Supporting Information

SI Text

Inference of the Persistence-Time Distribution from a Finite Observational Period. In what follows we provide a complete derivation of the mathematical treatment summarized in *Materials and Methods* of the main text.

When dealing with observational data, the effect of the finiteness of the observed time window on the measured species persistence times must be properly taken into account. In this section we show the analytical derivation of the distribution of two variables that can actually be measured from empirical data: (*i*) the persistence times τ' of species that emerge and go locally extinct within the observed time window ΔT_w and (*ii*) the variable τ'' that comprises the persistence times τ' and all the portions of species persistence times that are partially seen inside the observational time window but start or/and end outside (Fig. S1). By matching analytical and observational distributions for $p_{\tau'}(t)$ and $p_{\tau''}(t)$, it is possible to infer the persistence-time distribution $p_{\tau}(t)$. In the theoretical framework described in the main text, the probability ν of having a diversification event in a time step is constant; thus species emergence in the system due to migration or speciation is a uniform point Poisson process with rate $\lambda = \nu N$, where N is total number of individuals in the system and λ has the dimension of the inverse of a generation time. We term t_0 the emergence time of a species in the system, and T_0 and $T_f = T_0 + \Delta T_w$ the beginning and the end of the observational time window, respectively. A species emerged at time t_0 will be continuously present in a geographic region for its persistence time τ until its local extinction at time $t_0 + \tau$. We first analyze the distribution of τ'', the most complex case; the distribution of τ' will follow easily.

The variable τ'' can be expressed as function of the random variables τ and t_0, which are probabilistically characterized. We can distinguish four different cases (Fig. S1):

1. the species emerges and goes locally extinct within the time window;
2. the species emerges during the observations and it is still present at the end of the time window;
3. the species emerges before the beginning of the observations and goes locally extinct within the time window;
4. the species is always present for all the duration of the observations;

or, mathematically,

$$\tau'' = \begin{cases} \tau, & \text{if } T_0 \leq t_0 \leq T_f \text{ and } t_0 + \tau \leq T_f \\ T_f - t_0, & \text{if } T_0 \leq t_0 \leq T_f \text{ and } t_0 + \tau > T_f \\ t_0 + \tau - T_0, & \text{if } 0 < t_0 < T_0 \text{ and } T_0 \leq t_0 + \tau \leq T_f \\ T_f - T_0, & \text{if } 0 < t_0 < T_0 \text{ and } t_0 + \tau > T \end{cases}.$$

We express the probability of observing τ'' conditional on a persistence time of duration τ as

$$\begin{aligned} p_{\tau''}(t|\tau) = \frac{1}{\mathcal{N}}\Big(& \langle \delta(\tau - t)\Theta(t_0 - T_0)\Theta(T_f - (t_0 + \tau))\Theta(T_f - T_0 - \tau)\rangle \\ & + \langle \delta(T_f - t_0 - t)\Theta(t_0 - T_0)\Theta(T_f - t_0)\Theta(t_0 - (T_f - \tau))\rangle \\ & + \langle \delta(t_0 + \tau - T_0 - t)\Theta(t_0)\Theta(T_f - t_0 - \tau)\Theta(T_0 - t_0) \\ & \times \Theta(t_0 - T_0 + \tau)\rangle + \langle \delta(T_f - T_0 - t)\Theta(t_0)\Theta(T_0 - t_0) \\ & \times \Theta(t_0 - (T_f + \tau))\Theta(\tau - (T_f + T_0))\rangle \Big), \end{aligned} \tag{S1}$$

where the operator $\langle \cdot \rangle$ is the ensemble average with respect to the random variable t_0, $\delta(x)$ and $\Theta(x)$ are the delta of Dirac distribution and the Heaviside function, respectively. $\mathcal{N}$ is the normalization. Solving the ensemble averages, the previous equation reads

