METHOD

ECOLOGY LETTERS cnrs WILEY

The kinship matrix: inferring the kinship structure of a population from its demography

Christophe F. D. Coste[1] | **François Bienvenu[2,3,4,5]** | **Victor Ronget[6]** | **Juan-Pablo Ramirez-Loza[7]** | **Sarah Cubaynes[7]** | **Samuel Pavard[6]**

[1]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Mathematics and Computer Science, Transilvania University of Braşov, Braşov, Romania

[3]Institut des Sciences de l'Evolution de Montpellier, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France

[4]UMR AGAP, Université de Montpellier, CIRAD, INRAE, L'institut Agro, Montpellier, France

[5]Department of Statistics, University of Oxford, Oxford, United Kingdom

[6]Unité Eco-anthropologie (EA), Muséum National d'Histoire Naturelle, CNRS, Université Paris Diderot, Paris, France

[7]CEFE, CNRS, Univ. Montpellier, Univ. Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

Correspondence
Christophe F. D. Coste, Center for Biodiversity Dynamics, Department of Biology, NTNU, Trondheim, Norway.
Email: christophe.f.d.coste@ntnu.no

François Bienvenu, Facultatea de Matematică şi Informatică, Universitatea Transilvania din Braşov, Str. Iuliu Maniu nr. 50, Braşov, Romania.
Email: francois.bienvenu@normalesup.org

Editor: Tim Coulson

Abstract

The familial structure of a population and the relatedness of its individuals are determined by its demography. There is, however, no general method to infer kinship directly from the life cycle of a structured population. Yet, this question is central to fields such as ecology, evolution and conservation, especially in contexts where there is a strong interdependence between familial structure and population dynamics. Here, we give a general formula to compute, from any matrix population model, the expected number of arbitrary kin (sisters, nieces, cousins, etc) of a focal individual *ego*, structured by the class of *ego* and of its kin. Central to our approach are classic but little-used tools known as genealogical matrices. Our method can be used to obtain both individual-based and population-wide metrics of kinship, as we illustrate. It also makes it possible to analyse the sensitivity of the kinship structure to the traits implemented in the model.

KEYWORDS
familial structure, kinship demography, population projection matrices, relatedness, structured populations

Christophe F. D. Coste and François Bienvenu are co-first authors.

INTRODUCTION

The demography of a population determines its level of inbreeding and its kinship structure, that is, the network of kin relationships between the individuals of the population. Population geneticists have long used non-overlapping generation models to study the effects of demography on kinship structure, for instance to estimate the frequency of consanguineous marriage (Barrai et al., 1962; Hajnal, 1963). Models with overlapping generations have also been considered—from the seminal work of Goodman et al., (1974) to the recent advances by Caswell (2019, 2020)—but they focus on specific kin relationships. Thus, a general method has yet to be developed to infer kinship structure from population dynamics in structured populations, and little is known about the influence of specific life-history traits (or vital rates in general) on inbreeding levels, inclusive fitness and kinship/familial structures. Yet, those questions are central to fields such as ecology, demography, evolution, genetics and conservation.

In this article, we pave the way towards such a general method by showing how classic tools from the study of matrix population models can be combined in a new way to infer kinship structure from steady-state population dynamics. More specifically, we explain how to compute the expected number of kin of a focal individual *ego*, as a function of the classes of *ego* and of its kin, directly from the projection matrix of a population. Examples of applications include studying how inbreeding depression leads to genetic Allee effects in populations of small effective sizes or modelling the eco-evolutionary demography of social species (where the vital rates of individuals may strongly depend on cooperative and competitive interactions with kin). Readily usable implementations of our methods are available at DOI:10.5281/zenodo.4680716, see SM.1.

A brief history of kinship inference in structured populations

The first demographers to work on the estimation of kin distribution from the demographic rates of a structured population based their computations on age-structured, one-sex populations at the demographic steady state, and focused on a few key kin relationships (Le Bras, 1973; Goodman et al., 1974, discussed in Pavard & Coste, 2020). Their method, still widely used today, consists in computing the distribution of maternal age-at-childbirth from the Euler–Lotka equation (Lotka, 1939), and then use it to derive, for each age-class, the probability that the mother/grandmother of a focal individual is alive, as well as the expected number of daughters, sisters, aunts and nieces of that individual.

Shortly after, mathematicians attacked the problem from the field of branching processes. First, for unstructured populations without overlapping generations, using Bienaymé–Galton–Watson processes (Pullum, 1982; Waugh, 1981); then, by adding age as a category and considering multitype Crump–Mode–Jagers processes (e.g. Nerman & Jagers, 1984). However, the technical nature of these papers and their focus on sometimes abstract results has made them go largely unnoticed by demographers. For instance, despite being directly relevant to the matter, the work by Joffe and Waugh (1985) is only cited nine times in the mathematical literature—and not once in a biology or demography journal.

As of today, the inferred kin frequencies provided by Goodman et al., (1974), which have been in constant use since their publication, are still the state of the art (Pavard & Coste, 2020). These take the form of a limited number of *ad hoc* formulas, each specific to a kin relationship, that only apply to age-structured populations where at most one offspring can be produced at a time. Recently, Caswell (2019, 2020) has put this framework in matrix form and extended it to age × stage models, by giving a system of recursive equations expressing the number of kin at time $t + 1$ as a function of the number of kin at time t. However, before the present work computing the expected number of any kin in any life cycle remained an open problem and a necessary step in the development of a general theory of the interplay between kinship and demographic processes (Pavard & Coste, 2020).

Specific challenges to overcome

Inferring the complete kinship structure of an arbitrary structured population poses several challenges. First, in general class-structured models, the distribution of the class at birth is not as easily inferred as in age-structured populations, where it is obtained from the Euler–Lotka equation. We solve this problem by using the genealogical Markov chains first introduced by Demetrius to define population entropy (Demetrius, 1974, 1975). These somewhat little-known tools have recently proved useful to tackle various questions (Bienvenu et al., 2017; Bienvenu & Legendre, 2015).

The fact that several offspring can be produced at the same time constitutes another complication, because this prevents segregating genealogies as done in Goodman et al., (1974) and Caswell (2019, 2020). Accounting for the possibility of same-litter sisters requires special care and fine-grained information on reproduction.

Finally, the main difficulty is arguably the generalisation of the calculations to arbitrary kin relationships. This requires combining two different timescales: the demographic timescale (where time is expressed in fixed projection intervals, such as 1 year) and the genealogical one (where time is expressed in number of generations).

We do so by working out all the genealogies and associated trajectories of individuals in the life cycle that correspond to a given kin relationship.

The kinship matrices

In order to describe the kinship structure of a population, we need a generic notation for arbitrary kin relationships. In one-sex populations, we can use the following: we say that two individuals (i_1, i_2) are (g, q)-*kin* when their most recent common ancestor is separated from i_1 by g generations and from i_2 by q generations. For instance, the $(1, 0)$-kin of the focal individual *ego* is its mother, and its $(0, 1)$-kin are its daughters; similarly, its $(1, 1)$-kin are its sisters; its $(1, 2)$-kin its nieces; its $(2, 1)$-kin its aunts; etc (see Figure 1a). The same convention was used by Atkins (1974) and Pullum (1982), but with the order of g and q reversed.

We describe the kinship structure of a class-structured population through the *kinship matrices* $\mathbf{K}(g, q)$, where the (i, j)-th entry of $\mathbf{K}(g, q)$ is the expected number of (g, q)-kin alive in class i of *ego* in class j. From the point of view of *ego*—which is what matters for most applications—these matrices completely characterise the (expected) kinship structure of the population. An example of the genealogy of a focal individual, together with the corresponding population model and kinship matrix for daughters, is given in Figure 1b,c.

