Supplementary Information for Manuscript "Turnover and accumulation of genetic diversity across large time-scale cycles of population isolation and connection"

Nicolas Alcala[1*], **Séverine Vuilleumier**[1,2]

[1] *Department of Ecology and Evolution, University of Lausanne, Biophore CH-1015 Lausanne, Switzerland*

[2] *Institute of Microbiology, Lausanne University Hospital and University of Lausanne, CH-1011 Lausanne, Switzerland*

$*$ *corresponding author: Nicolas.Alcala@unil.ch*

Supplementary Material S1: Solution to the equations describing the genetic diversities within a period of isolation or connection

In this section, we derive the solution to the recurrence equations describing the dynamics of within- and between-population genetic diversities using the equations describing corresponding identities from [1] (p. 64) .

From [1], within- and between-population genetic identities at generation $P+1$, $f_s(P+1)$ and $f_b(P+1)$, follow:

$$\begin{cases} f_s(P+1) = (1-\mu)^2[a(c+(1-c)f_s(P)) + (1-a)f_b(P)] \\ f_b(P+1) = (1-\mu)^2[b(c+(1-c)f_s(P)) + (1-b)f_b(P)] \end{cases} \tag{S1.1}$$

where a, b and c are described in the main text (equations 3b-d). When $m > 0$ (resp. $m = 0$), equations S1.1 describe the dynamics of genetic identities during a connection period (resp. isolation period).

From S1.1, we can write the equations describing the dynamics of genetic diversities during isolation and connection periods, using the relationship $h = 1-f$, $h_s(P) = 1-f_s(P)$ and $h_b(P) =$

$1 - f_b(P)$, and denoting $\mathbf{H}_c(P)$ and $\mathbf{H}_i(P)$ the genetic diversities vector $\mathbf{H}(P) = \begin{pmatrix} h_s(P) \\ h_b(P) \end{pmatrix}$ during connection ($m > 0$) and isolation periods ($m = 0$), respectively. We obtain:

$$\mathbf{H}_c(P+1) = \mathbf{A}_c\mathbf{H}_c(P) + \mathbf{B}_c \tag{S1.2a}$$

$$\mathbf{H}_i(P+1) = \mathbf{A}_i\mathbf{H}_i(P) + \mathbf{B}_i \tag{S1.2b}$$

where $\mathbf{B}_c = (1-\mu)^2 \begin{pmatrix} 1-ac \\ 1-bc \end{pmatrix}$, $\mathbf{B}_i = (1-\mu)^2 \begin{pmatrix} 1-c \\ 1 \end{pmatrix}$ and $\mathbf{A}_c$ and $\mathbf{A}_i$ are defined in the main text (equations 3a and 4).

Equations S1.2a and S1.2b are both inhomogenous matrix difference equations. As matrix $(I - \mathbf{A}_c)$ (resp. $(I - \mathbf{A}_i)$) is invertible, equation S1.2a (resp. S1.2b) has an equilibrium value, $\widehat{\mathbf{H}}_c{=}(I - \mathbf{A}_c)^{-1}\mathbf{B}_c$ (resp. $\widehat{\mathbf{H}}_c{=}(I - \mathbf{A}_c)^{-1}\mathbf{B}_c$). Also, equations S1.2a and S1.2b each have a solution, which are presented in the main text equation 2.

Supplementary Material S2: The dynamics of genetic diversities across cycles under the panmictic connection periods approximation

In this section, we derive dynamics of the genetic diversities under cycles of connection and isolation in the case where populations are panmictic during the connection periods (i.e., $m = \frac{n-1}{n}$). We demonstrate that genetic diversity changes monotonically across cycles; in addition, we show that the relaxation time of genetic diversity when periods are short correspond to that under constant migration.

In this case, matrix $\mathbf{A}_c$ simplifies to:

$$\mathbf{A}_c = (1-\mu)^2 \begin{pmatrix} \frac{1}{n}(1-c) & \frac{n-1}{n} \\ \frac{1}{n}(1-c) & \frac{n-1}{n} \end{pmatrix} \tag{S2.1}$$

and its eigenvalues simplify to:

$$\begin{aligned} \lambda_1 &= (1-\mu)^2(1-c') \\ \lambda_2 &= 0 \end{aligned} \tag{S2.2}$$

where $c' = 1/2nN$ is the rate of genetic drift in a panmictic population of size nN.