$$\begin{aligned} p_{\tau''}(t|\tau) = \frac{1}{\mathcal{N}}\Big(& \delta(t - \tau) \int_{T_0}^{T_f - \tau} \Theta(T_f - T_0 - \tau)dt_0 \\ & + \Theta(T_f - T_0 - t)\Theta(\tau - t) \\ & + \Theta(T_f - T_0 - t)\Theta(T_0 + t - \tau)\Theta(\tau - t) \\ & + \delta(t - (T_f - T_0))\min[T_0, T_0 - (T_f - \tau)] \\ & \times \Theta(\tau - (T_f - T_0))\Big). \end{aligned} \tag{S2}$$

Marginalizing with respect to τ we obtain the probability distribution of τ'':

$$p_{\tau''}(t) = \int_0^\infty p_{\tau''}(t|\tau)p_\tau(\tau)d\tau. \tag{S3}$$

Eq. **S2** combined with Eq. **S3** yields

$$\begin{aligned} p_{\tau''}(t) = \frac{1}{\mathcal{N}}\Big(& (T_f - T_0 - t)p_\tau(t)\Theta(T_f - T_0 - t) \\ & + \Theta(T_f - T_0 - t)\int_{t>0}^\infty p_\tau(\tau)d\tau \\ & + \Theta(T_f - T_0 - t)\int_{t>0}^{T_0 + t} p_\tau(\tau)d\tau \\ & + \delta(t - (T_f - T_0))\int_{T_f - T_0}^\infty \min[T_0, T_0 - (T_f - \tau)]p_\tau(\tau)d\tau\Big). \end{aligned} \tag{S4}$$

The last term of Eq. **S4** gives an atom probability in $t = \Delta T_w = T_f - T_0$ corresponding to the fraction of species that are always present during the observational window.

Recalling Eq. **S1** and Eq. **3**, the normalization constant $\mathcal{N}$ reads

$$\begin{aligned} \mathcal{N} = \int_0^\infty \big(& \langle \Theta(t_0 - T_0)\Theta(T_f - (t_0 + \tau))\Theta(T_f - T_0 - \tau)\rangle \\ & + \langle \Theta(t_0 - T_0)\Theta(t_0 - (T_f - \tau))\Theta(T_f - t_0)\rangle \\ & + \langle \Theta(t_0)\Theta(T_f - (t_0 + \tau))\Theta(T_0 - t_0)\Theta(t_0 - (T_0 - \tau))\rangle \\ & + \langle \Theta(t_0)\Theta(T_0 - t_0)\Theta(t_0 - (T_f - \tau))\rangle \big)p_\tau(\tau)d\tau, \end{aligned} \tag{S5}$$

which simplifies to

$$\begin{aligned} \mathcal{N} = & \int_0^{T_f - T_0} (T_f - T_0 - \tau)p_\tau(\tau)d\tau \\ & + \int_0^\infty \min[T_f - T_0, \tau]p(\tau)d\tau \\ & + \int_0^\infty (\min[T_0, T_f - \tau] - \max[0, T_0 - \tau])p(\tau)d\tau \\ & + \int_{T_f - T_0}^\infty \min[T_0, T_0 - T_f + \tau]p_\tau(\tau)d\tau. \end{aligned} \tag{S6}$$

When comparing analytical and observational distributions, we assume that the system is at stationarity and unaffected by initial conditions; i.e., T_0 is far from the beginning of the process. Mathematically this is obtained taking the limit $T_0, T_f \to +\infty$ with $T_f - T_0 = \Delta T_w$ in Eq. **S4**, which finally takes the form

$$p_{\tau''}(t) = \frac{1}{\mathcal{N}} \bigg((\Delta T_w - t) p_\tau(t) \Theta(\Delta T_w - t)$$

