{eq}} \\ &= y_{10}\mathbf{U}_1\lambda_1^t + y_{20}\mathbf{U}_2\lambda_2^t + \mathbf{F}^{\mathrm{eq}}. \end{aligned} \tag{A4}$$

Further, denoting $\mathbf{C}_1 = y_{10}\mathbf{U}_1$ and $\mathbf{C}_2 = y_{20}\mathbf{U}_2$ leads to Equation 4a.

$\mathbf{F}_t$ changes according to two exponential decay functions, λ_1^t and λ_2^t. Their rates of change are $d\lambda_1^t/dt = \ln(\lambda_1)\times e^{\ln(\lambda_1)\mathrm{t}}$ and $d\lambda_2^t/dt = \ln(\lambda_2)\times e^{\ln(\lambda_2)\mathrm{t}}$, respectively. Thus $r_1 = \ln(\lambda_1)$ and $r_2 = \ln(\lambda_2)$ are the decay constants that determine the rate of change of functions λ_1^t and λ_2^t.

Therefore, the eigenvalues of matrix $\mathbf{A}$ can be used to compute the rates of change of genetic diversity. As $\lambda_1 > \lambda_2$ and both eigenvalues are <1, we have $|r_2| > |r_1|$, and thus $\mathbf{C}_2\lambda_2^t$ tends to 0 faster than $\mathbf{C}_1\lambda_1^t$. r_2 determines the transient rate of change of genetic diversity, while r_1 determines the asymptotic rate of change of genetic diversity.

We now want to simplify the expression of the rates of change of genetic diversity. To do so, we can rewrite Equation 5 as

$$\begin{aligned} \lambda_1 &= \frac{(1-\mu)^2}{2}\left[a(1-c) + 1 - b + (1 - a(1-c) + b)\sqrt{1 - \frac{4bc}{(1-a(1-c)+b)^2}}\right] \\ \lambda_2 &= \frac{(1-\mu)^2}{2}\left[a(1-c) + 1 - b - (1 - a(1-c) + b)\sqrt{1 - \frac{4bc}{(1-a(1-c)+b)^2}}\right]. \end{aligned}$$

With $4bc \ll (1 - a(1-c) + b)^2$ (as $m \ll 1$ and $N \gg 1$), and given that $\sqrt{1-x} = 1 - \frac{1}{2}x + o(x)$, Equation 5 further simplifies to

$$\begin{aligned} \lambda_1 &= (1-\mu)^2\left(1 - \frac{bc}{1 - a(1-c) + b}\right) \\ \lambda_2 &= (1-\mu)^2\left(a(1-c) - b + \frac{bc}{1 - a(1-c) + b}\right). \end{aligned}$$

Considering that migration rates and mutation rates are small, we can neglect terms in m^2, $1/N^2$, μ^2, m/N, $m\mu$, and μ/N, which leads to

$$\lambda_1 = 1 - 2\mu - 1/\left[2nN\left(1 + \frac{n-1}{nM}\right)\right]$$
$$\lambda_2 = 1 - 2\mu - 2m\frac{n}{n-1} - \frac{1}{2N} + 1/\left[2nN\left(1 + \frac{n-1}{nM}\right)\right] \tag{A5}$$

and thus

$$r_1 = ln\left(1 - 2\mu - 1/\left[2nN\left(1 + \frac{n-1}{nM}\right)\right]\right)$$
$$r_2 = ln\left(1 - 2\mu - 2m\frac{n}{n-1} - \frac{1}{2N} + 1/\left[2nN\left(1 + \frac{n-1}{nM}\right)\right]\right). \tag{A6}$$

Assuming that migration and mutation rates are small, and that local population sizes are large ($N \gg 1$), Equation A6 simplifies to Equation 6 as $\ln(1 - x) = -x + o(x)$.

Respective length of the asymptotic and transient dynamics periods

We denote t_1 and t_2 as the times needed for λ_1^t and λ_2^t to be reduced to a value α, where $\alpha \in]0; 1]$,

$$\lambda_1^{t_1} = \alpha$$
$$\lambda_2^{t_2} = \alpha, \tag{A7}$$

which leads to

$$t_1 = \frac{\ln(\alpha)}{r_1}$$
$$t_2 = \frac{\ln(\alpha)}{r_2}. \tag{A8}$$

Assuming small migration and mutation rates and large local population sizes, we can replace the expressions of r_1 and r_2 from Equation 6 into Equation A8, and we obtain Equation 8. $\mathbf{F}_t$ approximately follows

$$\mathbf{F}_t \simeq \begin{cases} \mathbf{F}^{eq} + \mathbf{C}_1\lambda_1^t + \mathbf{C}_2\lambda_2^t & \text{for } t < t_2 \\ \mathbf{F}^{eq} + \mathbf{C}_1\lambda_1^t & \text{for } t_2 < t < t_1 \\ \mathbf{F}^{eq} & \text{for } t_1 < t. \end{cases} \tag{A9}$$

Timescales separation

This section presents the conditions for $t_1 \gg t_2$. When $t_1 \gg t_2$, $\lambda_2^{t_2} \simeq 0$ and $\lambda_1^{t_2} \simeq 1$; thus, Equation A9 simplifies to

$$\mathbf{F}_t \simeq \begin{cases} \mathbf{F}^{eq} + \mathbf{C}_1 + \mathbf{C}_2\lambda_2^t & \text{for } t < t_2 \\ \mathbf{F}^{eq} + \mathbf{C}_1\lambda_1^t & \text{for } t_2 < t < t_1 \\ \mathbf{F}^{eq} & \text{for } t_1 < t. \end{cases} \tag{A10}$$

Equation A10 decomposes the dynamics of $\mathbf{F}_t$ into two timescales: a *transient period* of length t_2 and an *asymptotic period* of length $t_1 - t_2 \simeq t_1$. For $t > t_1$, the genetic identity is close to its equilibrium value $\mathbf{F}^{eq}$, so t_1 can be interpreted as the duration of the disequilibrium period.