Consequently,

$$\mathbf{\Gamma}_c = (1-\mu)^{4P}(1-c')^{P-1} \begin{pmatrix} \frac{1}{n}(1-c)^{P+1} & \frac{n-1}{n} \\ \frac{1}{n}(1-c)^{P+1} & \frac{n-1}{n} \end{pmatrix} \tag{S2.3}$$

with a first eigenvalue $\lambda_c = (1-\mu)^{4P}(1-c')^{P-1}(\frac{1}{n}(1-c)^{P+1}+\frac{n-1}{n})$, and a second eigenvalue which is null. Therefore, the changes of genetic diversity across cycles are monotonic, and in particular, the successive peaks of genetic diversity generated by connection events have a monotonically changing size.

So we have, for $k \geq 1$:

$$h_{c,s}^{(k)} = h_{c,b}^{(k)} = \lambda_c^k(\frac{\frac{1}{n}(1-c)^{P+1}h_s^{(0)} + \frac{n-1}{n}h_b^{(0)}}{\frac{1}{n}(1-c)^{P+1} + \frac{n-1}{n}} - h_c^*) + h_c^* \tag{S2.4}$$

where h_c^* is the equilibrium value of the cycles both within- and between-population.

Interestingly, when P tends to 0 the relaxation time of $h_{c,s}$ and h_c, b, $log(\delta)/log(\lambda_c)$, tends to the expected relaxation time during connection, P_C (equation 10). This demonstrates that when periods are short, the relaxation time is approximately the same as under constant migration.

Supplementary Material S3: Dynamics of genetic diversity under stochastic cycles of isolation and connection

We showed that values of P, P_W and P_I determine the behavior of genetic diversity under cycles of isolation and connection. In this section, we consider that each period P is a random variable. We study the two following quantities: the probabilities that the isolation period is shorter than P_W, $\mathbb{P}(P < P_W)$, and the probability that it is longer than P_I, $\mathbb{P}(P > P_I)$. Two scenarios are considered. Scenario **A** assumes that the probability of an event of isolation (resp. connection) is the same for each generation (i.e. independent of the generation t) but follows a geometric distribution. Scenario **B** assumes that the length of isolation and connection periods are regular but have a Gaussian noise, generating variance around the mean period P (i.e. dependent of the generation t). Thus, Scenario A considers time-homogeneous stochastic changes and Scenario B

considers time-inhomogeneous stochastic changes (following [2]).

Dynamics of genetic diversity under scenario A

Under scenario A, we assume that the probability p to switch from isolation state to connection state at a given generation is independent of the current generation t. Thus, the sequence of isolation and connection events is modelled as a two states time-homogeneous Markov process. Under such a scenario, the duration of each period (corresponding to the waiting time until state switch), P, follows a geometric distribution of parameter p, p being the probability of the occurrence of the isolation or connection event (so the mean period is $\bar{P} = 1/p$). Thus we have:

$$\begin{cases} \mathbb{P}(P < P_W) = 1 - (1-p)^{P_W} \\ \mathbb{P}(P > P_I) = (1-p)^{P_I} \end{cases} \tag{S3.1}$$

Which yields

$$\begin{cases} \mathbb{P}(P < P_W) > 1 - \epsilon_W \Leftrightarrow p > 1 - \epsilon_W^{1/P_W} \\ \mathbb{P}(P > P_I) > 1 - \epsilon_I \Leftrightarrow p < 1 - (1-\epsilon_I)^{1/P_I} \end{cases} \tag{S3.2}$$

Where ϵ_W and ϵ_I correspond to the probability that a random period P is larger than P_W and lower than P_I, respectively. Values of ϵ_W and ϵ_I close to 0 lead to a behavior of genetic diversities that follows what is expected under the short-period and long-period domains, respectively.