$$+ \Theta(\Delta T_w - t) \int_{t>0}^{\infty} p_\tau(\tau) d\tau$$

$$+ \Theta(\Delta T_w - t) \int_{t>0}^{\infty} p_\tau(\tau) d\tau$$

$$+ \delta(t - \Delta T_w) \int_{\Delta T_w}^{\infty} (\tau - \Delta T_w) p_\tau(\tau) d\tau \bigg), \qquad \text{[S7]}$$

where $\mathcal{N}$ simplifies to

$$\mathcal{N} = \Delta T_w + \langle \tau \rangle - 2\Delta T_w P_\tau(\Delta T_w) + 2 \bigg(\int_0^{\Delta T_w} (P_\tau(t) - t p_\tau(t)) dt \bigg), \qquad \text{[S8]}$$

with $P_\tau(t) = \int_t^{+\infty} p_\tau(\tau) d\tau$ being the exceedance cumulative distribution of the persistence-time probability density function.

The variable τ' comprises only the first of the four cases listed in Eq. **S1**. Thus the probability distribution $p_{\tau'}(t)$ follows directly from the first term of Eq. **S7**:

$$p_{\tau'}(t) = \frac{1}{\mathcal{N}'} (\Delta T_w - t) p_\tau(t) \Theta(\Delta T_w - t), \qquad \text{[S9]}$$

where the normalization constant $\mathcal{N}'$ is equal to

$$\mathcal{N}' = \int_0^{\Delta T_w} (\Delta T_w - \tau) p_\tau(\tau) d\tau, \qquad \text{[S10]}$$

which completes the derivation.

Imperfect Detection. While studying animal communities on the basis of presence/absence (or count) data, imperfect detection of species is a source of concern, because animal species are routinely sampled with a detection probability <1. This represents a well-known issue for the breeding bird dataset under analysis, as convincingly shown in the literature (1–7). The aggregation procedure of breeding bird data from route to cell level explained in the main text reduces the probability of imperfect detection. We term p_r the route-level detection probability, i.e., the probability that a species is recorded given that it is present in the surveyed route. The probability of a pseudoabsence is therefore one minus the detection probability. Following ref. 7, the detection probability at cell level p_c can be expressed as $p_c = 1 - (1 - p_r)^n$, where n is the number of routes comprised in the cell where the species is present. Therefore, the probability of detection at cell scale increases rapidly with n. Note that in our finest scale of analysis ($A = 10,000$ km^2), cells comprise an average of six routes.

We have explicitly tested the sensitivity of our results to imperfect detection. Starting from the data at route level, we have added, for every year and every route, species randomly chosen among the assemblage of species observed in that route in the whole observational window (41 y). The number of species added is chosen so that the resulting route-level detection probability has a constant value p_r. The data so modified have been then aggregated at cell level and analyzed to derive the persistence-time distribution. Given the randomness in the choice of the added species, the operation has been repeated 1,000 times. An analysis on species imperfect detection in the Breeding Bird Survey dataset (7) has estimated an average route-level detection probability of about 0.8 for rare species and $p_r \simeq 0.9$ for more common species. We have run our analysis with the conservative

value $p_r = 0.8$ for all species. Fig. S2 shows the comparison between the probability distribution of the variables τ' and τ'' derived from the original breeding bird data and those derived after the addition of possibly undetected species as explained above. Fig. S2 refers to the finest scale of analysis, the more affected by imperfect detections. As expected, even with a low route-level detection probability ($p_r = 0.8$) imperfect detections have no significant impact on the resulting persistence-time distribution.

We also note that the problem of imperfect detection is far less relevant for the herbaceous plant dataset, given the sampling methodology and the small spatial scale of analysis (8).

Persistence-Time Distributions for Breeding Bird Passeriformes Species. We characterize the persistence-time distribution of a subset of the breeding bird data. Specifically we analyze the passeriformes species, which comprise 282 out of the 644 species included in the dataset. Remarkably, also the persistence times of the passeriformes species prove to be best fitted by a power-law distribution with an exponent ($\alpha = 1.76 \pm 0.05$, Fig. S3) close to the scaling exponent of the whole dataset ($\alpha = 1.83 \pm 0.02$).

