The Kinship Formula:
inferring the numbers of all kin from any structured population projection model

Christophe F. D. Coste[1,2,3]

[1]Biosciences Department, University of Swansea, SA1 8EN, UK
[2]Centre for Biodiversity Dynamics, Department of Biology, NTNU, NO-7491, Trondheim, Norway
[3]Unité Eco-anthropologie (EA), Muséum National d'Histoire Naturelle, CNRS, Université Paris Diderot, F-75016, Paris, France.

Abstract

Structured population projection models are fundamental to many fields of science. They enable abundance forecasting for populations categorized by various traits such as age (for demography), patch (for spatial ecology), genotype (for genetics), infectious stage (for epidemiology) or capital (economics). The demography of a structured population, determined by the transition rates (e.g., survival, fertility) between its various states, also shapes its relatedness – or kinship – structure. This structure (a probabilistic genealogy) is crucial for understanding how individuals are related to the rest of the population and affects effective population size, inclusive fitness, inbreeding, pedigrees, relatedness, familial structures, etc. Despite its significance, the relationship between demography and kinship remains under-explored. By incorporating the generation number as a trait into the population structure, we derive the Kinship Formula, yielding the expected number of any kin for any structured population. This formula is simple to implement and fast to compute, even for complex models. Most importantly, it promises significant theoretical advances. From the Kinship Formula, one can, for instance, assess the impact of embedded processes (e.g., dispersal, inheritance, growth) on kinship, compute mean population relatedness and the eventual number of kin (including kin already dead or not born yet). The Kinship Formula derived here stems from a one-sex constant environment framework. Its simplicity should allow for extensions to include environmental and demographic stochasticity as well as two-sex models.

Corresponding author:
Christophe F.D. Coste
E-mail: `c.f.d.coste@swansea.ac.uk`
Telephone: +33 6 32 51 33 11
Department of Biosciences
Swansea University, Singleton Park Campus Swansea, SA2 8PP, UK.
ORCID identifier: `https://orcid.org/0000-0003-3680-5049`

Data accessibility statement: No new data are used. All data used to illustrate the method are public. The code related to the main text (R and Matlab) can be found in Supplementary Materials (S.M.VIII)

Introduction

Structured population projection models are crucial for many scientific fields. They enable the projection, over time, of populations categorised by one or several traits. This can be, for instance, a population of humans structured by age, sparrows by stage and body mass, susceptible individuals – in an epidemic – structured by infectious status, households by age and capital, drosophilas by genotype, cells by cancerous status, meta-populations by geographical patches. From the transition rates between all pair of states, one can predict population dynamics, future trait distribution, extinction probability of the population, and more. Such analyses are the bread and butter of ecology, epidemiology, genetics, anthropology, micro-economics and certain areas of physics.

The structured demography of a population not only informs about its future but also sheds light on its past history and current kinship (or relatedness) structure. The number of kin (or collaterals) an individual has is determined by its own reproduction, as well as the survival and reproduction of its ancestors and their descendants (1). Understanding the influence of population dynamics on relatedness structure is crucial across many fields and central to the emerging field of Kinship Demography (2). For example, in genetics, kinship structure affects inbreeding levels and effective population size (3; 4). In behavioral ecology, it influences how individuals may compete or cooperate and it determines their inclusive fitness (5; 6; 7). In anthropology and micro-economics, it combines with cultural norms to define familial structures (8; 9; 10). In epidemiology, the "kin" – individuals infected by a common source – can form clusters which are crucial for epidemic predictions (see, e.g., 11; 12; 13; 14). In population ecology and demography, the ability to predict kinship structures becomes increasingly crucial with the rising availability of pedigrees and genealogies (see, e.g., 15; 16; 17).

For simple, unstructured models with non-overlapping generations, population geneticists have developed tools to study the distributions of ancestors and collaterals (18; 19; 20; 21; 22). For age-structured, generation-overlapping populations, Le Bras published a method to trace probabilistic ancestry (23). Goodman, Keyfitz and Pullum extended this work (24) to provide the expected number of a few kin (ancestors *and* collaterals: sisters, aunts, mother, grand-mother, daughters and cousins) in a landmark study (2; 25). However this method – together with its matrix form (26) – focuses solely on age-structured populations and treats each kin type separately, which reduces its generalizability. Researchers interested in populations structured by traits other than age have therefore often resorted to simulations (or micro-simulations, via, e.g., the SocSim software, see 27; 28). Because of these limitations, this approach hinders Kinship Demography from fully exploring the impacts of specific processes (demographic, ecological, cultural, genetic) on relatedness, inbreeding or familial structure. It does not allow, for instance, to understand how dispersal, growth or inheritance affect the kinship structure of a population. To address this gap, Coste and collaborators recently published a closed-form formula for predicting the number

of any kin within any structured population (29). While this formula represents significant progress for Kinship Demography, it lacks the simplicity and generalizability needed to be the theoretical bridge between demography and kinship that has been called for by demographers (24), mathematicians (30) and evolutionary ecologists (2). Although more general than the aforementioned alternative approaches, it remains bulky and complex to analyse theoretically.

Here, we introduce a model that structures the population both by the original traits and by generation. This leads to a substantial simplification of Coste et al's formulation (29). By leveraging the genealogical Markov chains – developed by Demetrius (31; 32) and studied by Tuljapurkar (33; 34) – and the generation structure, we derive a simple formula (the Kinship Formula, eq.3) for the expected number of any kin from any structured population projection model. The Kinship Formula applies to close but also (very) distant kin and provides the structure of the kin network (e.g., their spatial dispersion or age distribution). Unlike other approaches, it enables the exact computation of the expected number of kin for a stage-structured population, which we illustrate for a ground squirrel population. The Kinship Formula facilitates rapid computation (even for a very complex model) and should prove crucial for sound extensions to more intricate frameworks (e.g., stochasticity, two-sex models). We demonstrate its utility by calculating the eventual number of kin. Further we show how the unstructured kinship (the number of kin irrespective of their state), when compared with the kinship of the unstructured population, enables analysis of the effects of traits on kinship. Finally we provide a link between the individual measure of the number of kin and the population measure of relatedness.