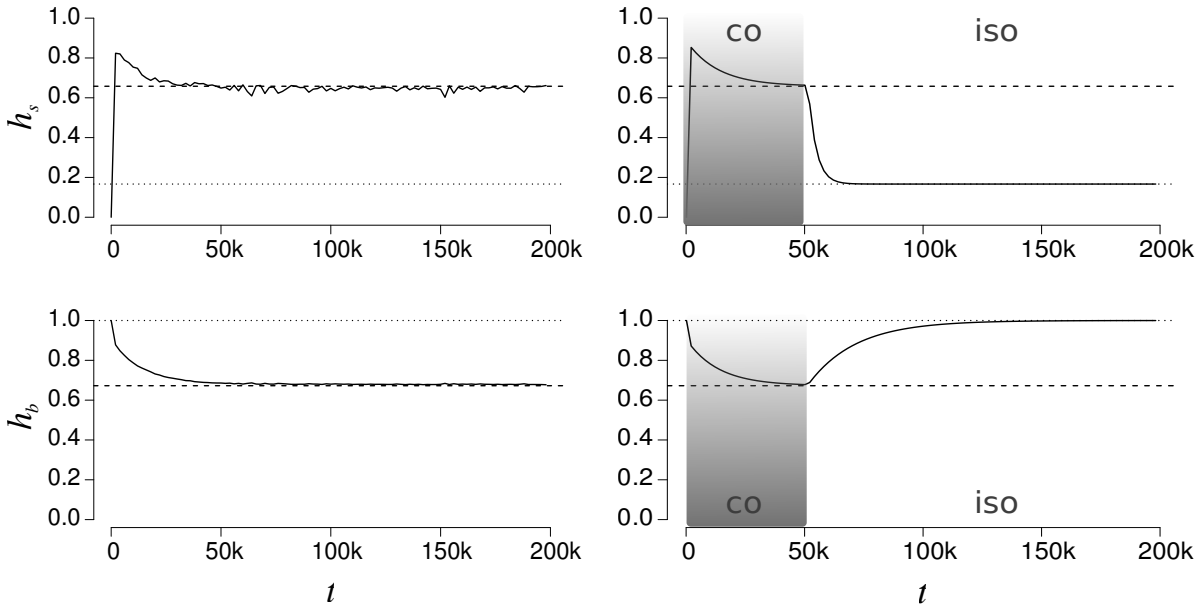
Using the expression of P_W and P_I from equations 10 and 11, we obtain the following approximation for conditions (S3.2):

$$\begin{cases} \mathbb{P}(P < P_W) > 1 - \epsilon_W \Leftrightarrow p > (1/2N + 2\mu)\dfrac{log(\epsilon_W)}{log(1-\alpha)} \\ \mathbb{P}(P > P_I) > 1 - \epsilon_I \Leftrightarrow p < 2\mu\dfrac{log(1-\epsilon_I)}{log(\alpha)} \end{cases} \tag{S3.3}$$

Where α is a value that determines the difference between genetic diversity during the isolation period and the expected genetic diversity in equilibrium isolated populations (by default, we use $\alpha = 0.05$).

From equation S3.3, it is interesting to see that the values of probability p which determine the shape of the equilibrium trajectory are approximately linear functions of $1/2N$ and μ. Thus,

as population size, N, increases, the probability that the trajectories of genetic diversity belong to the short-period domain increases. Similarly, as mutation decreases, μ, the probability that the trajectories of genetic diversity belong to the short-period domain increase and the probability that they belong to the large-period domain decreases. Numerical simulations confirm that when conditions from equation S3.3 are met, under scenario A, the genetic diversity reaches the equilibrium trajectory predicted under deterministic periods of isolation and connection, even though periods are stochastic (see figure S3.1).



(a) $p > 1 - \epsilon_W^{1/P_W}$ **(b)** $p < 1 - (1 - \epsilon_I)^{1/P_I}$

Figure S3.1 Illustration of the impact of stochastic period length (scenario A, geometric distribution of parameter p) on the trajectories of within- (h_s) and between-population (h_b) genetic diversities during cycles of isolation and connection considering (a) short and (b) long expected periods $\mathbb{E}[P] = 1/p$. The dashed and dotted lines represent the expected equilibrium value when populations are connected and isolated, respectively. In (a), $p > 1 - \epsilon_W^{1/P_W}$ (expected short-periods) and both h_s and h_b tend to the connection equilibrium (dashed line). In (b), $p < 1 - (1 - \epsilon_I)^{1/P_I}$ (expected long-periods) and genetic diversities reach their expected equilibrium value at the end of each connection period and isolation period. Parameters are $M = 40$, $n = 10$, $N = 2,000$, $\mu = 2.5 \times 10^{-5}$, ϵ_W=ϵ_I=0.05. (a) $p = 0.02$ ($\mathbb{E}[P] = 50$), (b) $p = 5.10^{-7}$ ($\mathbb{E}[P] = 2.10^6$).

Dynamics of genetic diversity under scenario B

Under scenario B, the period P follows a truncated normal distribution (between 0 and ∞). The period of the fluctuations P (i.e. the waiting time until a switch from one state to another) follows

a discretized normal distribution of mean $\bar{P}$ and variance σ^2:

$$f(P=t) = \begin{cases} \frac{K}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\bar{P})^2}{2\sigma^2}} & \text{if } t > 0 \\ 0 & \text{else} \end{cases} \tag{S3.4}$$

where $K = \sum_{t=0}^{+\infty} f(P{=}t)$ is a normalization constant taking into account the truncation and discretization of the distribution of P.