Equation A10 is true if a t exists such that $\lambda_1^t \simeq 1$ and $\lambda_2^t \simeq 0$. For simplicity, we consider that $\lambda_1^t \simeq 1$ if $\lambda_1^t \geq 0.95$ and that $\lambda_2^t \simeq 0$ if $\lambda_2^t \leq 0.05$. Thus, Equation A10 is a good approximation if a t exists such that

$$\begin{matrix} \lambda_1^t \geq 0.95 \\ \lambda_2^t \leq 0.05 \end{matrix} \Rightarrow \begin{cases} t \leq \dfrac{\ln(0.95)}{\ln(\lambda_1)} \\ \lambda_2^t \leq 0.05. \end{cases} \tag{A11}$$

Thus, showing that

$$\lambda_2^{\ln(0.95)/\ln(\lambda_1)} \leq 0.05 \tag{A12}$$

demonstrates that for $t = \ln(0.95)/\ln(\lambda_1)$, we have $\lambda_1^t \geq 0.95$ and $\lambda_2^t \leq 0.05$. This provides a sufficient proof of proposition (A11).

We can demonstrate that proposition (A12) depends only on the ratio $t_1/t_2 = \ln(\lambda_2)/\ln(\lambda_1)$ (see definitions of t_1 and t_2 Equation 7). Proposition (A12) leads to

$$\begin{aligned}
&\lambda_2^{\ln(0.95)/\ln(\lambda_1)} \leq 0.05 \\
&\Leftrightarrow e^{\ln(0.95)[ln(\lambda_2)/\ln(\lambda_1)]} \leq 0.05 \\
&\Leftrightarrow \ln(0.95)\frac{t_1}{t_2} \leq \ln(0.05) \\
&\Leftrightarrow \frac{t_1}{t_2} \geq \frac{\ln(0.05)}{\ln(0.95)}.
\end{aligned}$$

Therefore, the condition $t_1/t_2 \geq \ln(0.05)/\ln(0.95)$ is necessary and sufficient to prove propositions (A12) and (A11) and the validity of Equation A10. As $\ln(0.05)/\ln(0.95) \simeq 58.4$, Equation A10 is valid when $t_1/t_2 > 58.4$. Considering that m and μ are small, and that N is large, this ratio is approximately equal to (from Equation 8)

$$\frac{t_1}{t_2} \simeq \frac{2\mu + 2m(n/[n-1]) + (1/2N) - (1/2N_e)}{2\mu + (1/2N_e)}. \tag{A13}$$

From Equation A13 we can derive the conditions of the timescales separation of the dynamics of genetic diversity (*i.e.*, the parameter values for which $t_1 \gg t_2$). When $n > 14$, differences are the highest ($t_1 \gg t_2$), in the domain where $\mu \ll 1/2N$ and also when $\mu \gg 1/2N$ and $m > \mu$. When $n \leq 14$, the same conditions apply for $t_1 \gg t_2$ except in a restricted domain where $m \simeq 1/2N$ ($m \gg (n-1)/2nN$ or $m \ll (n-1)/2nN$ are required for $t_1 \gg t_2$; see Figure A1). Indeed, denoting $A = \ln(0.05)/\ln(0.95)$, $t_1/t_2 > A$ implies that

$$M > \frac{A + 1 - 2n + (A-1)n\theta + \sqrt{D}}{2(n^2/(n-1))} \quad \text{or} \quad M < \frac{A + 1 - 2n + (A-1)n\theta - \sqrt{D}}{2(n^2/(n-1))}, \tag{A14a}$$

where

$$D = (A+1)(A+1-4n) + 2n\theta\left(A^2 - 1\right) + n^2(A-1)^2\theta^2. \tag{A14b}$$

For $\theta \gg 1$, we can neglect terms that do not contain θ, and conditions (A14) simplify to

$$M > (A-1)\frac{n-1}{n}\theta, \tag{A15}$$

which simplifies to $M \gg \theta$.

For $\theta \ll 1$, terms that contain θ can be neglected in Equation A14, which yields the following conditions:

$$M > \frac{A + 1 - 2n + \sqrt{(A+1)(A+1-4n)}}{2(n^2/(n-1))} \quad \text{or} \quad M < \frac{A + 1 - 2n - \sqrt{(A+1)(A+1-4n)}}{2(n^2/(n-1))} \tag{A16}$$

Therefore, Equation A10 is not valid when conditions of Equation A16 are not met. This domain is centered around $M = ((n-1)/n)$, as

$$\frac{A + 1 - 2n + \sqrt{(A+1)(A+1-4n)}}{2(n^2/(n-1))} \quad \text{and} \quad \frac{A + 1 - 2n + \sqrt{(A+1)(A+1-4n)}}{2(n^2/(n-1))}$$

equalize to $M = ((n-1)/n)$, for $A = 4n - 1$. The size of this domain decreases when n increases (Figure A1), and Equation A10 is valid for any value of M in the domain $\theta \ll 1$ when $n > 14$ (as conditions (A16) are relaxed when $4n > A + 1$, with $A = (\ln(0.05)/\ln(0.95))$).