Persistence-Time Distributions in a Individual-Based Neutral Model: The Role of the Dispersal Range. We have investigated the effects of dispersal ranges wider than nearest neighbors on persistence-time distributions. To that end, we have implemented the individual-based neutral (9, 10) model with uniform-dispersal with varying radii r. We have simulated the model in regular lattices in d spatial dimensions comprising N sites each of them occupied by one individual. To avoid edge effects, periodic boundary conditions have been prescribed. The model assumes neutral dynamics; therefore, all the basic ecological processes reproduced in the model (birth, death, dispersal, colonization, and diversification) are equivalent for all the individuals of all species. The system is assumed to be always saturated; i.e., no available resources or sites are left unexploited. At each time step, a randomly selected individual dies and the resources are freed up and available for colonization by another individual. With probability ν, the diversification rate, the empty site is colonized by individual of a species not currently present in the system. The diversification is a rate per birth that accounts for both speciation and immigration from surroundings communities. With the residual probability $1 - \nu$ the empty site is occupied by on offspring of an individual randomly sampled among all the sites that are less distant than r from the empty site, where r is measured in units of lattice spacing. For $r = 1$ we reduce to the nearest-neighbor dispersal case (von Neumann neighborhood). After the system has reached a statistically steady state, we start measuring the persistence times of 10^5 species, tracking their abundances from emergence to extinction.

Fig. S4 shows the persistence-time exceedance probability distribution for different dispersal radii. When the range of dispersal is small compared to the typical size of species' cluster, the persistence-time distributions tend to have, after a transient regime, the same scaling found in the nearest-neighbor case. As expected, as the radius of dispersal increases, persistence-time distributions approach the scaling of the global dispersal.

Persistence-Time Distributions in a Individual-Based Competition-Survival Trade-Off Model. We have investigated the resulting persistence-time distributions once the neutral assumption is relaxed and differences among species are considered. In this context, we have implemented on regular lattices a competition-survival trade-off model (11–13) where species have different mortality rate μ. We assume a trade-off between mortality and competitive advantage so that species with higher mortality rates have a higher probability of outcompeting species with lower mortality rate in the engagement for the colonization of an empty site. Oper-

ationally, each species s is labeled by its mortality rate μ_s. At every time step an individual, randomly selected among all the individuals in the system, dies with a probability equal to the mortality rate of the species it belongs to. If a death occurs the empty site is colonized by one individual of a species not currently present in the system with probability ν, the diversification rate. The mortality rate of the new species is sampled from a uniform distribution between 0 and 1. With the remaining probability $1 - \nu$, the empty site is colonized by an offspring of an individual chosen among the nearest neighbors with probability proportional to their mortality rate.

Fig. S5 shows the comparison between persistence-time distributions emerging from the neutral and the competition-survival trade-off model implemented in a two-dimensional lattice. Also in the trade-off model the persistence-time distribution, after a transient regime, exhibits a power-law behavior with an exponent ($\alpha = 1.80 \pm 0.03$) close to that found under neutrality ($\alpha = 1.82 \pm 0.01$). Noticing that persistence time in the trade-off model has a larger mean value $\langle \tau \rangle$ and recalling that the mean number of species S in the system at a certain time is $S = \nu N \langle \tau \rangle$ (14), we conclude that trade-off mechanisms can facilitate the coexistence of species, a result already suggested in the literature (10, 11).