We can show that this distribution of waiting time corresponds to a time-inhomogeneous Markov process with two states (isolation and connection), and derive the corresponding transition probabilities $p(t)$ as a function of time t. Indeed, starting from a given state (either connection or isolation), the distribution of P is linked to the transition probabilities $p(t)$ through the relation:

$$f(P=t) = (1-p(1))(1-p(2))...(1-p(t-1))p(t) \tag{S3.5}$$

From equation S3.4 and S3.5, we have:

$$p(0) = f(P=0) = \frac{K}{\sigma\sqrt{2\pi}} e^{-\frac{\bar{P}^2}{2\sigma^2}} \tag{S3.6}$$

In addition, from equations S3.4 and S3.5, we have:

$$p(t+1)\frac{1-p(t)}{p(t)} = \frac{f(P=t+1)}{f(P=t)} = e^{\frac{-2(t-\bar{P})-1}{2\sigma^2}} \tag{S3.7}$$

which yields

$$p(t+1) = \frac{p(t)}{1-p(t)} e^{\frac{-2(t-\bar{P})-1}{2\sigma^2}} \tag{S3.8}$$

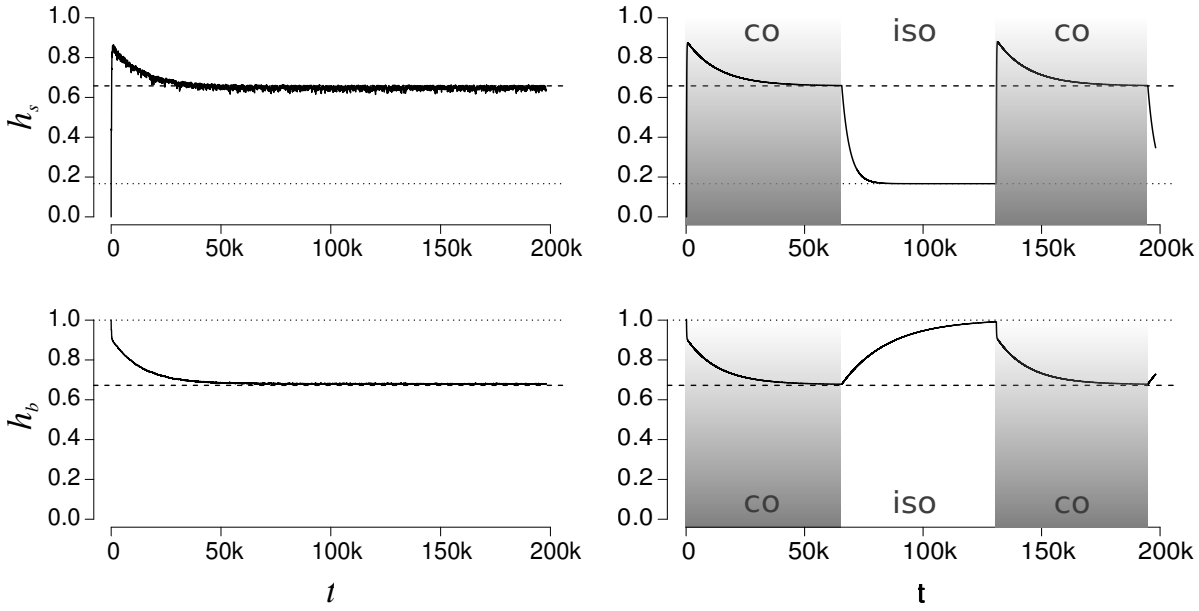
Using equation S3.8 recursively, starting from the expression of $p(0)$ given in equation S3.6, leads to all values of $p(t)$.

Assuming that the probabilities, ϵ_W and ϵ_I, that the period P is lower than P_W and larger than P_I are small, respectively (i.e., $\epsilon_W = \epsilon_I = 0.05$) and using the 5% and 95% quantiles of a normal

distribution of parameters $\overline{P}$ and σ ($\overline{P} - 1.64\sigma$ and $\overline{P} + 1.64\sigma$, respectively) we have:

$$\begin{cases} \mathbb{P}(P < P_W) > 1 - \epsilon_W \Leftrightarrow \overline{P} + 1.64\sigma < P_W \\ \mathbb{P}(P > P_I) > 1 - \epsilon_I \Leftrightarrow \overline{P} - 1.64\sigma > P_I \end{cases} \tag{S3.9}$$

Interestingly, we can see from equation S3.9 that for a given $\overline{P}$, increasing σ decreases the probability to reach the short-period domain and the intermediate-period domain, as then $\overline{P} + 1.64\sigma$ (resp. $\overline{P} - 1.64\sigma$) becomes closer and possibly larger than P_W (resp. smaller than P_I). The gaussian noise does not change the qualitative behavior of genetic diversity through the cycles of isolation and connection (including the equilibrium trajectory) when conditions from equation S3.9 are met (see figure S1).