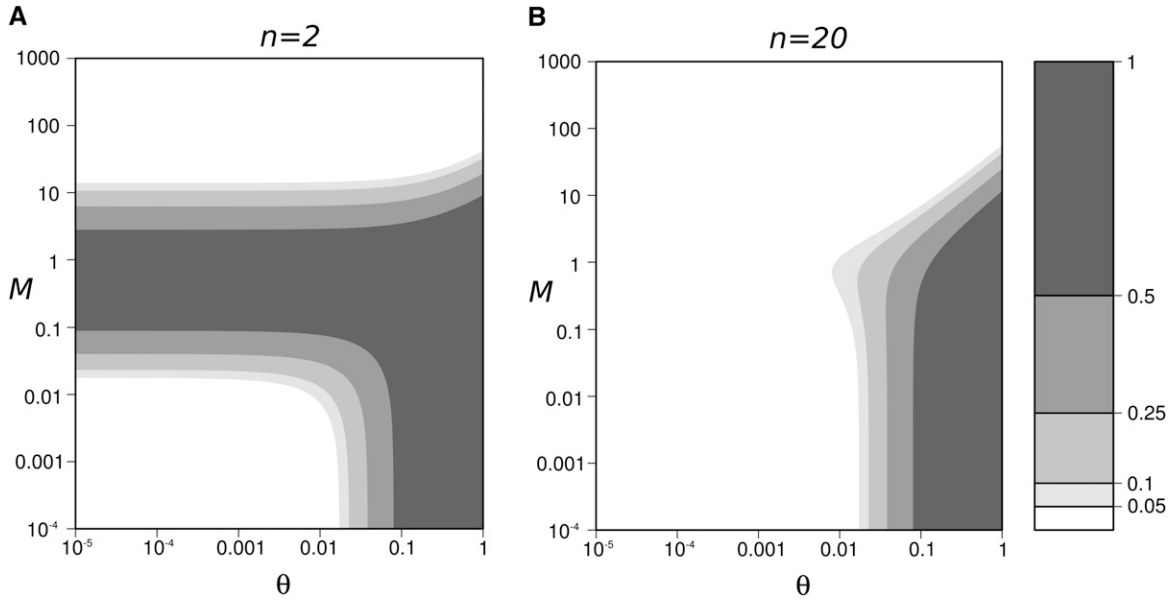

Figure A1 Domain of validity of Equation A10 (white contours), as a function of the strength of migration (M) and mutation (θ), (A) $n = 2$, (B) $n = 20$. The dark shading represent the domains where the validity of Equation A10 is poor (*i.e.*, the exact value of $\lambda_2^{\ln(0.95)/\ln(\lambda_1)} > 0.05$; from Equation 5).

Appendix B: Peak of Genetic Diversity Generated by a Reconnection Event

We derive first the value of the peak of within-population genetic diversity, ΔH_s, and the transient excess of between-population genetic diversity, ΔH_b, generated by a reconnection event. Second, we characterize their dependency on the migration rate, mutation rate, size, and number of populations.

Derivation of ΔH

We denote the vector of genetic diversity excess at the end of the transient dynamics phase as $\mathbf{\Delta H} = \begin{pmatrix} \Delta H_s \\ \Delta H_b \end{pmatrix}$.

The genetic diversity at the end of the transient dynamics phase is approximately $\mathbf{H}_{t_2} = 1 - (\mathbf{F}^{eq} + \mathbf{C}_1\lambda_1^{t_2})$ (from Equation A9 when $t \simeq t_2$). Thus $\mathbf{\Delta H}$ is approximately

$$\begin{aligned} \mathbf{\Delta H} &= \mathbf{H}_{t_2} - \mathbf{H}^{eq} \\ &= 1 - \left(\mathbf{F}^{eq} + \mathbf{C}_1\lambda_1^{t_2}\right) - (1 - \mathbf{F}^{eq}) \\ &= -\mathbf{C}_1\lambda_1^{t_2}. \end{aligned} \tag{B1}$$

The peak of genetic diversity depends on $\mathbf{C}_1 = y_{10}\mathbf{U}_1$ (from Equations A2 and A3). To derive the value of $\mathbf{C}_1$, we compute $\mathbf{U}_1$ and y_{10}. The eigenvector $\mathbf{U}_1 = \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix}$ is associated with λ_1 and follows

$$\mathbf{A}\mathbf{U}_1 = \lambda_1\mathbf{U}_1. \tag{B2}$$

Equation B2 leads to the condition that $u_{21} = -((a(1-c) - (1-b) - \sqrt{\Delta_{red}})/2(1-a))u_{11}$, where $\Delta_{red} = (1 - a(1-c) + b)^2 - 4bc$. We set $u_{11} = 2(1-a)$, leading to

$$\mathbf{U}_1 = \begin{pmatrix} 2(1-a) \\ 1 - a(1-c) - b + \sqrt{(1-a(1-c)+b)^2 - 4bc} \end{pmatrix}.$$