The rest of this article is organised as follows: First, we explain how to compute the kinship matrices directly from the projection matrix of the population. We then illustrate the method in Box 2 using the life cycle of the ground squirrel *Spermophilus dauricus* from Luo and Fox (1990); other detailed examples of applications can be found in SM.2. Finally, we conclude by discussing the limitations, implications and possible extensions of our method.
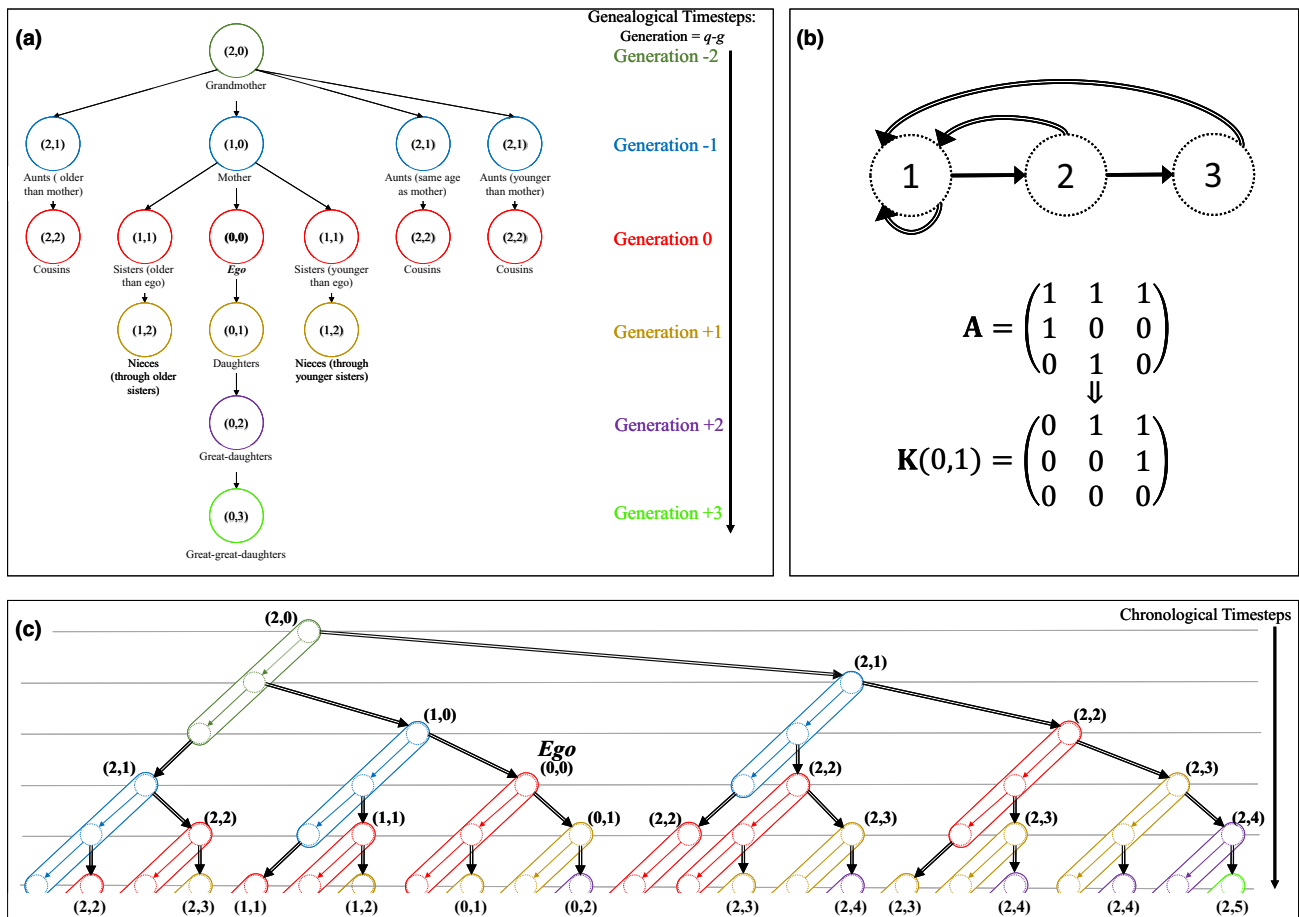

FIGURE 1 (a) Kin relationships, as described in terms of (g, q)-kin of the focal individual *ego*. Each line (and colour) corresponds to a generation. To get to the (g, q)-kin of *ego*, one has to go "up" g generations in order to reach the most recent common ancestor of *ego* and its (g, q)-kin; and then "down" q generations to reach the (g, q)-kin themselves. (b) Simple example of an age-structured life cycle, with the corresponding projection matrix and kinship matrix $\mathbf{K}(0, 1)$ giving the expected number of daughters of *ego*. In this toy model, all individuals survive to age 3 and then die, and they produce exactly one offspring each year. Thus, individuals aged 2 have exactly one daughter (aged 1) and individuals aged 3 have exactly two daughters (one aged 1 and one aged 2). (c) The genealogy of *ego* for the population described in (b), with the corresponding kin structure. Each dotted circle corresponds to an individual at a given age, and each solid oval to an individual's life trajectory. The colors correspond to the generation with respect to *ego*, as in (a). Here, we have assumed that *ego* was born when its mother was aged 1, and that the mother of *ego* was born when its grandmother was aged 2

METHOD

Our method is applicable to populations at the demographic steady state whose dynamics are governed by a matrix population model with a primitive (i.e. irreducible and aperiodic) projection matrix $\mathbf{A} = (a_{ij})$. It can also be used with some reducible matrices (see SM.3), but here we assume primitivity, for simplicity. Recall that a_{ij} is the per-capita contribution of an individual in class j at time t to the abundance of class i at time $t + 1$. The asymptotic growth rate λ is given by the dominant eigenvalue of $\mathbf{A}$ and the stable class distribution by the corresponding right eigenvector $\mathbf{w} = (w_i)$ such that $\sum_i w_i = 1$.

To compute the kinship matrix $\mathbf{K}(g, q)$, we decompose the projection matrix $\mathbf{A}$ into its survival and reproductive components—that is, into $\mathbf{A} = \mathbf{S} + \mathbf{F}$, where the *survival matrix* $\mathbf{S} = (s_{ij})$ is such that s_{ij} is the probability that an individual of class j survives into class i, and the *fertility matrix* $\mathbf{F} = (f_{ij})$ is such that f_{ij} is the expected number of offspring of class i produced by an individual of class j. Mathematically, we need reproduction and survival to be independent (it is otherwise not possible to retrace the life trajectory of an ancestor and then use this trajectory to compute its number of offspring). This assumption is never applicable in post-breeding models, where individuals have to survive a full projection interval before they can reproduce. We therefore recommend using pre-breeding models whenever possible and avoiding post-breeding models with our method. Birth-flow models can also be used as long as the covariance between survival and reproduction is kept in mind as a potentially significant source of discrepancy. Finally, note that one way to deal with covarying reproduction and survival (e.g. due to trade-offs) is to add (*survival*, *fertility*) classes to the model.

Finally, to take into account models with several offspring per time step, we need the *same-litter newborn sisters matrix* $\mathbf{Z} = (z_{ij})$, where z_{ij} is the expected number of same-litter sisters of *ego* that have just been born in class i, knowing that *ego* is in class j. Unlike $\mathbf{S}$ and $\mathbf{F}$, the matrix $\mathbf{Z}$ is not always directly available, and to compute it one will typically need to make additional assumptions on the offspring distribution. We give expressions of $\mathbf{Z}$ under three of the most common scenarios: when at most one offspring is produced by projection interval; when there is only one class of offspring: and when the total number of offspring follows a Poisson distribution and each offspring is independently allocated to a newborn class.

The three matrices $\mathbf{S}$, $\mathbf{F}$ and $\mathbf{Z}$ are all we need to compute any kinship matrix $\mathbf{K}(g, q)$, as summarised in Box 1. Python and MATLAB implementations of the method are provided (see SM.1). In the rest of this section, we detail the reasoning behind the expression of $\mathbf{K}(g, q)$ and give examples of aggregated metrics of the kinship or relatedness structure of a population that can be derived from it. Concrete examples of application are given in Box 2 and SM.2.

The genealogical Markov chain

The central tool of our approach is the genealogical Markov chain introduced by Demetrius (1974, 1975) to define population entropy. Despite its early introduction in the field of matrix population models, this Markov chain has remained mostly associated with population entropy (see, e.g. Tuljapurkar, 1982, 1993) and has so far failed to become part of the standard toolbox of population ecologists and demographers. However, it has recently proved useful to study questions such as the generation time (Bienvenu & Legendre, 2015) and the optimal aggregation of classes (Bienvenu et al., 2017). Our work provides yet another example of the usefulness of this tool.