The Framework

Consider a general, structured and generation-overlapping, population. The structure may consist of one or several traits (see S.M.I, for the concept of traits in projection models). We limit ourselves to one-sex models here and to processes that are constant in time. To be as general as possible, we consider a discrete structure and a discrete time-step (see S.M.II). The discrete structure consists of s classes (or states). This population can be projected over time, via $\tilde{M}_{i,j}$, the stochastic process yielding the number of individuals in class i stemming from an individual in class j, over the chosen time-step. Process $\tilde{M}_{i,j}$ can be decomposed into a survival ($\tilde{S}_{i,j}$, "producing" the same individual) and a fertility ($\tilde{F}_{i,j}$, producing new individuals) components: $\tilde{M}_{i,j} = \tilde{F}_{i,j} + \tilde{S}_{i,j}$. The deterministic expectations of this process – we denote $M_{i,j} = E(\tilde{M}_{i,j}) = E(\tilde{F}_{i,j}) + E(\tilde{S}_{i,j}) = F_{i,j} + S_{i,j}$ – projects the deterministic population vector $\boldsymbol{n}$ over time: $n_i(t+1) = \sum_j M_{i,j} n_j(t)$, which, in matrix form, is $\boldsymbol{n}(t+1) = \mathbf{M}\boldsymbol{n}(t) = (\mathbf{F} + \mathbf{S})\boldsymbol{n}(t)$. Matrix $\mathbf{M}$ is the so-called population projection matrix (35), which elements, the $M_{i,j}$ are called the transition rates. We consider, without any loss of generality, that the processes of reproduction and survival are independent and identically distributed (see S.M.III). Notation-wise, stochastic processes are noted with a tilde, and expectations without, vectors (lower case) and matrices (upper case) are in bold while scalars are in

plain text. The population projection matrices $\mathbf{S}$ and $\mathbf{F}$ are the building blocks of the Kinship Formula (eq.3) inferring the expected number of kin (structured by kin type and by class) of focal individual *ego* (categorized by its class). They will allow us, in a generation-structured population (eqs.1 and 2) which makes it possible to consider all kin of *ego* simultaneously, to go back in time towards the distribution of the ancestors of *ego* via their genealogical Markov chains, and from there to count the descendants of these ancestors.

Genealogical Markov Chains At the demographic stable state, the population growth rate and the vector of relative abundances (denoted respectively λ and $\boldsymbol{w}$) correspond to the maximum eigenvalue of $\mathbf{M}$ and its associated right-eigenvector (scaled to sum to 1): $\mathbf{M}\boldsymbol{w} = \lambda\boldsymbol{w}$. While $\mathbf{M}$ projects the population forward in time, its genealogical Markov chain (31; 32) projects it backwards in time. The transition matrix of the genealogical Markov chain associated with $\mathbf{M}$ is $\mathbf{P}$, where $P_{ij} = \frac{M_{ji}\,w_i}{\lambda\,w_j}$ is the probability that the ancestor (via survival or reproduction) of an individual in class j was in class i at the previous time-step (see appendix 1). Similarly to the projection matrix $\mathbf{M}$, the transitions of $\mathbf{P}$ can be decomposed according to survival and fertility: $\mathbf{P} = \mathbf{P_S} + \mathbf{P_F}$ (eq.13, in app.1).

Same-litter daughters and sisters When projecting individuals forward in time, we must consider the number of "same-litter sisters" – the sisters born at the same time than *ego* – which depends on the stochastic fertility process $\tilde{\mathbf{F}}$ (and not only deterministic $\mathbf{F}$), a fact already accounted for by Goodman, Keyfitz and Pullum (24). Another component of the Kinship Formula is therefore the number of "same-litter-daughters" $F^{(j)}_{i,k}$, the expected number of supplemental offspring (in class i) produced by a mother (in class k) knowing she has produced one (in class j) (eq.14 in app.2). In an operation, that mimics, at the level of one generation, the general pattern yielding the Kinship Formula, we generate, from $\mathbf{P_F}$ and $F^{(j)}_{i,k}$, the same-litter sister matrix $\mathbf{Z}$. Its $Z_{i,j}$ element is the number of "same-litter sisters" in class j of *ego* in class i (29) and corresponds to $Z_{i,j} = \sum_k F^{(j)}_{i,k} P_F(k,j)$ (see eq.15). In many cases, in practice, the number of same-litter daughters does not depend on the class of the individual known to have been produced (see below and S.M.IV), which yields a simpler Kinship Formula (eq.8).

Generation structured population We now incorporate an additional trait into the population structure: the *generation number*. We consider a population vector that is 2-dimensional, that is, made of two traits: the first trait corresponds to the s classes of the original structure (which can itself correspond to several traits that we consider combined into one) and the second trait to the *generation number* (see S.M.I). We turn this second dimension, which is in principle relative and infinite, into an absolute and finite metric by setting the generation number of the *most recent common ancestor* of the kin and *ego* to 1 and by providing trait *generation number* with an arbitrary number (g_m) of classes. The multitrait population vector (or tensor) needs to be vectorised (to be made one-dimensional) in order to be projected over time: individuals in class c_i and generation g_i are to be found in the ith entry of $\mathbb{n}$, the new population vector, where $i = c_i + s(g_i - 1)$; we will denote this entry equivalently $\mathbb{n}_{(c_i,g_i)}$ or $\mathbb{n}_i$.

We denote $\mathbb{F}$ and $\mathbb{S}$ the generation-structured matrices projecting $\mathbb{n}$ forward in time and $\mathbb{P}_{\mathbb{F}}$ and $\mathbb{P}_{\mathbb{S}}$ the generation-structured projectors backward in time. Such multitrait population projection matrices take the appearance of block-matrices; here, they are of size $s.g_m \times s.g_m$ made of $g_m \times g_m$ blocks of size $s \times s$. When the trait added on top of an existing structure does not affect the demographic processes, as for trait *generation number* here, the new model can be built simply using the tensor (or Kronecker) product $\otimes$. Here the generation number is constant via survival and increases by 1 via reproduction, and thus

$$\mathbb{S} = \mathbf{I}_{g_m} \otimes \mathbf{S} \quad \text{and} \quad \mathbb{F} = \mathbf{J_1} \otimes \mathbf{F}, \tag{1}$$

with $\mathbf{J_1}$ representing a zero matrix of size $g_m \times g_m$ except for ones on the sub-diagonal and $\mathbf{I}_{g_m}$ the identity matrix of size g_m. Block-matrix $\mathbb{M} = \mathbb{F} + \mathbb{S}$ has therefore $\mathbf{S}$ as its diagonal blocks and $\mathbf{F}$ as its sub-diagonal blocks; for $g_m = 3$, this is:

$$\mathbb{M} = \mathbb{F} + \mathbb{S} \text{ with } \mathbb{F} = \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{F} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{F} & \mathbf{0} \end{array}\right] \text{ and } \mathbb{S} = \left[\begin{array}{c|c|c} \mathbf{S} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{S} \end{array}\right]$$

Similarly, we have

$$\mathbb{P}_{\mathbb{S}} = \mathbf{I}_{g_m} \otimes \mathbf{P_S} \ , \ \mathbb{P}_{\mathbb{F}} = \mathbf{J_1}^T \otimes \mathbf{F} \text{ and } \mathbb{Z} = \mathbf{J_2} \otimes \mathbf{Z}, \tag{2}$$

with $\mathbf{J_2}$ the zero $g_m \times g_m$ zero-matrix but for a 1 in the (2,2) position.