Similarly, we can determine the eigenvector associated with λ_2:

$$\mathbf{U}_2 = \begin{pmatrix} 2(1-a) \\ 1 - a(1-c) - b - \sqrt{(1-a(1-c)+b)^2 - 4bc} \end{pmatrix}.$$

Denoting the initial value of genetic identity $\mathbf{F}_0 = \begin{pmatrix} F_{s,0} \\ F_{b,0} \end{pmatrix}$ and replacing the expression of $\mathbf{U}_1$ and $\mathbf{U}_2$ in Equation A3 leads to

$$\mathbf{Y}_0 = \begin{pmatrix} -\dfrac{\left(1-a(1-c)-b-\sqrt{\Delta_{\text{red}}}\right)}{4(1-a)\sqrt{\Delta_{\text{red}}}} & \dfrac{1}{2\sqrt{\Delta_{\text{red}}}} \\ \dfrac{1-a(1-c)-b+\sqrt{\Delta_{\text{red}}}}{4(1-a)\sqrt{\Delta_{\text{red}}}} & -\dfrac{1}{2\sqrt{\Delta_{\text{red}}}} \end{pmatrix} \begin{pmatrix} F_{\text{s},0}-F^{\text{eq}}_{\text{s,con}} \\ F_{\text{b},0}-F^{\text{eq}}_{\text{b,con}} \end{pmatrix}.$$

Assuming isolation equilibrium for the initial identity leads to $F_{\text{s},0} = F^{\text{eq}}_{\text{s,iso}}$ and $F_{\text{b},0} = 0$; thus, $\mathbf{\Delta H}$ simplifies to

$$\mathbf{\Delta H} = \left[\left(F^{\text{eq}}_{\text{s,iso}} - F^{\text{eq}}_{\text{s,con}}\right)\frac{1-a(1-c)-b-\sqrt{\Delta_{\text{red}}}}{4(1-a)\sqrt{\Delta_{\text{red}}}} + \frac{F^{\text{eq}}_{\text{b,con}}}{2\sqrt{\Delta_{\text{red}}}}\right]\mathbf{U}_1\lambda_1^{t_2}. \tag{B3}$$

With $4bc \ll (1 - a(1 - c) + b)^2$ (as $m \ll 1$ and $N \gg 1$), and given that $\sqrt{1-x} = 1 - \frac{1}{2}x + o(x)$, $\mathbf{\Delta H}$ further becomes

$$\mathbf{\Delta H} = -\frac{1}{2(1-a(1-c)+b-2bc/(1-a(1-c)+b))}\left[\frac{\left(F^{\text{eq}}_{\text{s,iso}} - F^{\text{eq}}_{\text{s,con}}\right)}{n-1}\frac{1-a(1-c)+b-c}{1-a(1-c)+b} - F^{\text{eq}}_{\text{b,con}}\right]\mathbf{U}_1\lambda_1^{t_2}. \tag{B4}$$

Assuming that migration and mutation rates are always small, we can neglect terms in m^2, μ^2, $1/N^2$, $m\mu$, μ/N, and m/N, and Equation B4 simplifies to

$$\mathbf{\Delta H} = -\left[\frac{\left(F^{\text{eq}}_{\text{s,iso}} - F^{\text{eq}}_{\text{s,con}}\right)}{n-1}\frac{M[n/(n-1)]}{1+M[n/(n-1)]} - F^{\text{eq}}_{\text{b,con}}\right]\begin{pmatrix} \dfrac{M}{1+[n/(n-1)]M-2N/N_{\text{e}}} \\ \dfrac{1+M-N/N_{\text{e}}}{1+[n/(n-1)]M-2N/N_{\text{e}}} \end{pmatrix}\lambda_1^{t_2}. \tag{B5}$$

By replacing the term $\lambda_1^{t_2}$ (Equation 8 with $\alpha = 0.05$) in Equation B5, we obtain Equation 9.

Equation 9 provides a good approximation of the size of the peak of within-population genetic diversity ΔH_{s} and of the transient excess of between-population genetic diversity ΔH_{b} in the entire parameter domain. Figure B1, A and C, and Figure B2, A and C, represent the exact (solid line) and approximate (dashed line; from Equation 9) values of ΔH_{s} and ΔH_{b}, respectively, as a function of θ and M, for $n = 2$ and $n = 20$; we can see that the true and approximate values are very close. Figure B1, B and D, and Figure B2, B and D, represent the absolute error resulting from the use of Equation 9 as an approximation of ΔH_{s} and ΔH_{b}, respectively, instead of its exact value, for $n = 2$ and $n = 20$. Discrepancies between Equation 9 and the true values of ΔH_{s} and ΔH_{b} can first come from the assumption that $m \ll 1$, $\mu \ll 1$, and $N \gg 1$ and second from the assumption of the existence of t_2 such that $|C_2\lambda_2^{t_2}| \ll |C_1\lambda_1^{t_2}|$.