The transition matrix of the genealogical Markov chain associated with $\mathbf{A}$ is the matrix $\mathbf{P} = (p_{ij})$, where

$$p_{ij} = \frac{a_{ij} w_j}{\lambda w_i}. \tag{1}$$

This Markov chain describes the sequence of classes that we encounter as we go "up" the genealogy of the population by following the lineage of an individual, backwards in time. That is, p_{ij} is the probability that an individual in class i at time t was either alive and in class j at time $t - 1$; or else had its mother in class j at that time. Thus, the probability distribution of the class of the ancestor (in the sense of a younger self or a genealogical ancestor), t time steps ago, of an individual currently in class i is given by $\mathbf{e}_i \mathbf{P}^t$, where $e_i(j) = 1$ if $j = i$ and 0 otherwise (note that the formalism of Markov chain uses right multiplication to project probability distributions, as in $\mathbf{x}(t + 1) = \mathbf{x}(t) \mathbf{P}$, whereas population projection matrices use left multiplication). See Bienvenu et al., (2017) for more on genealogical Markov chains.

As we did with $\mathbf{A}$, we can split the transitions of the genealogical matrix into a survival and a reproductive component. We then have $\mathbf{P} = \mathbf{P}_\mathbf{S} + \mathbf{P}_\mathbf{F}$, where $\mathbf{P}_\mathbf{S} = (p_\mathbf{S}(i,j))$ and $\mathbf{P}_\mathbf{F} = (p_\mathbf{F}(i,j))$ are given by

$$p_\mathbf{S}(i,j) = \frac{s_{ij} w_j}{\lambda w_i} \text{ and } p_\mathbf{F}(i,j) = \frac{f_{ij} w_j}{\lambda w_i}. \tag{2}$$

Accordingly, $p_\mathbf{S}(i,j)$ is the probability that an individual in class i at time t was alive and in class j at time $t - 1$, while $p_\mathbf{F}(i,j)$ is the probability that it was born between times $t - 1$ and t to a mother in class j. Thus, for instance, the (i,j)-th entry of $\mathbf{P}_\mathbf{S} \mathbf{P}_\mathbf{F}$ is the probability that the mother of an individual from class i was in class j two time steps ago *and* that the first transition when going up the genealogy was a survival transition, while the second one was a reproductive transition.

We are now able to: (1) project the class of the ancestor of *ego* backwards in time (within generations using $\mathbf{P}_\mathbf{S}$ and between generations using $\mathbf{P}_\mathbf{F}$); and (2) project the descent of an ancestor forwards in time (within generations

---

Box 1   Expression of the kinship matrix $\mathbf{K}(g,q)$

• Input: let $\mathbf{A} = \mathbf{S} + \mathbf{F}$ be the projection matrix of the population, where $\mathbf{S}$ is the survival matrix and $\mathbf{F}$ is the fertility matrix. Let $\mathbf{Z}$ be the same-litter newborn sisters matrix (see main text).
• Compute the genealogical matrices $\mathbf{P_S} = (p_\mathbf{S}(i,j))$ and $\mathbf{P_F} = (p_\mathbf{F}(i,j))$ defined by

$$p_\mathbf{S}(i,j) = \frac{s_{ij}w_j}{\lambda w_i} \text{ and } p_\mathbf{F}(i,j) = \frac{f_{ij}w_j}{\lambda w_i},$$

where λ is the asymptotic growth rate (dominant eigenvalue of $\mathbf{A}$) and $\mathbf{w}$ is the stable class distribution vector (dominant right eigenvector of $\mathbf{A}$).
• For all integers $n \geq 0$ and $\ell \geq 1$, let $\mathscr{C}(n,\ell)$ denote the set of vectors $\mathbf{k} = (k_1, ..., k_\ell)$ of non-negative integers such that $k_1 + \cdots + k_\ell = n$ (when $\ell = 0$ or $n < 0$, let $\mathscr{C}(n,\ell) = \varnothing$).
• For any $\mathbf{k} = (k_1, ..., k_\ell)$, let $\mathbf{Q}(\mathbf{k}) = \mathbf{P_S}^{k_1}\mathbf{P_F}\cdots\mathbf{P_F}\mathbf{P_S}^{k_\ell}$ and

$$\mathbf{U}(t,g) = \sum_{\mathbf{k} \in \mathscr{C}(t-g,g)} \mathbf{Q}(\mathbf{k}),$$

with the convention that $\mathbf{U}(0,0) = \mathbf{I}$, the identity matrix, and that $\mathbf{U}(t,g) = \mathbf{0}$ when the sum is empty.
• For any $\mathbf{k} = (k_1, ..., k_\ell)$, let $\mathbf{R}(\mathbf{k}) = \mathbf{S}^{k_1}\mathbf{F}\cdots\mathbf{F}\mathbf{S}^{k_\ell}$ and

$$\mathbf{D}(t,q) = \sum_{\mathbf{k} \in \mathscr{C}(t-1-q,q+1)} \mathbf{R}(\mathbf{k}),$$

with the convention that $\mathbf{D}(t,q) = \mathbf{0}$ when the sum is empty.
• The kinship matrix $\mathbf{K}(g,q)$, whose (i,j)-th entry is the expected number of (g,q)-kin of class i of a focal individual of class j, is $\mathbf{K}(0,0) = \mathbf{I}$ and, for $(g,q) \neq (0,0)$,

$$\mathbf{K}(g,q) = \sum_{t \geq 0} \left(\mathbf{D}(t,q)\,\mathbf{S}\mathbf{P_F}^\mathsf{T} + \mathbf{D}(t,q-1)\,\mathbf{Z} + \sum_{\alpha \geq 1} \mathbf{D}(t+\alpha,q-1)\,\mathbf{F}\left(\mathbf{P_S}^\mathsf{T}\right)^\alpha \mathbf{Y}(g) \right) \mathbf{U}(t,g)^\mathsf{T},$$

where $\mathbf{Y}(g) = \mathbf{I}$ if $g = 0$ and $\mathbf{P_F}^\mathsf{T}$ otherwise.

---

using $\mathbf{S}$ and between generations using $\mathbf{F}$). With this, our strategy to compute the number of (g,q)-kin of *ego* is to go "up" the genealogy to locate its g-th ancestor; and then "down" to compute the q-th descendants of that ancestor. However, to account for same-litter sisters, we also need to look at the $(g-1)$-th ancestor of *ego*.

Distribution of the $(g-1)$-th ancestor of *ego*

Let us start by assuming that $g \geq 1$ (the case $g = 0$ will be treated separately later). For $t \geq 1$, let $\mathbf{U}(t,g)$ be the matrix whose (i,j)-th entry is the probability that $t-1$ time steps ago the $(g-1)$-th ancestor of *ego* was in class j, given that *ego* is currently in class i. To compute $\mathbf{U}(t,g)$, we need to consider all the ways that $g-1$ generations can pass in the span of $t-1$ time steps. For any integers $n \geq 0$ and $\ell \geq 1$, let

$$\mathscr{C}(n,\ell) = \{(k_1, ..., k_\ell) \in \mathbb{N}^\ell : k_1 + \cdots + k_\ell = n\} \quad (3)$$

denote the set of vectors of length ℓ whose components are non-negative integers that sum to n (with $\mathscr{C}(n,\ell) = \varnothing$ when $\ell = 0$ or $n < 0$). Such vectors are known as *weak compositions of the integer n into ℓ parts*, and there are efficient algorithms to list them (see SM.4). The weak compositions $\mathbf{k} \in \mathscr{C}(t-g,g)$ exactly encode the sequences of $t-1$ "survival" or "reproduction" events that contain $g-1$ "reproduction": indeed, $\mathbf{k} = (k_1, ..., k_g)$ corresponds to the sequence

$$\underbrace{\text{survival}, ..., \text{survival}}_{k_1 \text{ times}}, \text{reproduction}, \cdots,$$

$$\text{reproduction}, \underbrace{\text{survival}, ..., \text{survival}}_{k_g \text{ times}},$$

Thus, letting $\mathbf{Q}(\mathbf{k}) = \mathbf{P_S}^{k_1}\mathbf{P_F}\cdots\mathbf{P_F}\mathbf{P_S}^{k_g}$—whose (i,j)-th entry is the probability that, starting from class i, after $k_1 + ... + k_g + g - 1$ steps one ends in class j after having encountered k_1 survivals, then one reproduction, then k_2 survivals, etc—we have

Box 2 Illustration

We illustrate our method on the most common model in animal ecology: the extended Leslie model, that is, an age-structured model with a class that regroups all individuals older than a certain age. More complex examples are given in SM.2, to illustrate and discuss specific points; but once an appropriate projection matrix has been obtained the method is always the same. Recall that we need to consider a pre-breeding model—that is, in the case of a three-class model, a projection matrix of the form

$$
\mathbf{A} = \begin{pmatrix} s_0 m_1 & s_0 m_2 & s_0 m_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & s_3 \end{pmatrix},
$$

where classes 1 and 2 correspond to individuals that are about to turn 1 year and 2 years old, respectively, and class 3 corresponds to older individuals; s_0 is the probability that newborns survive their first year, s_i is the probability that class-i individuals survive to the next year and m_i is the average number of (female) newborns to which they give birth.