Characterisation of kin relationships and the Kinship Matrix We aim to derive a mathematical object, we call the Kinship Matrix $\mathbb{K}$, that, like $\mathbb{F}$ or $\mathbb{P}_{\mathbb{S}}$, is generation-structured and which $\mathbb{K}((c_i, g_i), (c_j, g_j))$ entry represents the expected number of (c_i, g_i) descendants of the ancestor (in generation 1) of *ego* in state (c_j, g_j). These are referred to as the (g_i, g_j)-kin (in class c_i) of *ego* (in class c_j). This approach provides a complete characterisation of all possible kin relationships: the (g_i, g_j)-kin are the g_i-descendants of the g_j-ancestor of *ego*, where the g_i-descendants lie $g_i - 1$ generations down from the ancestor (e.g., the 3-descendants are the grand-daughters) and the g_j-ancestor is to be found $g_j - 1$ generations up from *ego* (e.g., the 2-ancestor is the mother). Thus, the $(1, 2)$-kin is the mother, the $(3, 1)$-kin the grand-daughters and the $(3, 3)$-kin the (first) cousins. We illustrate this characterisation of kin relationships in fig.1, which corresponds to a figure used for the same purpose in (36; 37; 38) but with an indexation that starts at $(1, 1)$ for *ego* (instead of $(0, 0)$ in these papers). This characterisation corresponds to the various blocks of the Kinship Matrix: the (g_i, g_j)-block of $\mathbb{K}$ corresponds to the number of (g_i, g_j)-kin. In particular, the first block-row of $\mathbb{K}$ corresponds to the ancestors of *ego* (including itself), and each related block-column corresponds to the descendants of that ancestor (see eq.9 for a numerical illustration).

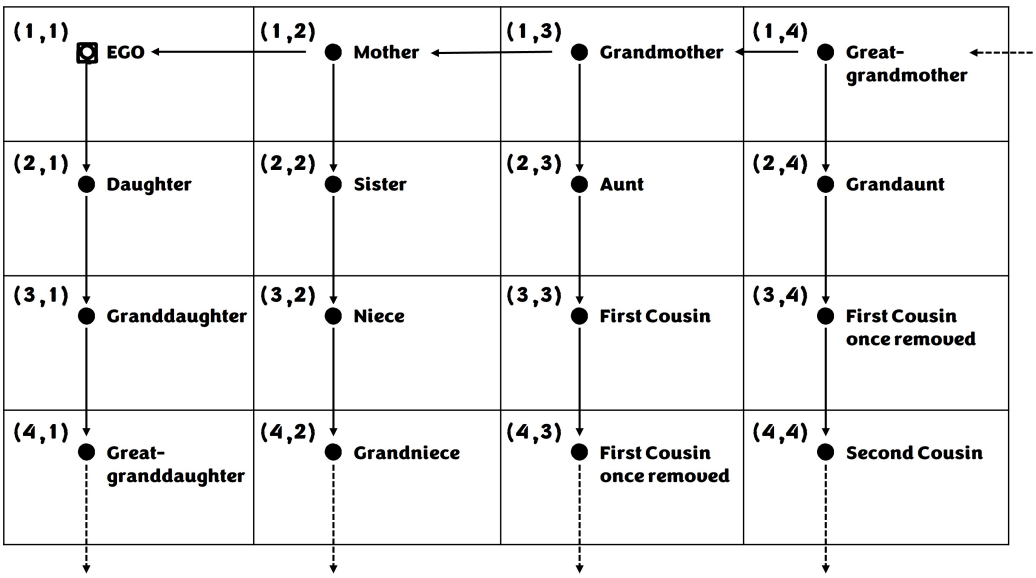

Figure 1: Characterisation of (g_i, g_j)-kin relationships, adapted from (36).

Results

The Kinship Formula

Theorem 1. If $\mathbf{S}$ converges (i.e., if $\lim_{t\to+\infty} \mathbf{S}^t = \mathbf{0}$), the Kinship Matrix $\mathbb{K}$ is the unique solution of

$$\mathbb{K} = \mathbb{MKP} + (\mathbb{FJP}_{\mathbb{S}} + \mathbb{SJP}_{\mathbb{F}} + \mathbb{Z}), \tag{3}$$

where $\mathbb{J} = \mathbf{J_3} \otimes \mathbf{I}_s$ is the zero $sg_m \times sg_m$ block-matrix but for the diagonal of the first block made of 1s and $\mathbf{J_3}$, the zero $g_m \times g_m$ matrix with 1 in the (1,1) position.

We demonstrate, below, that $\mathbb{K}$ is indeed a solution of eq.3 and, in appendix 5, we establish the existence and uniqueness of this solution, provided that $\mathbf{S}$ converges. This fundamental condition for most projection models implies that every individual dies eventually. The formulation of eq.3 does not allow for a direct computation of $\mathbb{K}$ (it can be found on both sides of the equation) which we provide in the next section (eqs.5,6).

proof and interpretation Equation 3 states that the kin of *ego* (the $\mathbb{K}$ on the left-hand side) are the descendants of the kin of *ego*'s immediate ancestor (its ancestor one time-step ago), denoted as $\mathbb{B} = \mathbb{MKP}$, plus additional kin denoted as $\mathbb{A} = (\mathbb{FJP}_{\mathbb{S}} + \mathbb{SJP}_{\mathbb{F}} + \mathbb{Z})$.
Consider $\mathbb{K}((c_i, g_i), (c_j, g_j))$ representing the number of (g_i, g_j)-kin in class c_i of *ego* in class c_j. The immediate ancestor of *ego* is in class c_a and generation g_a with probability $\mathbb{P}((c_a, g_a), (c_j, g_j))$. There are two cases: either *ego*'s immediate ancestor is *ego* itself (if alive at the previous time-step) – then $g_a = g_j$ and $\mathbb{P}((c_a, g_a), (c_j, g_j)) = P_s(c_a, c_j)$ – or *ego* has just been born and *ego*'s immediate ancestor is *ego*'s mother – then $g_a = g_j - 1$ and $\mathbb{P}((c_a, g_a), (c_j, g_j)) = P_f(c_a, c_j)$. Note that if $g_j = 1$, we do not consider *ego*'s mother's kin as the $(g_j, 1)$-kin of *ego* are its direct descendants. At the stable state,

the number of (g_b, g_a)-kin in class c_b of the immediate ancestor (in class c_a) is $\mathbb{K}((c_b, g_b), (c_a, g_a))$. We project the kin of the immediate ancestor towards the present day via $\mathbb{M}$: the expected number of (g_i, g_j)-kin in class c_i for *ego* in class c_j corresponding to the immediate descendants of its immediate ancestor is $\sum_{c_b, c_a} \sum_{g_b=g_i-1}^{g_i} \sum_{g_a=g_j-1}^{g_j} \mathbb{M}((c_i, g_i), (c_b, g_b)) \mathbb{K}((c_b, g_b), (c_a, g_a)) \mathbb{P}((c_a, g_a), (c_j, g_j)) = \mathbb{B}((c_i, g_i), (c_j, g_j))$. If $g_a > 1$, that is, if *ego*'s immediate ancestor is not the most recent common ancestor of *ego* and the kin, then this counts all (g_i, g_j)-kin of *ego* today and $\mathbb{K}((c_i, g_i), (c_j, g_j)) = \mathbb{B}((c_i, g_i), (c_j, g_j))$. If $g_a = 1$ and *ego*'s ancestor is *ego* itself ($g_j = 1$) then $\mathbb{B}$ counts all kin of *ego* today but for those that *ego* has just produced which correspond to $\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}}$. If $g_a = 1$ and *ego*'s ancestor is *ego*'s mother ($g_j = 2$) then $\mathbb{B}$ counts all kin of *ego* today but for those that *ego*'s mother has just produced (*ego*'s same-litter sisters) and *ego*'s mother herself (*ego* is not a kin of *ego*); the sum of these two components corresponds to $\mathbb{Z} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}}$. Overall, we account for these aditional kin via $\mathbb{A} = (\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z})$, where dummy matrix $\mathbb{J}$ ensures that the number of additional kin is non-zero only for $g_a = 1$. Therefore, $\forall c_i, c_j, g_i, g_j, \quad \mathbb{K}((c_i, g_i), (c_j, g_j)) = (\mathbb{A} + \mathbb{B})((c_i, g_i), (c_j, g_j))$, which completes the proof.