Maximum peak of diversity after a connection event

The peak of genetic diversity increases monotonously with M ($d(\mathbf{\Delta H})/dM > 0$ for any value of m, μ, n, and N) and decreases monotonously with θ ($d(\mathbf{\Delta H})/dM < 0$ for any value of m, μ, n, and N). The peak of diversity is maximized for intermediate values of n and reached when

$$\frac{d(\mathbf{\Delta H})}{dn} = 0. \tag{B6}$$

If we neglect terms in μ^2, m^2, $1/N^2$, μm, μ/N, and m/N, and assuming small θ and high M, solving Equation B6 using Equation 10 for the peak of genetic diversity leads to Equation 11. When $n = n^*$, Equation 10 leads to a peak of genetic diversity of

$$\Delta H^{\max}|_{n^*} = \frac{1}{1+2\sqrt{\theta}}.$$

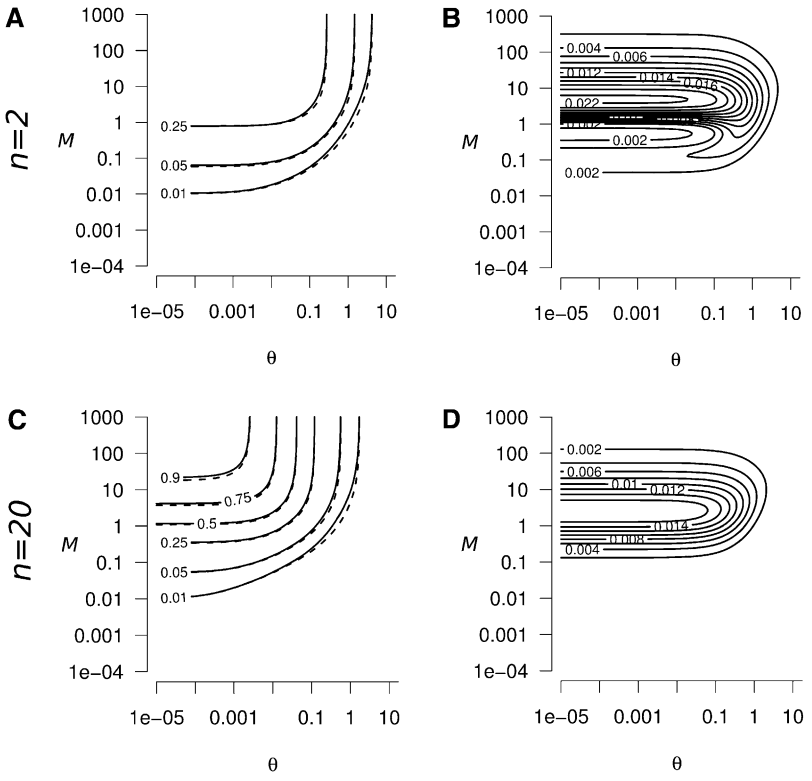

Figure B1 (A and C) Exact (solid lines) and approximate (dashed lines, from Equation 9) values of the peak of genetic diversity ΔH_s, as a function of the strength of migration (M) and mutation (θ). In both A and C, the exact and approximate values of ΔH_s are very close. (B and D) Absolute error when using Equation 9 to approximate ΔH_s, as a function of M and θ. The maximum absolute error is reached when $M \simeq 5$ and $\theta < 1$ in both B and D. The error decreases when $M \gg 5$ or $M \ll 5$. The absolute error increases when n decreases, but remains weak: (B) the maximum absolute error is 0.025 for $n = 2$, and (D) 0.018 for $n = 20$. Consequently, Equation 9 is a good approximation for the peak of genetic diversity whatever the parameter values of θ, M, and n considered.

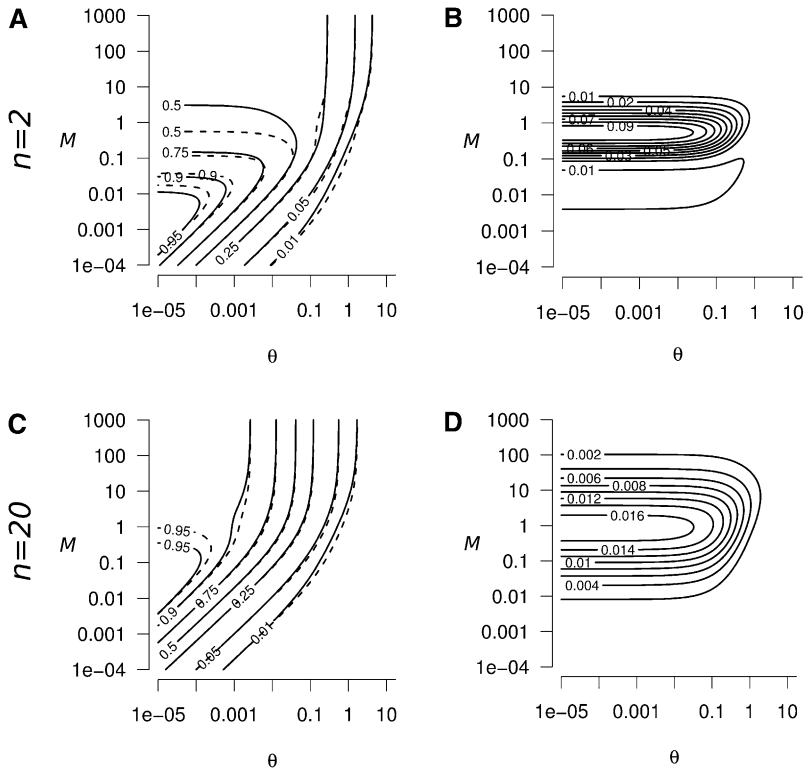

Figure B2 (A and C) Exact (solid lines) and approximate (dashed lines, from Equation 9) values of the transient excess of between-population genetic diversity ΔH_b, as a function of the strength of migration (M) and mutation (θ). In both A and C, the exact and approximate values of ΔH_b are very close. (B and D) Absolute error when using Equation A9 to approximate ΔH_b, as a function of M and θ. The maximum absolute error is reached when $M \simeq 1$ and $\theta < 1$, the absolute error is 0.09 for $n = 2$ (B), and 0.016 for $n = 20$ (D). Equation 9 is a good approximation for ΔH_b whatever the parameter values of θ, M, and n considered.