Using the life table for the ground squirrel *Spermophilus dauricus* from Luo and Fox (1990), we get the following survival and fertility matrices:

$$
\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 \\ 0.57 & 0 & 0 \\ 0 & 0.54 & 0.46 \end{pmatrix} \text{ and } \mathbf{F} = \begin{pmatrix} 0.23 & 0.51 & 0.84 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
$$

The asymptotic growth rate is $\lambda = 1.00$ and the stable class distribution is $\mathbf{w} = (0.47, 0.27, 0.27)$. From this, Equation (2) gives

$$
\mathbf{P_S} = \begin{pmatrix} 0 & 0 & 0 \\ 1.00 & 0 & 0 \\ 0 & 0.54 & 0.46 \end{pmatrix} \text{ and } \mathbf{P_F} = \begin{pmatrix} 0.23 & 0.29 & 0.48 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
$$

For instance, the (3,2)-th entry of $\mathbf{P_S}$ shows that, at the demographic steady state, 54% of class-3 individuals were in class 2 the previous year. Similarly, the (1, 1)-th entry of $\mathbf{P_F}$ indicates that a yearling has a 23% probability of having been born to a yearling the previous year. Since there is a single newborn class, we can use Equation (10) to compute the matrix $\mathbf{Z}$ under arbitrary assumptions on the fertilities. But here, in the absence of additional information, we will simply assume that they correspond to Poisson random variables, so that $\mathbf{Z} = \mathbf{FP_F^T}$. This yields

$$
\mathbf{Z} = \begin{pmatrix} 0.60 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
$$

Let us now compute $\mathbf{K}(1, 0)$, whose (i, j)-th entry is the probability that the mother of an individual of class j is currently alive and in class i. Using $t_{max} = 9$ and $\alpha_{max} = 0$, which as per SM.6 gives an approximation error $\left\| \hat{\mathbf{K}} - \mathbf{K} \right\|_1 < 10^{-2}$, we get

$$
\mathbf{K}(1, 0) = \begin{pmatrix} 0 & 0 & 0 \\ 0.13 & 0 & 0 \\ 0.38 & 0.24 & 0.08 \end{pmatrix}.
$$

This matrix indicates, for instance, that a yearling has a 13% probability that its mother is alive in class 2, and 38% that it is alive in class 3. Note that kin are structured by their *current* class. Thus, it is not possible to have yearling mothers here: even if the mother of *ego* was a yearling when *ego* was born, by now it will have moved to an adult class.

Box 2 Continued

Figure 2a gives a graphical representation of $\mathbf{K}(2,1)$, the kinship matrix for aunts. The total number of aunts (irrespective of their class) of an individual sampled according to the stable class distribution is $\mathbf{1K}(2,1)\mathbf{w} = 0.36$, and numerically it is possible to compute its elasticities to the entries of the projection matrix, that is, $\partial \ln\mathbf{1K}(2,1)\mathbf{w}/\partial \ln a_{ij}$. These are plotted in Figure 2b, which shows that the number of aunts is most affected by relative changes in the survival probabilities s_1 and s_3.

Finally, we get sensibly the same results when taking the values of t_{max} and α_{max} suggested by Equation (16), with the maximal age $\hat{\omega} = 6$ reported in Luo and Fox (1990). This is also the case with other models that we studied (see SM.2), which tends to suggest our simple heuristic to chose t_{max} and α_{max} based on a maximal age is adequate.

In SM.12, we compare the kinship matrices $\mathbf{K}(1,0)$ and $\mathbf{K}(2,1)$ to the output of individual-based simulations, confirming our formulas.
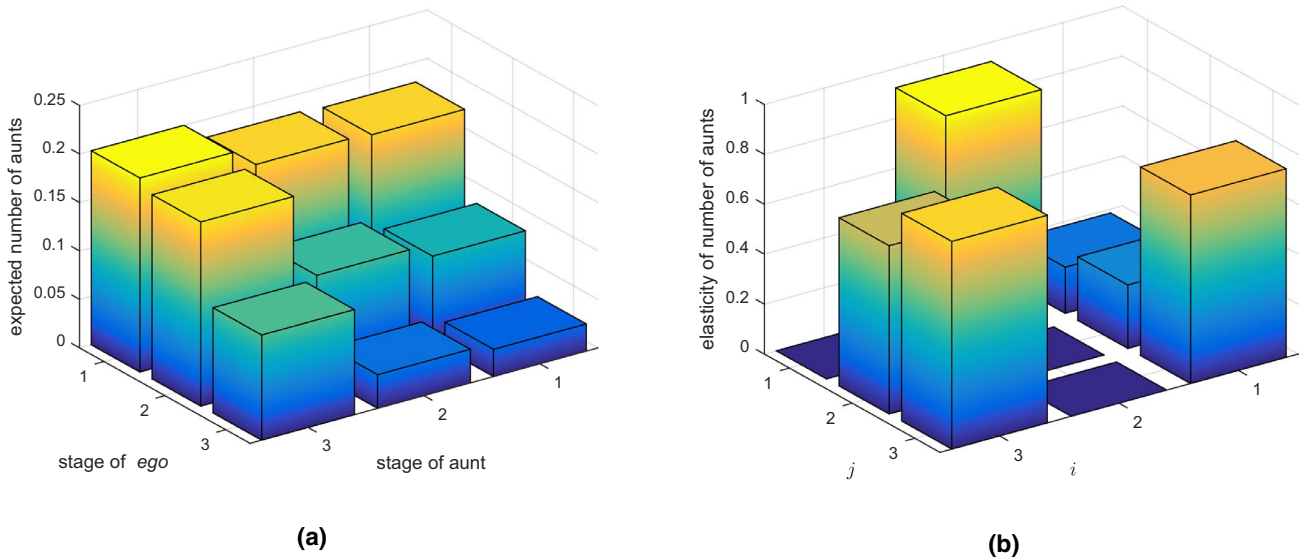

(a)

(b)

FIGURE 2 (a) Graphical representation of the matrix $\mathbf{K}(2,1)$, whose entries give the expected number of aunts of *ego* as a function of its class and that of its aunts; (b) the elasticities of the expected total number of aunts of *ego* (sampled at random in the stable population), $\partial\ln\mathbf{1K}(2,1)\mathbf{w}/\partial\ln a_{ij}$

$$\mathbf{U}(t,g) = \sum_{\mathbf{k}\in\mathscr{C}(t-g,g)} \mathbf{Q}(\mathbf{k}), \qquad (4)$$

with the convention that $\mathbf{U}(t,g) = \mathbf{0}$ when the sum is empty (e.g., if $t < g$).

As examples of the formula, note that for $g = 1$ and $t \geq 1$, $\mathbf{U}(t,1) = \mathbf{P}_\mathbf{S}^{t-1}$, and for $g = 2$ and $t = 2, 3, 4, \ldots$

$$\mathbf{U}(2,2) = \mathbf{P}_\mathbf{F}, \quad \mathbf{U}(3,2) = \mathbf{P}_\mathbf{F}\mathbf{P}_\mathbf{S} + \mathbf{P}_\mathbf{S}\mathbf{P}_\mathbf{F},$$
$$\mathbf{U}(4,2) = \mathbf{P}_\mathbf{F}\mathbf{P}_\mathbf{S}^2 + \mathbf{P}_\mathbf{S}\mathbf{P}_\mathbf{F}\mathbf{P}_\mathbf{S} + \mathbf{P}_\mathbf{S}^2\mathbf{P}_\mathbf{F}, \quad \ldots$$

Finally, letting $\mathbf{U}(t,g) = (u_{ij}(t,g))$, the probability, given that *ego* is currently in class i, that its g-th ancestor gave birth to its $(g-1)$-th ancestor in class ℓ, from class k and between times t and $t-1$ before the present is $u_{i\ell}(t,g)\,p_\mathbf{F}(\ell,k)$. The reason why we need to look at both the $(g-1)$-th and the g-th ancestor of *ego*, rather than at its g-th ancestor only, is that we will need this information to account for the (g,q)-kin *ego* that are descendants

of same-litter sisters of its $(g-1)$-th ancestor. Note that the probability, given that *ego* is currently in class i, that its g-th ancestor gave birth to its $(g-1)$-th ancestor from class k and between times t and $t-1$ before the present is the (i,k)-th entry of $\mathbf{U}(t,g)\mathbf{P}_\mathbf{F}$.