The three-way decomposition of $\mathbb{A}$ corresponds to that of kin via older, younger and same-litter sisters implemented already in the foundational works of Kinship Demography (24; 37; 29). Distinguishing between older and younger sisters is essential because the knowledge of the ancestor's status varies when considering potential siblings born before (the ancestor is known to be alive) of after (the ancestor may not be alive) *ego*'s birth. Additionally, the inclusion of the same-litter daughters is necessary due to their aforementioned dependence on the stochastic reproductive process (29). This formulation of the Kinship Formula (eq.3) lays the groundwork for extending the framework to incorporate individual stochasticity and compute the variance and other moments of the number of kin.

Computation of the Kinship Matrix and alternative formulations

We will now present an alternative expression of the Kinship Formula enabling direct computation of $\mathbb{K}$. First, we simplify eq.3 by substituting the components of $\mathbf{P}$ by their expressions as functions of $\mathbf{M}$'s components (see eq.13 in app.1). Let $\mathbf{D}_{\boldsymbol{w}}$ be the diagonal matrix with $\boldsymbol{w}$ on its diagonal and $\mathbb{D}_{\boldsymbol{w}} = \mathbf{I}_{g_m} \otimes \mathbf{D}_{\boldsymbol{w}}$. Letting $Z^{\dagger}_{i,j} = \sum_k w_k \left(Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k}) + F_{i,k}F_{j,k} - \delta(i,j)F_{j,k} \right)$ (eq.16 in app.3), $\mathbb{Z}^{\dagger} = \mathbf{J_2} \otimes \mathbf{Z}^{\dagger}$, and $\mathbb{K}_{\boldsymbol{w}} = \mathbb{K}\mathbb{D}_{\boldsymbol{w}}$, we arrive at an equivalent formulation of the Kinship Formula where the transition matrix of the genealogical Markov chain is not explicit (see appendix 3 for details):

$$\lambda \mathbb{K}_{\boldsymbol{w}} = \mathbb{M}\mathbb{K}_{\boldsymbol{w}}\mathbb{M}^T + (\mathbb{F}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T + \mathbb{S}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T + \mathbb{Z}^{\dagger}), \tag{4}$$

While eq.4 may seem less intuitive than its counterpart eq.3 due to the absence of the genealogical Markov chain, it offers greater analytical tractability. This expression should facilitate the computation of the sensitivity of the kinship structure to the demographic rates and the integration of varying environments. Notably, when expressed as $\mathbb{K}_{\boldsymbol{w}} = \left(\frac{\mathbb{M}}{\sqrt{\lambda}}\right) \mathbb{K}_{\boldsymbol{w}} \left(\frac{\mathbb{M}}{\sqrt{\lambda}}\right)^T + \frac{1}{\lambda}(\mathbb{F}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T + \mathbb{S}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T + \mathbb{Z}^{\dagger})$, equation 4 becomes a

Lyapunov (or Stein) equation (39), extensively studied in signal processing and control theory (40). The use of vectorization to solve Lyapunov equations is a classic technique (41) and we will now proceed with this operation.

Vectorization of the Kinship Formula

Let us now introduce the $\boldsymbol{vec}$ operator which turns a matrix into a vector by "stacking" its columns into a column vector (42). From eq.4, this operator allows to provide an alternative and equivalent version of the Kinship Formula (see apps.4,5):

$$\boldsymbol{vec}\left(\mathbb{K}_{\boldsymbol{w}}\right) = \left(\lambda\mathbf{I} - \left(\mathbb{M}\otimes\mathbb{M}\right)\right)^{-1}\left(\left(\mathbb{S}\otimes\mathbb{F}+\mathbb{F}\otimes\mathbb{S}\right)\boldsymbol{vec}\left(\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\right)+\boldsymbol{vec}\left(\mathbb{Z}^{\dagger}\right)\right). \tag{5}$$

From this, we can readily obtain the Kinship Matrix, using the inverse of the $\boldsymbol{vec}$ operator, denoted $\boldsymbol{vec}^{-1}_{a,b}$, which turns a vector $\boldsymbol{d}$ of size ab into an $a\times b$ matrix such that $\boldsymbol{vec}(\boldsymbol{vec}^{-1}_{a,b}(\boldsymbol{d})) = \boldsymbol{d}$:

$$\mathbb{K} = \boldsymbol{vec}^{-1}_{sg_m,sg_m}\left(\boldsymbol{vec}\left(\mathbb{K}_{\boldsymbol{w}}\right)\right)\mathbb{D}^{-1}_{\boldsymbol{w}}, \tag{6}$$

which is simple to implement and fast to compute. We illustrate this (below and in S.M.V) and provide the code to compute $\mathbb{K}$ from eq.6, for both Matlab and R, in S.M.VIII.

The Kinship formula as an infinite sum

The Kinship Formula can also be equivalently expressed as (see S.M.VI)

$$\mathbb{K} = \sum_{t=0}^{\infty}\mathbb{M}^t\left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}}+\mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}}+\mathbb{Z}\right)\mathbb{P}^t, \tag{7}$$

Computationally, this formulation can only provide the exact number of kin when the sum is finite, i.e., when a maximum age ω is embedded in the projection matrix (such that $\mathbf{S}^{\omega} = \mathbf{0}$). In this case $\mathbb{K}$ simplifies to: $\mathbb{K} = \sum_{t=0}^{\omega(g_m-1)-1}\mathbb{M}^t\left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}}+\mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}}+\mathbb{Z}\right)\mathbb{P}^t$. While this equation is of lesser interest than its alternatives, it remains useful. For instance, we employ this this approach when computing the eventual number of kin (see below and S.M.VII).