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.147785/-/DC1

Peak and Persistent Excess of Genetic Diversity Following an Abrupt Migration Increase

Nicolas Alcala, Daniela Streit, Jérôme Goudet, and Séverine Vuilleumier

DOI: 10.1534/genetics.112.147785

FILE S1

GENETIC DIVERSITY EQUILIBRIUM

In this Supporting Information File, we provide for eq. 3 the conditions of existence of a genetic diversity equilibrium value, we derive its value and finally determine its stability.

Condition of existence of an equilibrium: Eq. 3 is a non-homogeneous linear matrix recurrence equation, with an additive constant vector $\mathbf{B}$. If $\mathbf{I}-\mathbf{A}$ is invertible, there is a unique equilibrium value for eq. 3:

$$\mathbf{F}^{eq} = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} \tag{S1.1}$$

and from any initial value $\mathbf{F}_0$, genetic identities after t generations follow:

$$\mathbf{F}_t = \mathbf{A}^t(\mathbf{F}_0 - \mathbf{F}^{eq}) + \mathbf{F}^{eq} \tag{S1.2}$$

$\mathbf{I}-\mathbf{A}$ is invertible if and only if the determinant of $\mathbf{I}-\mathbf{A}$ is not null, which is true whenever $\mu \neq 0$. This condition is always met in our model.

Equilibrium value: We can express the equilibrium value S1.1 as a function of the migration rate m, mutation rate μ, population size N and number of populations n:

$$\begin{aligned}
\mathbf{F}^{eq} &= (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} \\
&= \begin{pmatrix} 1-a(1-c)(1-\mu)^2 & -(1-a)(1-\mu)^2 \\ -b(1-c)(1-\mu)^2 & 1-(1-b)(1-\mu)^2 \end{pmatrix}^{-1} (1-\mu)^2 \begin{pmatrix} ac \\ bc \end{pmatrix} \\
&= \frac{1}{ac(\frac{1}{(1-\mu)^2}-1)+bc+(\frac{1}{(1-\mu)^2}-1)(\frac{1}{(1-\mu)^2}+b-a)} \begin{pmatrix} ac(\frac{1}{(1-\mu)^2}-1)+bc \\ \frac{bc}{(1-\mu)^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{1+\frac{(2-\mu)}{c(1-\mu)^2}\mu\frac{\frac{(2-\mu)}{(1-\mu)^2}\mu+nb}{a\frac{(2-\mu)}{(1-\mu)^2}\mu+b}} \\ \frac{b}{b(1-\mu)^2+(2-\mu)\mu(a+c^{-1}(\frac{(2-\mu)}{(1-\mu)^2}\mu+nb))} \end{pmatrix}
\end{aligned}$$

When mutation rates and migration rates are small, $2-\mu$ is close to 2, $(1-\mu)^2$ is close to 1, $a = (1-m)^2 + \frac{m^2}{n-1}$ is close to $1-2m$, and $b = \frac{1-a}{n-1}$ is close to $\frac{2m}{n-1}$. Therefore, taking into account that $c = \frac{1}{2N}$, the expression above further simplifies to:

$$\mathbf{F}^{eq} = \begin{pmatrix} \frac{1}{1+4N\mu\frac{2\mu+2m\frac{n}{n-1}}{2\mu+\frac{2m}{n-1}}} \\ \frac{\frac{2m}{n-1}}{\frac{2m}{n-1}+2\mu(1-2m+2N(2\mu+2m\frac{n}{n-1}))} \end{pmatrix}$$

Denoting $\theta = 4N\mu$ and $M = 4Nm$, for the scaled mutation and migration rate, respectively, the above expression can be described as:

$$\mathbf{F}^{eq} = \begin{pmatrix} \frac{1}{1+\theta(1+\frac{M}{\theta+\frac{M}{n-1}})} \\ \frac{M}{M+(n-1)\theta(1+\theta+\frac{n}{n-1}M)} \end{pmatrix} \tag{S1.3}$$

which is the equilibrium value that was first derived by MARUYAMA (1970). We can then derive the genetic diversities $\mathbf{H}_t = \begin{pmatrix} H_{s,t} \\ H_{b,t} \end{pmatrix} = \mathbf{1}-\mathbf{F}_t$, and the equilibrium genetic diversities $\mathbf{H}^{eq} = \mathbf{1}-\mathbf{F}^{eq}$.

Stability of the equilibrium: Equilibrium S1.1 is stable if and only if $\lim_{t\to\infty} \mathbf{A}^t = 0$, thus if and only if the absolute value of all eigenvalues of matrix $\mathbf{A}$ are below 1.