Expected number of q-th generation descendants

Let $\mathbf{D}(t,q)$ be the matrix whose (i,j)-th entry is the expected number of q-th generation descendants in class i, after $t-1$ time steps, of an individual in class j (not counting offspring produced before the first time step). As before, to compute $\mathbf{D}(t,q)$ we need to consider all the ways that q generations can pass in the span of $t-1$ time steps. For $\mathbf{k} = (k_1, \ldots, k_\ell)$, let $\mathbf{R}(\mathbf{k}) = \mathbf{S}^{k_1}\mathbf{F}\cdots\mathbf{F}\mathbf{S}^{k_\ell}$. Recall that the expected number of descendants in class i of an individual in class j, after $t-1$ time steps (and counting the initial individual) is the (i,j)-th entry of the matrix

$$\mathbf{A}^{t-1}=(\mathbf{S}+\mathbf{F})^{t-1}=\sum_{q=0}^{t-1}\sum_{\mathbf{k}\in\mathscr{C}(t-1-q,q+1)}\mathbf{R}(\mathbf{k}).$$

For instance, for $t-1=3$,

$$\mathbf{A}^3 = \underbrace{\mathbf{S}^3}_{q=0} + \underbrace{\mathbf{FS}^2 + \mathbf{SFS} + \mathbf{S}^2\mathbf{F}}_{q=1}$$
$$+ \underbrace{\mathbf{F}^2\mathbf{S} + \mathbf{FSF} + \mathbf{SF}^2}_{q=2} + \underbrace{\mathbf{F}^3}_{q=3}.$$

In these expressions, the terms corresponding to a fixed value of q correspond to the q-th generation descendants of the initial individual. In other words,

$$\mathbf{D}(t,q) = \sum_{\mathbf{k}\in\mathscr{C}(t-1-q,\,q+1)}\mathbf{R}(\mathbf{k}),\qquad(5)$$

again with the convention that $\mathbf{D}(t,q)=\mathbf{0}$ when the sum is empty (e.g. when $q<0$ or $t-1<q$).

Expected number of (g,q)-kin

Let *ego* be in class j, and recall that we assume $g\geq1$. For t ranging over the positive integers and k,ℓ over the classes of the model, the events "the g-th ancestor of *ego* gave birth to its $(g-1)$-th ancestor in class ℓ, from class k, and $t-1$ time steps before the present" form a complete system of events. Each of these events has probability $u_{j\ell}(t,g)\,p_{\mathbf{F}}(\ell,k)$. Thus, all we need in order to compute the expected number of (g,q)-kin of *ego* is the expected number

of q-th generation descendants of an individual that was in class k at time t before the present and had an offspring born in class ℓ between times t and $t-1$. For this, we need to distinguish between descendants of younger, older and same-litter sisters of the $(g-1)$-th ancestor of *ego* (see Figure 3).

Descendants of younger sisters of the $(g-1)$-th ancestor

Between times t and $t-1$ before the present, the g-th ancestor of *ego* survives from class k to class m with probability s_{mk}. From there, the expected number of q-th generation descendants that it leaves in class i after the remaining $t-1$ time steps is $d_{im}(t,g)$, the (i,m)-th entry of the matrix $\mathbf{D}(t,g)$. The expected number of class-i (g,q)-kin of *ego* through younger sisters of its $(g-1)$-th ancestor is, therefore, $\sum_{t\geq0}\sum_{k,\ell,m}d_{im}(t,q)\,s_{mk}\,u_{j\ell}(t,g)\,p_{\mathbf{F}}(\ell,k)$ —which, in matrix notation, is the (i,j)-th entry of

$$\mathbf{K}_y(g,q)=\sum_{t\geq1}\mathbf{D}(t,q)\,\mathbf{S}\,\mathbf{P}_{\mathbf{F}}^{\mathrm{T}}\mathbf{U}(t,g)^{\mathrm{T}}.\qquad(6)$$

Descendants of older sisters of the $(g-1)$-th ancestor

To compute the expected number of descendants of older sisters of the $(g-1)$-th ancestor of *ego*, we have to look at the g-th ancestor at times $t+\alpha$ before the present, for each $\alpha\geq1$, to see how many older sisters were born then
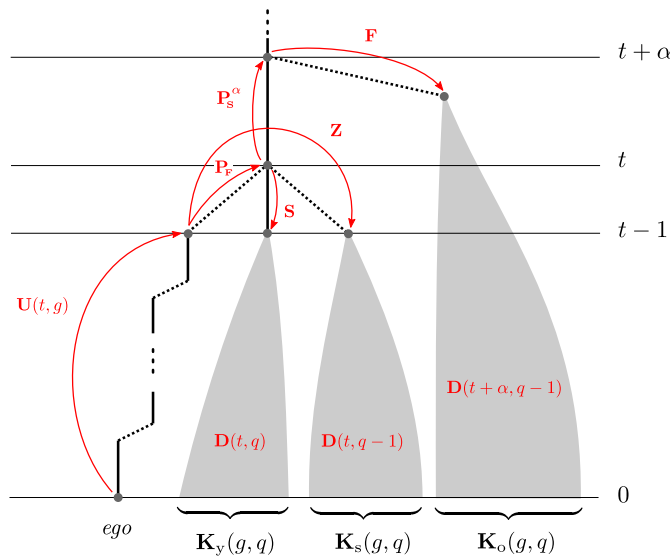

FIGURE 3 Genealogy of the population, showing the various contributions to the kinship matrix $\mathbf{K}(g,q)$. On the left is the lineage of *ego*, up to its g-th ancestor. The solid black lines correspond to survival and the dotted ones to reproduction. On the right are the q-th descendants of the g-th ancestor of *ego*, grouped according to whether they are descendants of younger, same-litter or older sisters of the $(g-1)-th$ ancestor of *ego*. The weights of the red arrows have to be multiplied along a path to get the corresponding contribution to the expected number of (g,q)-kin (matrix transposes have been omitted for simplicity)

and how many $(q-1)$-th generation descendants each of these left. Using the same reasoning as before, we get:

$$\text{For } g \geq 1, \mathbf{K}_{\mathrm{o}}(g, q) = \sum_{t \geq 1} \sum_{\alpha \geq 1} \mathbf{D}(t+\alpha, q-1) \mathbf{F} \left(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}}\right)^{\alpha} \mathbf{P}_{\mathbf{F}}^{\mathrm{T}} \mathbf{U}(t, g)^{\mathrm{T}}. \quad (7)$$

Descendants of same-litter sisters of the $(g-1)$-th ancestor

Unlike for descendants of older and younger sisters, for which knowing about the g-th ancestor of *ego* was sufficient, here we also need to know about its $(g-1)$-th ancestor. This is because knowing that the g-th ancestor gave birth to the $(g-1)$-th one t time steps before the present biases its number of offspring for this reproductive event. To understand this bias, we have to know the class in which the $(g-1)$-th ancestor was born. Indeed, if we sample an individual uniformly at random in class i and then look at its mother, then the mother is sampled proportionally to its number of offspring in class i. Let F_{ij} have the distribution of the number of offspring of class i of an individual of class j, and let $F_{ij}^{(\ell)}$ have the distribution of F_{ij} biased by $F_{\ell j}$, that is, $\mathbb{P}(F_{1j}^{(\ell)} = x_1, ..., F_{Nj}^{(\ell)} = x_N) = \frac{x_\ell}{f_{\ell j}} \mathbb{P}\left(F_{1j} = x_1, ..., F_{Nj} = x_N\right)$. Finally, let $\tilde{f}_{ij}^{(\ell)} = \mathbb{E}(F_{ij}^{(\ell)}) - \mathbb{1}_{\{i=\ell\}}$ (where $\mathbb{1}_{\{i=\ell\}} = 1$ if $i = \ell$ and 0 otherwise). With these quantities, the expected number of class-i (g, q)-kin of *ego* through same-litter sisters of its $(g-1)$-th ancestor is

$$\sum_{t \geq 1} \sum_{k, \ell, m} d_{im}(t, q-1) \tilde{f}_{mk}^{(\ell)} p_{\mathbf{F}}(\ell, k) u_{j\ell}(t, g)$$
$$= \sum_{t \geq 0} \sum_{\ell, m} d_{im}(t, q-1) z_{m\ell} u_{j\ell}(t, g),$$

where $z_{m\ell} = \sum_k \tilde{f}_{mk}^{(\ell)} p_{\mathbf{F}}(\ell, k)$. In matrix form, this is the (i, j)-th entry of

$$\mathbf{K}_{\mathrm{s}}(g, q) = \sum_{t \geq 1} \mathbf{D}(t, q-1) \mathbf{Z} \mathbf{U}(t, g)^{\mathrm{T}}. \quad (8)$$