Scaling Relations. In this section we show the derivation of the scaling relations used in the main text to obtain the specie area relationship. Assuming a persistence-time probability density function of the form

$$p_\tau(\tau) = \mathscr{A}\theta(\tau - t_{\min})\tau^{-\alpha}e^{-\nu\tau},$$

where ν is the diversification rate and $\mathscr{A} = \nu^{\alpha-1}/\Gamma(1 - \alpha, \nu t_{\min})$ is the normalization constant, we obtain that the mean persistence time can be expressed by

$$\langle \tau \rangle = \frac{\int_{t_{\min} \to 0}^{+\infty} \tau^{1-\alpha}e^{-\nu\tau}d\tau}{\int_{t_{\min} \to 0}^{+\infty} \tau^{-\alpha}e^{-\nu\tau}d\tau}$$

$$\approx \frac{\nu^{\alpha-2}}{t_{\min}^{1-\alpha} + \nu^{\alpha-1}} = \begin{cases} \frac{1}{\nu}, & \text{if } t_{\min}\nu \gg 1; \\ \frac{\nu^{\alpha-2}}{t_{\min}^{1-\alpha}}, & \text{if } t_{\min}\nu \ll 1. \end{cases}$$

In our case $\nu \ll 1$, and thus we obtain

$$\langle \tau \rangle \propto \nu^{\alpha-2},$$

which is a scaling relation between the mean persistence time and the diversification rate.

It is interesting to note that the relation $S = N\nu\langle\tau\rangle$ used in the main text to derive the species-area relationship can be alternatively derived by considering the species population dynamics. In fact, the mean number of species S at a certain time can be expressed as

$$S(t) = \frac{N}{\langle n(t) \rangle}, \tag{S11}$$

where N is the total number of individuals in the ecosystem and $\langle n \rangle$ is the averaged abundance of the species observed at a certain

time t. We term $p_n^{(s)}(t)$ the probability for the sth species of having an abundance n at time t. Different species have different random emergence times t_{0_s} that follow a Poisson distribution with frequency λ. Therefore $p_n^{(s)}(t)$ can be expressed as

$$p_n^{(s)}(t) = p_n^*(t - t_{0_s}),$$

where $p_n^*(t)$ is the probability for a single species to have abundance n at time t after its emergence. Under the neutral assumption $p_n^*(t)$ is species invariant. The mean population size at time t can be calculated as

$$\langle n(t) \rangle = \left\langle \frac{\sum_{s=1}^{D} \sum_{n=0}^{+\infty} p_n^*(t - t_{0_s})n}{\sum_{s=1}^{D} \sum_{n=1}^{+\infty} p_n^*(t - t_{0_s})} \right\rangle$$

$$\sim \frac{\left\langle \sum_{s=1}^{D} \sum_{n=0}^{+\infty} p_n^*(t - t_{0_s})n \right\rangle}{\left\langle \sum_{s=1}^{D} \sum_{n=1}^{+\infty} p_n^*(t - t_{0_s}) \right\rangle}, \tag{S12}$$

where $D(t)$ is the number of diversification events occurred until time t. Explicating the ensemble average ($\langle \cdot \rangle = \sum_D \frac{(\lambda t)^D}{D!}e^{-\lambda t}\int_0^t \prod_{s=1}^{D} \frac{dt_{0_s}}{t} \cdot$), using the fact that all the t_{0_s} are Poisson distributed with the same frequency λ and the Taylor expansion $\sum_{D \geq 1}(\lambda t)^{D-1}/(D-1)! = e^{\lambda t}$, Eq. **S12** simplifies to

$$\langle n(t) \rangle = \frac{\int_0^t dt' \sum_n np_n^*(t')}{\int_0^t dt' \sum_{n \geq 1} p_n^*(t')}. \tag{S13}$$