Particular Reproductive Cases

In some particular cases, that are likely to constitute the majority of cases in practice, the number of same-litter-daughters $F^{(j)}_{i,k}$ is independent from the class (j) of the offspring known to have been produced. This simplifies computation, allowing the number of same-litter-daughters to be represented by a matrix $\mathbf{F}^*$. In its generation-structure formulation, $\mathbb{F}^* = \mathbf{J_1}\otimes\mathbf{F}^*$, it allows computing directly $\mathbb{Z} = \mathbb{F}^*\mathbb{J}\mathbb{P}_{\mathbb{F}}$ and $\mathbb{Z}^{\dagger} = \mathbb{F}^*\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T$, which can then be incorporated into the various versions of the Kinship Formula. For eq. 3, this is $\mathbb{K} = \mathbb{M}\mathbb{K}\mathbb{P}+\left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}}+\mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}}+\mathbb{F}^*\mathbb{J}\mathbb{P}_{\mathbb{F}}\right)$ and for eq.5:

$$\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}}) = (\lambda\mathbf{I} - (\mathbb{M}\otimes\mathbb{M}))^{-1}(\mathbb{S}\otimes\mathbb{F} + \mathbb{F}\otimes\mathbb{S} + \mathbb{F}\otimes\mathbb{F}^*)\,\boldsymbol{vec}(\mathbb{JD}_{\boldsymbol{w}}) \tag{8}$$

We now consider three such cases (see S.M.IV). These are

Bernoulli Reproduction Individuals produce at most 1 offspring per time-step, hence $\mathbf{F}^* = \mathbf{0}$.

Independent Poisson reproduction Here, the number and class of supplemental offspring is independent from the number and classes of offspring already produced and therefore $\mathbf{F}^* = \mathbf{F}$.

One offspring class (say, class 1) then, from eq.14, $\mathbf{F}^*$ is such that $F^*_{1,i} = \frac{Var(\tilde{F}_{1,i})}{F_{1,i}} + F_{1,i} - 1$

Illustration

In S.M.V, we illustrate the computation of the number of kin for a population of ground squirrels, *Spermophilus dauricus*, structured by 3 stages. From the life table in (43) and for $g_m = 3$, we obtain the following Kinship Matrix (via eqs.8 and 6):

$$\mathbb{K} = \left[\begin{array}{ccc|ccc|ccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.13 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.38 & 0.24 & 0.08 & 0.11 & 0.05 & 0.02 \\
\hline
0 & 0.23 & 0.66 & 0.60 & 0.38 & 0.14 & 0.18 & 0.09 & 0.03 \\
0 & 0 & 0.24 & 0.22 & 0.34 & 0.16 & 0.18 & 0.10 & 0.03 \\
0 & 0 & 0.08 & 0.08 & 0.16 & 0.22 & 0.20 & 0.19 & 0.11 \\
\hline
0 & 0 & 0.19 & 0.18 & 0.32 & 0.35 & 0.33 & 0.30 & 0.18 \\
0 & 0 & 0.05 & 0.05 & 0.10 & 0.19 & 0.17 & 0.19 & 0.14 \\
0 & 0 & 0.02 & 0.02 & 0.03 & 0.11 & 0.10 & 0.14 & 0.16
\end{array}\right]
\begin{array}{l}
\left.\vphantom{\begin{array}{c}0\\0\\0\end{array}}\right\}\text{ancestors of } \textit{ego}\text{, kin is common ancestor (c.a.)} \\
\left.\vphantom{\begin{array}{c}0\\0\\0\end{array}}\right\}\text{kin is daughter (dtr) of c.a.} \\
\left.\vphantom{\begin{array}{c}0\\0\\0\end{array}}\right\}\text{kin is g-daughter (gdtr) of c.a.}
\end{array}
\tag{9}$$

$$\underbrace{\qquad\qquad}_{\text{descendants of } \textit{ego}\text{, } \textit{ego} \text{ is c.a.}}\quad\underbrace{\qquad\qquad}_{\text{desc. of mother, } \textit{ego} \text{ is dtr of c.a.}}\quad\underbrace{\qquad\qquad}_{\text{desc. of g-mother, } \textit{ego} \text{ is gdtr of c.a.}}$$

In the $(1,2)$ block of eq.9, highlighted in red, we find the expected number of mothers, i.e., the probability for *ego* to have its mother alive (see fig.1). The Kinship Matrix $\mathbb{K}$ indicates, for instance, that a yearling (an individual in class 1) has a 13% probability for its mother to be alive *and* in class 2. In the $(3,3)$ block, highlighted in blue, we observe the expected number of cousins: a class 2 individual, on average, has 0.19 cousins in class 2. The element in green, corresponds to the $((3,2),(1,3))$ entry of $\mathbb{K}$, representing the expected number of $(2,3)$-kin, or aunts, in class 3 for *ego* in class 1. Notably, the kin numbers in eq.9 are exact, contrasting with the approximation in (29).

Extensions

Unstructured Kinship

The Kinship Matrix $\mathbb{K}$ is structured by class and type of kin. It is also possible to consider the unstructured kinship of this structured population, meaning the number of kin regardless of their class for an *ego* chosen at random from the entire population. Let $\mathbb{K}^u(g_i, g_j)$ be the expected number of (g_i, g_j)-kin of such an individual. At the stable state, the probability for *ego* to be in class c_j is w_{c_j} and the number of its g_i kin is the sum over all c_i, therefore $\mathbb{K}^u(g_i, g_j) = \sum_{c_i, c_j} w_{c_j} \mathbb{K}((c_i, g_i), (c_j, g_j))$, or, in matrix form:

$$\mathbb{K}^u = \mathbb{J}_y \mathbb{K}_w \mathbb{J}_y^T, \tag{10}$$

with $\mathbb{J}_y = (\mathbf{I}_{g_m} \otimes \mathbf{1}_s^T)$ the $g_m \times s g_m$ matrix where the ith $g_m \times s$ block is zero but for the ith row filled with ones, $\mathbf{1}_s$ the vertical vector (of size s) of ones and ^{T} denoting matrix transposition.

The Unstructured Kinship Matrix $\mathbb{K}^u$ is symmetrical This is a key result from this manuscript:

Theorem 2. For *ego* drawn at random in the population at the stable state, its expected number of (g_i, g_j)-kin is equal to its expected number of (g_j, g_i)-kin: $\forall g_i, g_j, \mathbb{K}^u(g_i, g_j) = \mathbb{K}^u(g_j, g_i)$ or, in matrix form:

$$\mathbb{K}^u = \mathbb{K}^{u^T}.$$

In other words, the probability for an individual to have its mother alive is equal to the expected number of its daughters, and the expected number of nieces is equal to that of aunts. This result is intuitive: a (g_i, g_j) kin relationship for *ego* corresponds to a (g_j, g_i) kin relationship for the kin. In a population with n individuals, among the $n(n-1)$ (directional) relationship pairs, there will be as many (g_i, g_j) kin relationships as (g_j, g_i) ones: $n_{g_i, g_j} = n_{g_j, g_i}$. At the individual level, this implies that $\mathbb{K}^u(g_i, g_j) = \frac{n_{g_i, g_j}}{n(n-1)} = \frac{n_{g_j, g_i}}{n(n-1)} = \mathbb{K}^u(g_j, g_i)$.