What makes Equation (8) relevant is that the matrix $\mathbf{Z}$ has a concrete interpretation: as mentioned at the beginning of this section, z_{ij} is the expected number of same-litter sisters of a focal individual in class j that have just been born in class i. Thus, the matrix $\mathbf{Z}$ can be estimated from detailed census data. Alternatively, $\mathbf{Z}$ can be computed from the matrices $\mathbf{S}$ and $\mathbf{F}$ and the covariances $\text{Cov}\left(F_{ij}, F_{\ell j}\right)$. Indeed, elementary calculations give

$$\tilde{f}_{ij}^{(\ell)} = \text{Cov}\left(F_{ij}, F_{\ell j}\right) / f_{\ell j} + f_{\ell j} - \mathbb{1}_{\{i=\ell\}}, \quad (9)$$

from which $\mathbf{Z}$ can then be computed. A few special cases are worth pointing out. First, when at most one offspring is produced during a projection interval we have $\mathbf{Z} = \mathbf{0}$.

Second, if there is a single newborn class, say class 1, then $z_{ij} = 0$ if $i \neq 1$ or $j \neq 1$, and

$$z_{11} = \sum_k \left(\text{Var}\left(F_{1k}\right)/f_{1k} + f_{1k} - 1\right) p_{\mathbf{F}}(1, k). \quad (10)$$

This covers the majority of the models used in practice. Third, when the numbers of offspring produced by an individual in the various classes are independent Poisson variables,

$$\mathbf{Z} = \mathbf{F}\mathbf{P}_{\mathbf{F}}^{\mathrm{T}} \quad (11)$$

This formula can be used by default when the covariances of the fertilities are not available.

Expression of the kinship matrix

Combining Equations (6), (7) and (8), we get the following expression for the kinship matrix $\mathbf{K} = \mathbf{K}_{\mathrm{y}} + \mathbf{K}_{\mathrm{o}} + \mathbf{K}_{\mathrm{s}}$, which is valid for all $q \geq 0$ but for $g \geq 1$ only:

$$\mathbf{K}(g, q) = \sum_{t \geq 1} \left(\mathbf{D}(t, q) \mathbf{S} \mathbf{P}_{\mathbf{F}}^{\mathrm{T}} + \mathbf{D}(t, q-1) \mathbf{Z} \right.$$
$$\left. + \sum_{\alpha \geq 1} \mathbf{D}(t+\alpha, q-1) \mathbf{F} \left(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}}\right)^{\alpha} \mathbf{P}_{\mathbf{F}}^{\mathrm{T}}\right) \mathbf{U}(t, g)^{\mathrm{T}}.$$

For $g = 0$, the (g, q)-kin of *ego* are its q-th descendants (and *ego* itself when $q = 0$). Thus, $\mathbf{K}(0, 0) = \mathbf{I}$ and, for $q > 0$,

$$\mathbf{K}(0, q) = \sum_{\alpha \geq 1} \mathbf{D}(\alpha, q-1) \mathbf{F} \left(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}}\right)^{\alpha}. \quad (12)$$

Note that, compared to Equation (7), there is no factor $\mathbf{P}_{\mathbf{F}}^{\mathrm{T}}$ to the right of $(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}})^{\alpha}$.

In order to get a unique expression of $\mathbf{K}(g, q)$ that works for all $g \geq 0$, we let $\mathbf{U}(t, 0) = 0$ for $t \geq 1$, in agreement with Equation (4) and the definition of weak compositions in Equation (3); but for $t = 0$ we use the special convention $\mathbf{U}(0, 0) = \mathbf{I}$. Setting $\mathbf{Y}(g) = \mathbf{I}$ if $g = 0$ and $\mathbf{P}_{\mathbf{F}}^{\mathrm{T}}$ otherwise, this gives the following general expression for the kinship matrix, valid for any $(g, q) \neq (0, 0)$:

$$\mathbf{K}(g, q) = \sum_{t \geq 0} \left(\mathbf{D}(t, q) \mathbf{S} \mathbf{P}_{\mathbf{F}}^{\mathrm{T}} + \mathbf{D}(t, q-1) \mathbf{Z} \right.$$
$$\left. + \sum_{\alpha \geq 1} \mathbf{D}(t+\alpha, q-1) \mathbf{F} \left(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}}\right)^{\alpha} \mathbf{Y}(g)\right) \mathbf{U}(t, g)^{\mathrm{T}}. \quad (13)$$

Note that for $q = 0$, this formula reduces to

$$\mathbf{K}(g, 0) = \sum_{t \geq 1} \mathbf{S}^t \mathbf{P}_{\mathbf{F}}^{\mathrm{T}} \mathbf{U}(t, g)^{\mathrm{T}} \quad (14)$$

and gives the probability that the g-th ancestor of *ego* is alive, structured by the class of *ego* and of the g-th ancestor.

Computing the kinship matrix in practice

Equation (13) gives the kinship matrix as an infinite sum. However, in numerical applications it will have to be approximated by a finite sum, for instance, of the form

$$\widehat{\mathbf{K}}\left(g, q; t_{\max}, \alpha_{\max}\right) = \sum_{t=0}^{t_{\max}} \left(\mathbf{D}\left(t, q\right) \mathbf{S} \mathbf{P}_{\mathbf{F}}^{\mathrm{T}} + \mathbf{D}\left(t, q-1\right) \mathbf{Z} \right.$$

$$\left. + \sum_{\alpha=1}^{\alpha_{\max}} \mathbf{D}\left(t+\alpha, q-1\right) \mathbf{F}\left(\mathbf{P}_{\mathbf{S}}^{\mathrm{T}}\right)^{\alpha} \mathbf{Y}\left(g\right) \right) \mathbf{U}(t, g)^{\mathrm{T}}. \tag{15}$$

This raises the question of determining how many terms to sum, and of the precision of the resulting estimate. In SM.6, we give an explicit upper bound on $\left\| \mathbf{K}\left(g, q\right) - \widehat{\mathbf{K}}\left(g, q; t_{\max}, \alpha_{\max}\right) \right\|$ that can be used to chose $t_{\max}$ and $\alpha_{\max}$ so as to meet a target accuracy in the estimation. A procedure to do so automatically is implemented in our code (see SM.1). It should be pointed out, however, that the upper bound of SM.6 is rather crude and that the actual precision will typically be much better than it suggests.

Importantly, when there exists a maximal age ω in the model (i.e. ω such that $\mathbf{S}^{\omega-1} \neq \mathbf{0}$ and $\mathbf{S}^{\omega} = \mathbf{0}$), the sums in Equation (13) are, in fact, finite: indeed, in that case, $\mathbf{U}(t, g) = \mathbf{0}$ for all $t > g\omega$; $\mathbf{D}(t, g) = \mathbf{0}$ for all $t > (q+1)\omega$ and $\mathbf{P}_{\mathbf{S}}^{\alpha} = \mathbf{0}$ for all $\alpha \geq \omega$. As a result,

$$\mathbf{K}\left(g, q\right) = \widehat{\mathbf{K}}\left(g, q; t_{\max}, \alpha_{\max}\right),$$
$$\text{with} \begin{cases} t_{\max} = \min\left(g\omega, (q+1)\omega\right) \\ \alpha_{\max} = \omega - 1 \end{cases}. \tag{16}$$

This gives a useful heuristic to choose $t_{\max}$ and $\alpha_{\max}$ when the values obtained from the upper bound of SM.6 lead to prohibitively long calculations: first, find a reasonable estimate of the maximum age that one can expect to observe in the population, by picking $\widehat{\omega}$ such that the probability that an individual reaches age $\widehat{\omega}$ is less than some very small ε (e.g. by taking $\widehat{\omega} = \inf\left\{ t \geq 0 : \| \mathbf{S}^{t-1} \|_1 < \varepsilon \right\}$, where $\|\mathbf{X}\|_1 = \sup_j \sum_i |x_{ij}|$); then use the values of $t_{\max}$ and $\alpha_{\max}$ obtained by plugging $\widehat{\omega}$ in Equation (16).