From the definition of $p_n^*(t)$ it follows that

$$\sum_n np_n^*(t') = \langle n^* \rangle_t$$

is the mean population of a species after a time t from its emergence. It is easy to show (15) that the term $\langle n^* \rangle_t$ obeys to the deterministic equation $d\langle n^* \rangle_t/dt = -\nu\langle n^* \rangle_t$, whose solution is $\langle n^* \rangle_t = \langle n^* \rangle_0 e^{-\nu t}$. Thus the numerator of Eq. **S13** is simply $\int_0^t \langle n^* \rangle_0 e^{-\nu t'}dt' = \langle n^* \rangle_0(1 - e^{-\nu t})/\nu$. We now observe that $\sum_{n \geq 1} p_n^*(t')$ is the probability that the species has more than one individual at time t; i.e., the cumulative distribution of the persistence-time probability density function $P_\tau(t) = \sum_{n \geq 1} p_n^*(t) = \int_t^{+\infty} p_\tau(\tau)d\tau$. Therefore the denominator of Eq. **S13** can be written as $\int_0^t dt' \sum_{n \geq 1} p_n^*(t') = \int_0^t dt' \int_{t'}^{+\infty} p_\tau(\tau)d\tau$. Simplifying Eq. **S13** and taking its stationary limit, we obtain

$$\langle n \rangle t \to +\infty = \frac{1}{\nu\langle\tau\rangle}, \tag{S14}$$

where we have set $\langle n^* \rangle_0 = 1$. Finally, substituting Eq. **S14** into Eq. **S11** we alternatively obtain the relation $S(t) = N\nu\langle\tau\rangle$ used in the main text.

1. Boulinier T, Nichols J, Sauer J, Hines J, Pollock K (1998) Estimating species richness: The importance of heterogeneity in species detectability. *Ecology* 79:1018–1028.
2. Nichols J, Boulinier T, Hines J, Pollock K, Sauer J (1998) Estimating rates of local species extinction, colonization, and turnover in animal communities. *Ecol Appl* 8:1213–1225.
3. Nichols J, Boulinier T, Hines J, Pollock K, Sauer J (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conserv Biol* 12:1390–1398.
4. Cam E, Nichols J, Sauer J, Hines J, Flather C (2000) Relative species richness and community completeness: Birds and urbanization in the mid-Atlantic states. *Ecol Appl* 10:1196–1210.
5. Boulinier T, et al. (2001) Forest fragmentation and bird community dynamics: Inference at regional scales. *Ecology* 82:1159–1169.
6. Cam E, Nichols J, Sauer J, Hines J (2002) On the estimation of species richness based on the accumulation of previously unrecorded species. *Ecography* 25:102–108.
7. Alpizar-Jara R, et al. (2004) The relationship between species detection probability and local extinction probability. *Oecologia* 141:652–660.
8. Adler PB, Tyburczy WR, Lauenroth WK (2007) Long-term mapped quadrats from Kansas prairie: Demographic information for herbaceous plants. *Ecology* 88:2673–2673.