The eigenvalues of matrix A are the roots of the characteristic equation $\chi_{\mathbf{A}}(T) = T^2 - tr(\mathbf{A})T + det(\mathbf{A})$. They are presented in eq. 5. As $4bc \geq 0$ for all values of m, n and N, $\sqrt{(1-a(1-c)+b)^2-4bc} \leq 1-a(1-c)+b$. Thus from eq. 5, $\lambda_1 \leq (1-\mu)^2$. λ_1 is below 1 for all mutation rates strictly less than 1. As $\lambda_1 < 1$ and $\lambda_2 < \lambda_1$, both eigenvalues are below 1. In addition, λ_1 and λ_2 are positive for any possible migration rate, mutation rate, size and number of populations. Therefore $0 < \lambda_1 < 1$ and $0 < \lambda_2 < 1$, which proves that equilibrium S1.1 is always stable, and is reached whatever the initial identity $\mathbf{F}_0$ considered.

FILE S2

EFFECT OF AN ISOLATION EVENT ON GENETIC DIVERSITY

We show in this section that following an isolation event, within- and between-population genetic diversities change independently and that their times to reach equilibrium values are t_2 and t_1, respectively. When populations are isolated ($m = 0$), eq. 2 simplifies to:

$$\begin{cases} F_{s,t+1} = (c + (1-c)F_{s,t})(1-\mu)^2 \\ F_{b,t+1} = F_{b,t}(1-\mu)^2 \end{cases} \tag{S2.1}$$

In eq. S2.1, $F_{s,t+1}$ depends only on $F_{s,t}$ but not on $F_{b,t}$; similarly, $F_{b,t+1}$ depends only on $F_{b,t}$ but not on $F_{s,t}$. Thus, $F_{s,t}$ and $F_{b,t}$ both follow a one dimensional recurrence equation:

$$\begin{cases} F_{s,t} = F_s^{eq} + (F_{s,0} - F_s^{eq})((1-c)(1-\mu)^2)^t \\ F_{b,t} = F_{b,0}(1-\mu)^{2t} \end{cases} \tag{S2.2}$$

Therefore, when populations are isolated, within- and between-population genetic diversities change according to $(1-c)(1-\mu)^2$ and $(1-\mu)^2$, respectively. As when m=0, λ_1=$(1-\mu)^2$ and λ_2=$(1-c)(1-\mu)^2$ (from eq. 5), we can rewrite S2.2 as:

$$\begin{cases} F_{s,t} = F_s^{eq} + (F_{s,0} - F_s^{eq})\lambda_2^t \\ F_{b,t} = F_{b,0}\lambda_1^t \end{cases} \tag{S2.3}$$

We can conclude that when populations are isolated, F_s and F_b change according to λ_2^t and λ_1^t, respectively. This demonstrates that after isolation, within- and between-population genetic diversities reach their equilibrium value in t_2 and t_1 generations, respectively.

FILE S3

RELAXING THE COMPLETE ISOLATION HYPOTHESIS

Throughout the study, we consider a complete isolation event, where populations that were previously connected suddenly become completely isolated, and a reconnection event where previously completely isolated populations suddenly become connected. Nevertheless, we show in this Supporting Information File that the assumption of complete isolation can be relaxed. Indeed, the results for complete isolation are a very good approximation of results for a strong but incomplete isolation.

Equilibrium genetic identity under incomplete isolation: Incomplete isolation corresponds to a state where populations are connected through migration at a rate $0 < m << 1$. We characterize in this section the threshold value of migration rate under which the complete isolation equilibrium genetic identity is a good approximation of the incomplete isolation equilibrium value.

To do so, we can rewrite eq. S1.3 under the following form (isolating terms in M):

$$\mathbf{F}^{eq} = \begin{pmatrix} \frac{\theta+\frac{M}{n-1}}{\theta+\frac{M}{n-1}+\theta(\theta+\frac{M}{n-1}+M)} \\ \frac{M}{M+(n-1)\theta(1+\theta)+n\theta M)} \end{pmatrix} = \begin{pmatrix} \frac{1}{1+\theta} + (\frac{1}{1+n\theta} - \frac{1}{1+\theta})\frac{M}{M+(n-1)\theta\frac{1+\theta}{1+n\theta}} \\ (\frac{1}{1+n\theta})\frac{M}{M+(n-1)\theta\frac{1+\theta}{1+n\theta}} \end{pmatrix}$$

We know that an isolated population of size N has a within-population genetic identity equilibrium value of $\frac{1}{1+\theta}$ (KIMURA and CROW 1964), and a between-population equilibrium value of 0. Therefore, the genetic identity of isolated populations at equilibrium is $\mathbf{F}^{iso} = \begin{pmatrix} \frac{1}{1+\theta} \\ 0 \end{pmatrix}$. The equilibrium within- and between-population genetic identity of a panmictic population of size nN

is $\mathbf{F}^{pan} = \begin{pmatrix} \frac{1}{1+n\theta} \\ \frac{1}{1+n\theta} \end{pmatrix}$ (KIMURA and CROW 1964). Eq. S1.3 can be written:

$$\mathbf{F}^{eq} = \mathbf{F}^{iso} + (\mathbf{F}^{pan} - \mathbf{F}^{iso}) f_{n;\theta}(M) \tag{S3.1a}$$

with

$$f_{n;\theta}(M) = \frac{M}{M + (n-1)\theta\frac{1+\theta}{1+n\theta}} \tag{S3.1b}$$