Although there is no mathematical guarantee—that is, other than that given in SM.6—on the difference between $\mathbf{K}$ and $\widehat{\mathbf{K}}$ when using $t_{\max} = \min\left(g\widehat{\omega}, (q+1)\widehat{\omega}\right)$ and $\alpha_{\max} = \widehat{\omega} - 1$, this difference should be well within the modelling error of the projection model, as it will exclusively be due to individuals that live past $\widehat{\omega}$—which are allowed by the model but may not exist in practice. One way to test this is to compare the results obtained with the projection matrix $\mathbf{A}$ to those obtained with a model $\overline{\mathbf{A}}\left(\widehat{\omega}\right)$ where the maximal age has been capped (as discussed in SM.7). The biological relevance of $\overline{\mathbf{A}}\left(\widehat{\omega}\right)$ compared to that of $\mathbf{A}$ is something that can be assessed independently; and once this is done, the kinship matrix of $\overline{\mathbf{A}}\left(\widehat{\omega}\right)$ can be computed exactly using $t_{\max} = \min\left(g\widehat{\omega}, (q+1)\widehat{\omega}\right)$ and $\alpha_{\max} = \widehat{\omega} - 1$.

Aggregated measures of kinship and relatedness

Having a general expression for the number of any type of kin makes it possible to consider various aggregated measures of the kinship structure of a population. For instance, the matrix

$$\widetilde{\mathbf{K}}(d) = \sum_{g+q=d} \mathbf{K}\left(g, q\right), \tag{17}$$

gives the number of kin that are at distance d from *ego* in their genealogical tree. If $\lambda > 1$, $\widetilde{\mathbf{K}}(d)$ diverges as $d \to \infty$, whereas if $\lambda < 1$ it goes to zero. It is, thus, natural to wonder how $\|\widetilde{\mathbf{K}}(d)\|$ grows/decreases as a function of d. In SM.9, we show that it grows like θ_c^d, where

$$\theta_c = \inf\{\theta > 0 : \rho\left(\mathbf{S} + \theta^{-1}\mathbf{F}\right) \rho\left(\mathbf{P}_{\mathbf{S}} + \theta^{-1}\mathbf{P}_{\mathbf{F}}\right) < 1\} \tag{18}$$

and where ρ denotes the spectral radius. This characterisation makes it straightforward to compute θ_c numerically. In SM.9, we also explain why we expect to have

$$\theta_c \approx \frac{\lambda - s^\star}{\sqrt{\lambda - s^\star}} \text{ with } s^\star = \mathbf{v}\mathbf{S}\mathbf{w}, \tag{19}$$

where the vector of reproductive values $\mathbf{v}$ and the stable distribution $\mathbf{w}$ are such that $\mathbf{v}\mathbf{w} = 1$.

One of the notable feature of θ_c is that it describes the global kinship structure, at a large scale. This contrasts with the matrices $\widetilde{\mathbf{K}}(d)$, which can only be computed for relatively small values of d (see SM.10) and thus only give us access to the local, small-scale structure of the kinship.

In applications, θ_c could be a useful descriptor for questions ranging from life-history theory to inclusive fitness: for instance, Equation (19) shows that for a fixed value of $\lambda > 1$, the number of kin is maximised by increasing $s^\star$, that is, by K-strategies; but that if $\lambda < 1$ it is optimised by decreasing $s^\star$, that is, by r-strategies. The statistic θ_c could, thus, be used to compare life-history patterns in terms of the $r - K$ continuum, as an alternative to the generation time $T = 1/\left(\lambda - s^\star\right)$ (Bienvenu & Legendre, 2015). In the context of inclusive fitness, since the matrices $\mathbf{K}$ can be interpreted as measures of relatedness (see SM.11), θ_c is an indicator of how much relatedness is carried by various kin: for instance, the higher θ_c, the more relatedness is carried by close kin—but the smaller the relative contribution of close kin compared to that of distant ones.

Of course, $\widetilde{\mathbf{K}}(d)$ and θ_c are only two examples of aggregated measures of kinship that can be defined from the matrices $\mathbf{K}$. In general, one may want to consider various statistics of the form $\sum_{g,q} f(\mathbf{K}(g, q), g, q)$, for functions f and values of (g, q) that depend on the specific application. If the application requires summing on large values of g and q, this may not be possible because of prohibitive computation times. In that case, one possible strategy could be to first turn the projection matrix into a 1×1 matrix projecting the unstructured population (see

Bienvenu et al., 2017; Coste et al., 2017), and then use the efficiently computable formulas of $\mathbf{K}(g, q)$ for 1×1 models given in SM.8.

DISCUSSION

In this article, we have developed a unified framework to compute the expected number of kin of a focal individual (as a function of the vital traits of the focal individual and of its kin), directly from the demographic rates of a population. The main feature of our method is that it applies to any kin relationship (mother–daughter, aunt–niece, cousin–cousin, etc) and any structured population described by a matrix population model—be it structured by age (as illustrated in Box 2), by stage (SM.2.1), by patch (SM.2.3) or by any conceivable type of discrete class. These theoretical results come with readily usable implementations that make them available to any ecologist. Here, we discuss the main limitations and implications of this unified framework, and give some perspectives for future research.

Conceptual and practical limitations

Our work is set in the framework of matrix population models, and thus inherits its usual limitations. In particular, density dependence and environmental stochasticity—which are not considered by our input model—are likely to be two of the main sources of discrepancies between our theoretical predictions and reality. However, because our mathematical reasoning is based directly on the stochastic model underlying the matrix population model (rather than on its deterministic description of the average population dynamics), demographic stochasticity is not a limitation: it is taken into account in our calculations, and although our final formulas correspond to expected values with respect to this stochasticity, our methods could also be used to compute, for example, variances. This is discussed below.

Another important limitation of our approach is that it hinges on the assumption that the dynamics of the population has been at the steady state for an extended period of time. This limits the validity of our formulas in populations whose demography has been perturbed in the recent past—all the more so when studying distant kin relationships. Unfortunately, there does not seem to be an easy way to circumvent this using our techniques, and there may not be alternatives to recursive formulas for such models (Caswell & Song, 2021).

When applying our results, one should also pay attention to the fact that they are formulated for models where the fertilities are independent from adult survival. This assumption is required for the connection with the stochastic population model that implicitly underlies our calculations, but it is not compatible with post-breeding models. However, these can (usually easily) be reformulated as pre-breeding models, so this is not a major limitation of our work.

As implicit with matrix population models, we work in the context of monoparental genealogies. This setting could correspond either to a one-sex species or to a two-sex species with a female-based demographic model. In the latter case, it should be noted that our formulas are based on a notion of kinship that only considers matrilines, and will thus differ from the traditional notion of kinship in a complete pedigree: for instance, the daughters of the brothers of the focal individual will not be counted as its nieces. Although this is not a limitation of our results, this is something important to have in mind when interpreting them. Incidentally, this specific notion of kinship can, in some contexts, be interpreted as a direct measure of relatedness. Indeed, as explained in SM.11, in an idealised setting (namely, a large panmictic population where males and females have similar life histories and where no two individuals have the same two parents), discarding kin that are not based on matrilines has the same effect as the dilution of genetic material over successive generations. Thus, although our results were first and foremost developed to study kinship, they could have wider ranging implications in evolutionary biology, by working as a measure of relatedness.

Finally, a practical limitation of our results is that the computing time of our formulas increases rapidly when we consider increasingly distant kin, and can become prohibitively long—especially for organisms with a "slow" strategy on the slow-fast continuum. Because close kin are what matter in most situations, this is, in general, hardly going to be a problem. However, should very distant kin be of interest, we provide two partial solutions to prohibitive computing times: first, we use our results to derive an efficiently computable statistic to describe the large-scale kinship structure of a population; second, if one is ready to discard the structure of the population, then in SM.8 we give a simple formula for the expected number of kin that is trivial to compute in practice. Note that both of these results are made possible by the fact that we have a generic formula for the kinship, and would presumably not have been possible using ad hoc formulas for each type of kin.