proof The symmetry is absent from $\mathbb{K}$ itself, but present in $\mathbb{K}_w$. Indeed, as $\mathbb{D}_{\boldsymbol{w}}\mathbb{J} = \mathbb{J}\mathbb{D}_{\boldsymbol{w}}$, we have $\left(\mathbb{F}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T\right)^T = \mathbb{S}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T$. From eq.16, it is clear that $\mathbf{Z}^\dagger$ and therefore $\mathbb{Z}^\dagger$ are symmetrical. From eq.4, this implies that $\lambda\mathbb{K}_{\boldsymbol{w}} - \mathbb{M}\mathbb{K}_{\boldsymbol{w}}\mathbb{M}^T$ is symmetrical which leads to $\boldsymbol{vec}\left(\mathbb{K}_{\boldsymbol{w}} - \mathbb{K}_{\boldsymbol{w}}^T\right) = \left(\lambda\mathbf{I} - (\mathbb{M} \otimes \mathbb{M})\right)^{-1} \mathbf{0} = \mathbf{0}$ and therefore $\mathbb{K}_{\boldsymbol{w}} = \mathbb{K}_{\boldsymbol{w}}^T$. From eq.10, this further implies that $\mathbb{K}^{u^T} = \mathbb{J}_y \mathbb{K}_w^T \mathbb{J}_y^T = \mathbb{K}^u$, which completes the proof.

illustration For $g_m = 3$, the Unstructured Kinship Matrix for the ground squirrels population is (S.M.V):

$$\mathbb{K}^u = \begin{bmatrix} 0 & 0.32 & 0.07 \\ 0.32 & 0.80 & 0.41 \\ 0.07 & 0.41 & 0.58 \end{bmatrix} \tag{11}$$

From eq.11, we can infer that an individual taken at random in the population, is expected to have 0.80 sisters, 0.58 cousins, 0.41 nieces and aunts and 0.07 grand-daughters and grandmothers (i.e., the probability that its grandmother is alive is 0.07). We observe that $\mathbb{K}^u$ is indeed symmetrical. We can explore the number of kin at much larger g_m distances. In fig. 2, we show the expected (unstructured) number of kin for the population of ground squirrels (a declining population, $\lambda < 1$) for $g_m = 300$. Each curve corresponds to the number of (g_i, g_j) kin for a given g_j, that is, to the descendants of the g_j-ancestor (see fig.1). For $g_j = 1$ (in dark blue), it corresponds to the numbers of the various direct g_i-descendants of *ego*, which (from th.2) is also the number of its direct g_i-ancestors. Dots correspond to the elements already appearing in $\mathbb{K}^u$ for $g_m = 3$ of eq.11. After a "transient" period corresponding to close kin (low g_i and g_j), we observe that, for low or high g_i, *ego* has no (g_i, g_j) kin alive (they are are already dead or not born yet) and that $\mathbb{K}^u(g_i, g_j)$, for a given g_j, reaches a maximum at $g_i = g_j$, corresponding to *ego*'s g_i^{th} cousins (we point to an asymptotic analysis of kinship for unstructured populations in (29)).

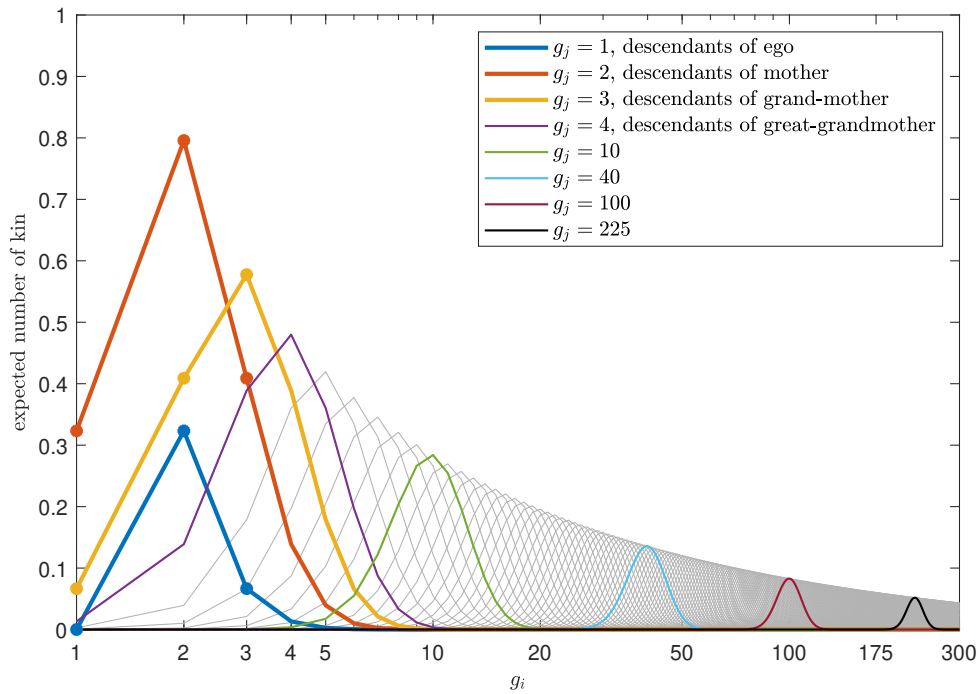

Figure 2: (unstructured) number of (g_i, g_j)-kin, $\mathbb{K}^u(g_i, g_j)$, for the ground squirrel population (illustration) for $1 \leq g_i, g_j \leq 300$ (note: the x axis, g_i, is on a *log* scale). Each curve represents the descendants of the g_j-ancestor of *ego* for a given g_j (see fig.1). Dots correspond to elements appearing in eq.11.

Unstructured kinship vs kinship of unstructured population The Kinship Matrix offers a novel avenue for application of the Trait Level Analysis developed in (44). This methodology assesses the sensitivity of a population model to a specific trait by comparing the model to its version without (or "folded on") that trait (see S.M.I). Within this multitrait theoretical framework, $\mathbb{K}^u$ represents the (*class* × *generation*)-matrix $\mathbb{K}$, *folded* over trait *class*. By contrasting $\mathbb{K}^u(\mathbf{M})$, the unstructured kinship of the structured model $\mathbf{M}$, with the kinship of the unstructured population $\mathbb{K}(\mathbf{M^u})$ derived from $\mathbf{M}$ *folded* over trait *class* (denoted $\mathbf{M^u}$), we gain insights into how the process(es) embedded in the *class* heterogeneity influence the kinship structure. We illustrate this approach using the ground squirrel population in

S.M.V, demonstrating that their survival senescence (together with the age-fertility pattern) leads to a reduction in the number of direct ascendants (mother, grand-mother) and descendants (daughters, grand-daughters) while simultaneously increasing the count of "indirect" kin, such as sisters, nieces, cousins and aunts.