$f_{n;\theta}(M)$ is similar to a Michaelis-Menten function (in the form $f(M) = vM/(M + M_T)$), with a maximum value v=1 and a threshold M_T=$(n-1)\theta\frac{1+\theta}{1+n\theta}$ (MICHAELIS and MENTEN 1913). We can predict by analogy to the Michaelis-Menten function that the behavior of the function depends on the relative value of M and M_T, with in our case a threshold value of:

$$M_T = (n-1)\theta\frac{1+\theta}{1+n\theta} \tag{S3.2}$$

If M is much below the threshold, $\mathbf{F}^{eq} \simeq \mathbf{F}^{iso}$: the genetic identity equilibrium is close to the isolation equilibrium. If M is larger than the threshold, $\mathbf{F}^{eq} \simeq \mathbf{F}^{pan}$: the genetic identity equilibrium is close to the panmictic equilibrium. When M=M_T, $\mathbf{F}^{eq}$=$\frac{\mathbf{F}^{iso}+\mathbf{F}^{pan}}{2}$: the genetic identity equilibrium is the mean of the panmictic and isolation equilibria. Fig. S3 illustrates the variations of $\mathbf{F}_s^{eq}$ and $\mathbf{F}_b^{eq}$ as a function of M.

Peak of genetic identity generated by a migration rate increase: We now consider an event of abrupt migration rate increase (from scaled rate M_0 to $M > M_0$). In this case, eq. B.5 becomes:

$$\mathbf{\Delta H} = -[\frac{(F_{s,iso}^{eq} - F_s^{pan})}{n-1}\frac{M\frac{n}{n-1}}{1+M\frac{n}{n-1}} - F_b^{pan}](f(M)-f(M_0)) \begin{pmatrix} \frac{M}{1+\frac{n}{n-1}M-\frac{2N}{Ne}} \\ \frac{1+M-\frac{N}{Ne}}{1+\frac{n}{n-1}M-\frac{2N}{Ne}} \end{pmatrix} 0.05^{\frac{t_2}{t_1}} \tag{S3.3}$$

The value of the peak of diversity depends on the relative value of M_0 and the threshold M_T.

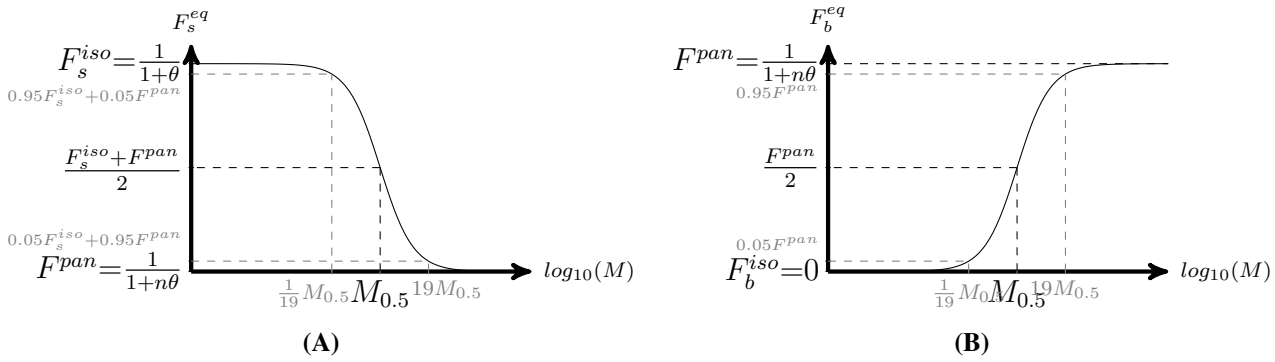

Figure S1 (A) Within-population F_s^{eq}, and (B) between-population F_b^{eq} genetic identity, as a function of the strength of migration M (scaled migration rate). F^{iso} and F^{pan} are the expected equilibria when all populations are isolated or panmictic, respectively. The inflexion point (transition between panmictic and isolation equilibrium) is $M_T{=}(n-1)\theta\frac{1+\theta}{1+n\theta}$ for both within- and between-population genetic identity. For $M < M_{0.05}{=}\frac{0.05}{0.95}M_T$ (M_T is 19 fold $M_{0.05}$), the genetic identity equilibrium is close to the isolation equilibrium.

Indeed, from eq. S3.1b and S3.3 we can see that $f(M_0){\simeq}0$ and eq. S3.3 and B.5 equalize when $M_0{<<}M_T$. Thus, an increase in migration rate above the threshold produces the same peak as a reconnection event. This demonstrates that genetic diversity peaks are generated whenever the scaled migration rate increases abruptly and crosses the threshold value $\mathbf{M_T}$. Thus, an increase in migration rate from a small M_0 to M produces approximately the same peak of genetic diversity as an increase from 0 to M ($M_0 < \frac{0.05}{0.95}M_T$ ensures that $f(M_0){<}0.05$).

LITERATURE CITED

KIMURA, M., and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

MARUYAMA, T., 1970 Effective number of alleles in a subdivided population. Theor Popul Biol **1**: 273–306.

MICHAELIS, L., and M. L. MENTEN, 1913 The kenetics of the inversion effect. Biochemische Zeitschrift **49**: 333–369.