Implications and perspectives

First and foremost, having a generic formula for arbitrary kin relationship makes it possible—and elementary—to define tailor-made measures of kinship to be used in a specific setting. For instance, anthropologists who would like to compare the number of lineal versus collateral kin can easily do so by summing the kinship matrices of their choice—or, more generally, combine them however they want to produce the best suited metric for their particular application.

Our methods can be used to compute, at least numerically, the sensitivity or the elasticity of the kinship structure to the traits implemented in the model. This is illustrated on a concrete example in Box 2. Because our formulas apply to any primitive projection matrix, they can be used to study models where classes are based on several traits, such as metapopulation models (Lebreton & Gonzáles-Dávila, 1993) or more general multitrait models (Coste et al., 2017; Coste & Pavard, 2020; Roth & Caswell, 2016). This makes it possible to study and compare the relative effects of different traits on the kinship. We illustrate this in SM.2.3, where we use a *stage × patch* model to study the effect of dispersal.

Finally, because our results make new use of classic tools to solve a long-standing problem, they could spur further theoretical developments in the field of matrix population models. An example of this is given by our introduction, in SM.5, of the previous-generation matrix (which is a dual version of the classic next-generation matrix of Cushing & Zhou, 1994) and of the Euler–Lotka matrices (which give two simple generalisations of the Euler–Lotka equation to arbitrarily structured populations). Despite being very natural from a mathematical and biological point of view, these quantities do not seem to have been identified as such before (although similar ideas can be found in Lebreton, 2005). Though not directly linked to kinship, these results are straightforward consequences of the framework we have developed to study it.

Our work offers several perspectives in terms of extensions, in particular to see if our methods can be extended beyond the framework of matrix population models. A natural candidate for these are integral projection models (Rees et al., 2014). Because the variations in the individual realisations of the vital rates have a relevant impact on kin structure (Coresh & Goldman, 1988; Tuljapurkar et al., 2020, 2021), another perspective is to go beyond the "expected-value" description of the kinship structure given here—namely, by deriving statistics that describe the demographic stochasticity underlying our calculations, instead of averaging it out. For instance, with our method, the classic results of Everett and Ulam (1948) on multitype Galton–Watson processes could be used to compute the variance of the number of kin. Finally, an important line of research that we plan to explore is the extension of our work to two-sex models. This is particularly challenging, as it requires leaving the framework of monoparental genealogies and will, therefore, have to be done using a combination of analytical and numerical methods.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13854.

OPEN RESEARCH BADGES

This article has earned Open Data badges. Data and materials are available at https://dsc.duq.edu/.

ORCID
Christophe F. D. Coste https://orcid.org/0000-0003-3680-5049
François Bienvenu https://orcid.org/0000-0002-5396-1193
Victor Ronget https://orcid.org/0000-0001-9991-0482
Juan-Pablo Ramirez-Loza https://orcid.org/0000-0002-9959-3592
Sarah Cubaynes https://orcid.org/0000-0002-3935-9825
Samuel Pavard https://orcid.org/0000-0002-6803-8123

REFERENCES
Atkins, J.R. (1974) On the fundamental consanguineal numbers and their structural basis. *American Ethnologist*, 1(1), 1–31.

Barrai, I., Cavalli-Sforza, L. & Moroni, A. (1962) Frequencies of pedigrees of consanguineous marriages and mating structure of the population. *Annals of Human Genetics*, 25(4), 347–377.

Bienvenu, F., Akçay, E., Legendre, S. & McCandlish, D.M. (2017) The genealogical decomposition of a matrix population model with applications to the aggregation of stages. *Theoretical Population Biology*, 115, 69–80.

Bienvenu, F. & Legendre, S. (2015) A new approach to the generation time in matrix population models. *The American Naturalist*, 185(6), 834–843.

Caswell, H. (2019) The formal demography of kinship: A matrix formulation. *Demographic Research*, 41, 679–712.

Caswell, H. (2020) The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Research*, 42, 1097–1146.

Caswell, H. & Song, X. (2021). The formal demography of kinship iii: Kinship dynamics with time varying demographic rates. *bioRxiv*.

Coresh, J. & Goldman, N. (1988) The effect of variability in the fertility schedule on numbers of kin. *Mathematical Population Studies*, 1(2), 137–156.

Coste, C.F.D., Austerlitz, F. & Pavard, S. (2017) Trait level analysis of multitrait population projection matrices. *Theoretical Population Biology*, 116, 47–58.

Coste, C.F.D. & Pavard, S. (2020) Analysis of a multitrait population projection matrix reveals the evolutionary and demographic effects of a life history trade-off. *Ecological Modelling*, 418, 108915.

Cushing, J.M. & Zhou, Y. (1994) The net reproductive value and stability in matrix population models. *Natural Resource Modeling*, 8(4), 297–333.

Demetrius, L. (1974) Demographic parameters and natural selection. *Proceedings of the National Academy of Sciences*, 71(12), 4645–4647.

Demetrius, L. (1975) Natural selection and age structured populations. *Genetics*, 79(3), 535–544.

Everett, C.J. & Ulam, S. (1948) Multiplicative systems, I. *Proceedings of the National Academy of Sciences*, 34(8), 403–405.

Goodman, L.A., Keyfitz, N. & Pullum, T.W. (1974) Family formation and the frequency of various kinship relationships. *Theoretical Population Biology*, 5(1), 1–27.

Hajnal, J. (1963) Concepts of random mating and the frequency of consanguineous marriages. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 159(974), 125–177.

Joffe, A. & Waugh, W.A.O. (1985) The kin number problem in a multitype Galton-Watson population. *Journal of Applied Probability*, 22(1), 37–47.

Le Bras, H. (1973) Parents, grands-parents bisaieux. *Population (French Edition)*, 1(1), 9–38.

Lebreton, J.D. (2005) Age, stages, and the role of generation time in matrix models. *Ecological Modelling*, 188(1), 22–29.

Lebreton, J.-D. & Gonzáles-Dávila, G. (1993) An introduction to models of subdivided populations. *Journal of Biological Systems*, 01(04), 389–423.

Lotka, A.J. (1939) *Théorie analytique des associations biologiques*. Paris: Hermann et Cie.

Luo, J. & Fox, B.J. (1990) Life-table comparisons between two ground squirrels. *Journal of Mammalogy*, 71(3), 364–370.

Nerman, O. & Jagers, P. (1984) The stable doubly infinite pedigree process of supercritical branching populations. *Zeitschrift fuer Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3), 445–460.

Pavard, S. & Coste, C.F.D. (2020) Goodman, Keyfitz & Pullum (1974) and the population frequencies of kinship relationships. *Theoretical Population Biology*, 133, 15–16.

Pullum, T.W. (1982) The eventual frequencies of kin in a stable population. *Demography*, 19(4), 549–565.

Rees, M., Childs, D.Z. & Ellner, S.P. (2014) Building integral projection models: A user's guide. *Journal of Animal Ecology*, 83(3), 528–545.

Roth, G. & Caswell, H. (2016) Hyperstate matrix models: extending demographic state spaces to higher dimensions. *Methods in Ecology and Evolution*, 7(12), 1438–1450.

Tuljapurkar, S.D. (1982) Population dynamics in variable environments. IV. Weak ergodicity in the Lotka Equation. *Journal of Mathematical Biology*, 14, 221–230.

Tuljapurkar, S. (1993) Entropy and convergence in dynamics and demography. *Journal of Mathematical Biology*, 31(3), 253–271.

Tuljapurkar, S., Zuo, W., Coulson, T., Horvitz, C. & Gaillard, J.M. (2020) Skewed distributions of lifetime reproductive success: Beyond mean and variance. *Ecology Letters*, 23(4):ele.13467.

Tuljapurkar, S., Zuo, W., Coulson, T., Horvitz, C. & Gaillard, J.M. (2021) Distributions of LRS in varying environments. *Ecology Letters*, 1328–1340.

Waugh, W.A.O. (1981) Application of the Galton-Watson process to the kin number problem. *Advances in Applied Probability*, 13(4), 631–649.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.