Kinship and Relatedness

The Kinship Formula (eq.3) provides the number of kin in a one-sex population. Consequently, it is possible to relate the Kinship Matrix to the relatedness along the one-sex lineages of the population (45). This requires defining a "relatedness distance" d corresponding to the length of the path connecting two individuals (29). In a (g_i, g_j) kin relationship, the kin is $(g_i - 1)$ generations away from the most recent common ancestor and *ego* is $(g_j - 1)$ generations away from that ancestor. Therefore $d(g_i, g_j) = (g_i - 1) + (g_j - 1) = g_i + g_j - 2$. The coefficient of relatedness for a (g_i, g_j) kin relationship is given by $r(g_i, g_j) = \frac{1}{2^{d(g_i,g_j)}}$ (46; 47), that is:

$$r(g_i, g_j) = \frac{1}{2^{g_i+g_j-2}} \tag{12}$$

Each relatedness coefficient $(\frac{1}{2^d})$ corresponds to the various entries on the $(d+1)^{th}$ anti-diagonal of $\mathbb{K}^u$ (eq.10). For $d = 2$ (i.e., $g_i + g_j = 4$), it accounts for the grand-mother, the sisters and the grand-daughters and for $g_m = 3$ (as per eq11), it corresponds to the main anti-diagonal.

mean relatedness These considerations extend the individual-centered kinship formula to population measures of relatedness. One common measure is the mean relatedness, which typically focuses on male or female sub-populations and considers the mean relatedness coefficient across all dyads (a dyad is a non-directional pair) for a given relatedness depth g_m. First, consider $\bar{r}_i$, the *individual* mean relatedness (48). It can be obtained from $r(g_i, g_j)$ and $\mathbb{K}^u$ via $\bar{r}_i = \sum_{g_i,g_j \leq g_m} r(g_i, g_j)\mathbb{K}^u_{g_i,g_j}$, which, from eq.12, yields

$$\bar{r}_i = \boldsymbol{c}^T \mathbb{K}^u \boldsymbol{c},$$

with vector $\boldsymbol{c}$ of length g_m such that $c_{g_i} = \frac{1}{2^{g_i-1}}$. From this individual relatedness, we can compute the population or *dyad* mean-relatedness $\bar{r}_p(n)$ of a population of size n. There are $n(n-1)$ directional pairs of individual and therefore $\frac{n(n-1)}{2}$ dyads in such a population. The total relatedness is $n\bar{r}_i$, in which each dyad is counted twice. Hence, the population mean relatedness is

$$\bar{r}_p = \frac{\bar{r}_i}{n-1} = \frac{\boldsymbol{c}^T \mathbb{K}^u \boldsymbol{c}}{n-1}.$$

Eventual number of kin

The Kinship Formula provides the *current* number of kin of *ego*, where both *ego* and the kin are alive. In human demography and anthropology, where all kin matter including those not alive at the same time than *ego*, it is valuable to know the *eventual* number of kin: their total number encompassing those

who died before *ego*'s birth and those who will be born after *ego*'s death. In S.M.VII, we show that this can be achieved by adding a class "death" to the original model structure. Applying this method to an unstructured, non-generation-overlapping population where individuals produce R_0 offspring per generation and the expected number of sisters is Z, we obtain for $g_m = 4$, the following *eventual* number of kin:

$$\mathbb{K}_e = \begin{bmatrix} 0 & 1 & 1 & 1 \\ R_0 & Z & Z & Z \\ R_0^2 & ZR_0 & ZR_0 & ZR_0 \\ R_0^3 & ZR_0^2 & ZR_0^2 & ZR_0^2 \end{bmatrix},$$

which is the same result obtained by Goodman et al.(24, fig 2) and by Pullum (37, p.550).

Conclusion and Discussion

In this article, we present the Kinship Formula (eq.3), which calculates the expected number of kin, for any kin relationship and for any generation-overlapping structured population, in a one-sex constant environment framework. This formula utilizes the genealogical Markov chain (31), a model projecting populations backward in time, to produce the Kinship Matrix (the expected number of all kin of *ego* structured by kin type and the classes of the kin and *ego*), thereby extending results previously restricted to age-structured populations (24; 26) to any structured population. For example, the Kinship Matrix can analyze relatedness within spatially structured populations at both the patch and meta-population levels. Thanks to its generation structure, the Kinship Formula is closed-form and provides the exact number of all kin, up to an arbitrary relatedness distance, simultaneously. Both the genealogical Markov chain and the generation-structure are likely to have further importance for Kinship Demography in particular and evolutionary demography in general. Generation-structured multi-type branching processes will be central to studying the variance and distribution of the number of kin, extending the expectation analyzed here (manuscript in prep.). The genealogical Markov chain should prove important for extending results from age-structured models to any structured population, such as the Euler-Lotka equation, a crucial tools for evolutionary studies (49; 29).

We expect the simplicity of the Kinship Formula to appeal to population ecologists who wish to infer relatedness frequencies in their study population and compare them to available pedigrees. Demographers and anthropologists will benefit from its ability to compute the exact number of kin without resorting to simulations. For evolutionary and behavioral ecologists, the formula facilitates comparisons of relatedness structures across various social species, enabling analysis from an inclusive fitness perspective (50). Its simplicity allows for large-scale theoretical investigations of relatedness, as demonstrated in fig.2, which will be crucial for geneticists studying the distribution of identical-by-descent alleles and breeding values in a population (51; 52) In essence the Kinship Formula allows to extend to generation-overlapping

structured populations the Kingman's coalescent (20). Additionally, the Kinship Formula is valuable for micro- and family-economists, as it can quantify the effects of inheritance systems on the distribution of economic capital across extended families. Epidemiologists will find it useful for deriving information on the likelihood of an infected person "descending" from a "cluster" within a spatially-structured compartmental model. The Kinship Formula's versatility arises from its applicability to any structured population, encompassing the two main types of individual heterogeneity: dynamic (classes are "stages") and fixed (classes are "types") heterogeneity (53; 54). It applies to (age- or) stage-structured models (55; 56), to genetic models where individuals are categorised by genotype or breeding value (51; 52), and to models combining both dynamic *and* fixed heterogeneity (57). As it allows to compare the unstructured relatedness to the relatedness of the unstructured population, the Kinship Formula is crucial for analyzing how different processes embedded within different traits affect the kinship network (44). For life historians, this is useful for understanding the consequences of trade-offs and the presence of various strategies on a population's relatedness landscape.

The kinship formula provides the expected number of kin at the demographic stable state, for a one-sex population. Its simplicity makes it a foundational tool for developing extensions to more realistic frameworks, in a process that would mirror the historical progress in projection models: understanding the deterministic long-term growth rate as the maximum eigenvalue of a projection model (58), allowed to incorporate varying environments (59), individual stochasticity (60), two sexes (61) and compute its sensitivity (62). Ongoing and future works aim to follow the same path for Kinship Demography. The theoretical studies enabled by the Kinship Formula, such as investigating long-distance kinship structure (a study crucial for genetics, initiated in 29), hold significant potential across various fields interested in kinship structure and the behavior of collaterals in dynamical processes.

Finally, while these future extensions will enhance our understanding of how demography affects kinship composition, they represent only one aspect of Kinship Demography. This field aims to establish a comprehensive theory linking population dynamics and relatedness by examining both the effects of demography on kinship and the feedback of kinship on demography (2). In population ecology, the presence of kin significantly impacts individual fitness in many animal species, particularly in social species like whales and great apes (63; 64). In human societies, the effects of kin presence are notable, such as the extended care required for young (called altriciality, see 65) and the costs associated with caring for frail older relatives (66). Incorporating these effects into population demography has, for instance, reassessed the selection pressures on reproduction and survival in the context of menopause (67) and susceptibility to late-onset diseases (68). The overarching goal of Kinship Demography is to generalize these ad-hoc studies to understand the feedback loop between demography and kinship, model populations via both kinship-dependent demography and demography-dependent kinship, and formulate a general theory of their co-evolution. These future investigations will rely heavily on the Kinship Formula

and emulate the work of Chu on the co-evolution of inter-generational transfers (69), and the general approach used by May (70) and others when studying density-dependent population (where density, like kinship, is both an input and an output of the population model).