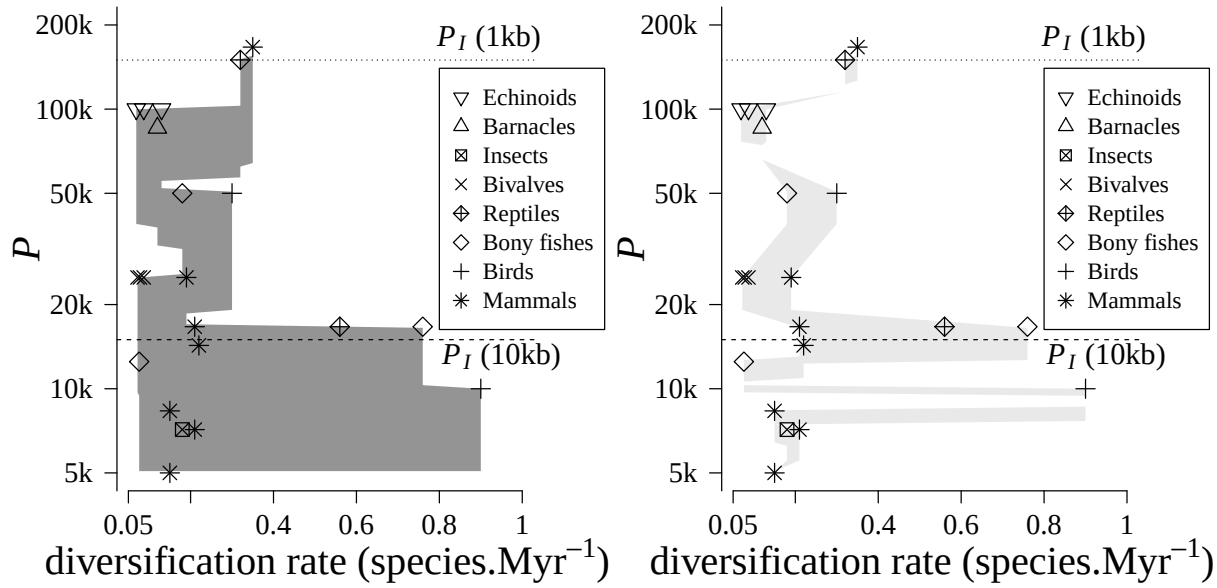
(a) $\overline{P} + 1.64\sigma < P_W$ **(b)** $\overline{P} - 1.64\sigma > P_I$

Figure S3.2 Illustration of the impact of stochastic period length (scenario B, duration following a normal distribution of mean $\overline{P}$ and standard deviation σ) on the within- (h_s) and between-population (h_b) genetic diversities during cycles of isolation and connection. The dashed and dotted lines represent the expected value at equilibrium when populations are connected and isolated, respectively. In (a), $\overline{P} + 1.64\sigma < P_W$ (expected short-periods), both h_s and h_b tend to the connection equilibrium (dashed line). In (b), $\overline{P} - 1.64\sigma > P_I$ (expected long-periods), genetic diversities reach their expected equilibrium value at the end of each connection period and isolation period. Parameters are $M = 40$, $n = 10$, $N = 2,000$, $\mu = 2.5 \times 10^{-5}$. (a) $\overline{P} = 100$, $\sigma = 50$ (b) $\overline{P} = 65,000$, $\sigma = 1,000$.

LITERATURE CITED

[1] Crow, J. F., 1986 *Basic concepts in population, quantitative, and evolutionary genetics.* WH Freeman and Company.

[2] Tuljapurkar, S., 1989 An uncertain life: demography in random environments. *Theoretical population biology* **35**, 227–294.



(a) 0.4 $log_{10}(years)$ window

(b) 0.1 $log_{10}(years)$ window

Figure S1 Impact of the window size on the estimated range of diversification rates (shaded area), as a function of the period length P, for species representative of the main animal orders experiencing environmental cycles of period $100,000$ years. (a) Sliding windows of size 0.4 $log_{10}(years)$; (b) Sliding windows of size 0.1 $log_{10}(years)$. Correlations between P and diversification rate are significant in both (a) and (b) (Spearman's test).