9. Hubbell S (2001) *The Unified Theory of Biodiversity and Biogeography.* (Princeton Univ. Press, Princeton, NJ).
10. Chave J, Muller-Landau H, Levin S (2002) Comparing classical community models: Theoretical consequences for patterns of diversity. *Am Nat* 159:1–23.
11. Tilman D (1994) Competition and biodiversity in spatially structured habitats. *Ecology* 75:2–16.
12. Buttel L, Durrett R, Levin S (2002) Competition and species packing in patchy environments. *Theor Pop Biol* 61:265–276.
13. Chave J (2004) Neutral theory and community ecology. *Ecol Lett* 7:241–253.
14. Rodriguez-Iturbe I, Cox D, Isham V (1987) Some models for rainfall based on stochastic point-processes. *Proc R Soc London Ser A* 410:269–288.
15. Pigolotti S, Flammini A, Marsili M, Maritan A (2005) Species lifetime distribution for simple models of ecologies. *Proc Natl Acad Sci USA* 102:15747–15751.
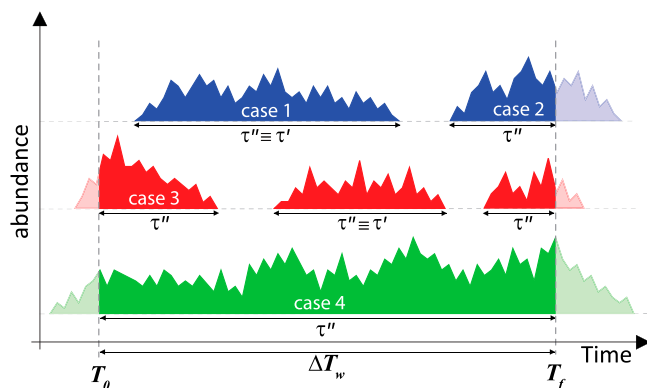
Fig. S1. Schematic representation of the variables that can be measured from empirical data over a time window ΔT_w: τ' (persistence times that start and end inside the observational window) and τ'' (which comprises the persistence times τ' and all the portions of species persistence times that are partially seen inside the observational time window but start or/and end outside). Cases 1, 2, 3, and 4 refer to the description provided in the text.
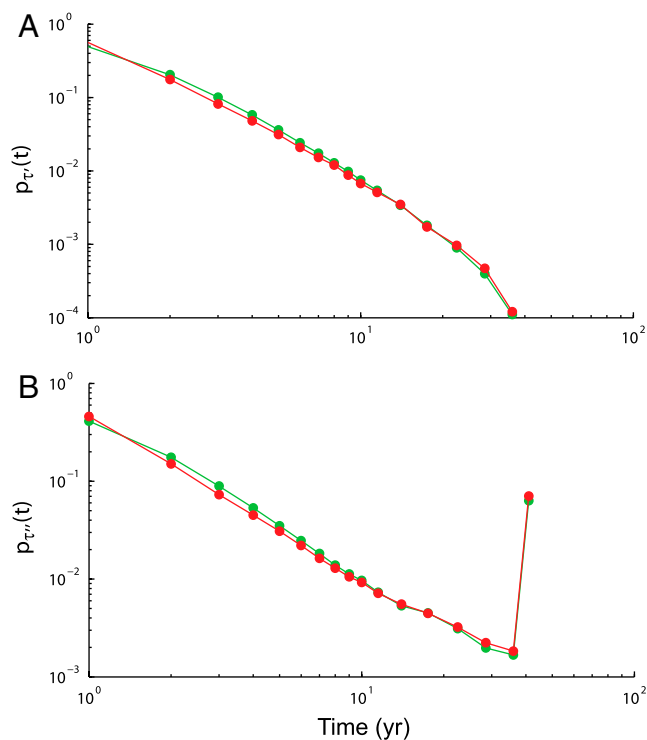
Fig. S2. Imperfect detections in the breeding bird dataset. Comparison between the probability distributions of the variables τ' (*A*) and τ'' (*B*) (see main text for explanation) derived from the original data (red) and those derived after the addition of possibly undetected species (green).

Fig. S3. Persistence-time distributions of breeding bird passeriformes species. Probability density function of τ' (green), τ'' (blue), and persistence times τ (red) (see main text for explanation). Filled circles and solid lines show observational distributions and fits, respectively. The values of the best fit exponent is $\alpha = 1.76 \pm 0.05$.

Fig. S4. Persistence-time exceedance probability distributions $P_\tau(t)$ for the neutral model implemented in a one-dimensional lattice. Different colors refer to different uniform-dispersal radii. For all simulations $\nu = 10^{-5}$ and time is expressed in generation time units (9).

Fig. S5. Comparison between the persistence-time exceedance probability distribution $P_\tau(t)$ for the neutral (green) and the competition-survival trade-off (blue) model implemented in a two-dimensional lattice. Note that in the power-law regime if $p_\tau(t)$ scales as $t^{-\alpha}$, $P_\tau(t) \propto t^{-\alpha+1}$. The scaling exponent α in the range $10^3 <$ Time $< 10^4$ is equal to 1.82 ± 0.01 for the neutral model and 1.80 ± 0.03 for the trade-off model. Errors are estimated through bootstrap method (random sampling with replacement). For all simulations $\nu = 10^{-5}$ and time is expressed in generation time units (9).