Acknowledgments and funding sources
This study was supported by the Chair "Modélisation Mathématique et Biodiversité of Veolia – Ecole Polytechnique – MNHN–Fondation X", by the Agence Nationale de la Recherche (grant ANR-18-CE02-0011, MathKinD), by the Research Council of Norway (Centre of Excellence grant SFF-III, project 223257) and by the Natural Environment Research Council (grant no. NE/W006731/1). We thank Samuel Pavard and Mike Fowler for helpful comments and suggestions.

Appendices

1 Genealogical Markov Chains

From λ and $\boldsymbol{w}$, the maximum eigenvalue of $\mathbf{M}$ and its associated right-eigenvector, the genealogical Markov chain of $\mathbf{M}$ (31; 32; 71; 34) allows to project the population backwards in time. We denote its transition matrix $\mathbf{P}$, where $P_{ij} = \frac{M_{ji}\, w_i}{\lambda\, w_j}$ is the probability that the ancestor (via survival or reproduction) of an individual in class j was in class i at the previous time-step. Decomposed according to survival and fertility processes, this is $\mathbf{P} = \mathbf{P_S} + \mathbf{P_F}$, where

$$P_{\mathbf{S}}(i,j) = \frac{S_{ji}\, w_i}{\lambda\, w_j} \text{ and } P_{\mathbf{F}}(i,j) = \frac{F_{ji}\, w_i}{\lambda\, w_j}. \tag{13}$$

Note that, in order for backwards projection to obey the same rule of left-multiplication as projection matrices, we write $\mathbf{P}$ as a left stochastic matrix; $\mathbf{P}$ is therefore the transpose of the matrix of the same name in, e.g., (31) or (29).

Genealogical Markov chain and distribution of maternal age The genealogical Markov chain generalises to any structure, the distribution of age of mothers (at birth-giving) used, for age-structured populations, in (24; 26). For a Leslie matrix (55; 72) – the projection matrix of a population structured by age – we get a $\mathbf{P_S}$ which is zero but for the sub-diagonal made of 1s (individuals older than the first age class necessary stem from their younger selves in the previous age class) and a $\mathbf{P_F}$, which is zero but for its first column which corresponds to the distribution of age of mothers.

2 Same-litter daughters and sisters

The number of same-litter-daughters $F^{(j)}_{i,k}$ is the expected number of supplemental offspring produced by a mother knowing she has produced one, structured by the class i of the supplemental offspring, the class k of the mother and the class j of the offspring known to have been produced:

$$F^{(j)}_{i,k} = \frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k})}{F_{j,k}} + F_{i,k} - \delta(i,j), \tag{14}$$

where $\delta(i,j) = 1$ if $i = j$ and 0 otherwise. This formula (proof in S.M.IV) generalises to a structured generation-overlapping population, the formula for unstructured non-generation overlapping populations of (24) to be found in paragraph 3 of the addendum.

Same-litter sister matrix From $\mathbf{P_F}$ and eq.14, we can generate $\mathbf{Z}$, the same-litter sister matrix, which element $Z_{i,j}$ is the expected number of sisters in class j of *ego* in class i, by browsing through all classes k in which *ego*'s mother may be as she gives birth to this litter: $Z_{i,j} = \sum_k F^{(j)}_{i,k} P_F(k,j)$, that is:

$$Z_{i,j} = \sum_k \left(\frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k})}{F_{j,k}} + F_{i,k} - \delta(i,j) \right) P_F(k,j) \tag{15}$$

3 Substituting the genealogical Markov chain into the kinship formula

Let $\mathbf{D}_{\boldsymbol{w}}$ be the matrix of size $s \times s$, with zeros everywhere but for its diagonal made up of the elements of $\boldsymbol{w}$, the vector of relative abundances. For any $s \times s$ matrix $\mathbf{A}$, $(\mathbf{A}\mathbf{D}_{\boldsymbol{w}})(i,j) = A_{i,j} w_j$. Therefore, from eq.13, $(\mathbf{P_S}\mathbf{D}_{\boldsymbol{w}})(i,j) = \frac{S_{ji}\, w_i}{\lambda} = (\frac{1}{\lambda}\mathbf{S}\mathbf{D}_{\boldsymbol{w}})(j,i)$. In other words, $\mathbf{P_S}\mathbf{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\left(\mathbf{S}\mathbf{D}_{\boldsymbol{w}}\right)^T = \frac{1}{\lambda}\mathbf{D}_{\boldsymbol{w}}\mathbf{S}^T$, where ^{T} denotes transposition. Similarly, we have $\mathbf{P_F}\mathbf{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\mathbf{D}_{\boldsymbol{w}}\mathbf{F}^T$. Now, let $\mathbb{D}_{\boldsymbol{w}} = \mathbf{I}_{g_m} \otimes \mathbf{D}_{\boldsymbol{w}}$. Then, from eqs1,2, classic Kronecker product properties (73) and the above, we get $\mathbb{P}_{\mathbb{F}}\mathbb{D}_{\boldsymbol{w}} = \left(\mathbf{J}_1^T \otimes \mathbf{P_S}\right)\left(\mathbf{I}_{g_m} \otimes \mathbf{D}_{\boldsymbol{w}}\right) = \frac{1}{\lambda}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T$. Similarly we have $\mathbb{P}_{\mathbb{S}}\mathbb{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T$ and $\mathbb{P}\mathbb{D}_{\boldsymbol{w}} = \mathbb{P}_{\mathbb{S}}\mathbb{D}_{\boldsymbol{w}} + \mathbb{P}_{\mathbb{F}}\mathbb{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\mathbb{D}_{\boldsymbol{w}}(\mathbb{S}^T + \mathbb{F}^T) = \frac{1}{\lambda}\mathbb{D}_{\boldsymbol{w}}\mathbb{M}^T$. Let us now multiply both sides of eq.3 by $\mathbb{D}_{\boldsymbol{w}}$ (on the right). From the above, we get $\mathbb{K}\mathbb{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\mathbb{M}\mathbb{K}\mathbb{D}_{\boldsymbol{w}}\mathbb{M}^T + \left(\frac{1}{\lambda}\mathbb{F}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T + \frac{1}{\lambda}\mathbb{S}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T + \mathbb{Z}\mathbb{D}_{\boldsymbol{w}}\right)$. Note that, in the particular reproductive cases, where $\mathbb{Z} = \mathbb{F}^\star\mathbb{J}\mathbb{P}_{\mathbb{F}}$, we have $\mathbb{Z}\mathbb{D}_{\boldsymbol{w}} = \frac{1}{\lambda}\mathbb{F}^\star\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T$. Call $\mathbb{K}_{\boldsymbol{w}} = \mathbb{K}\mathbb{D}_{\boldsymbol{w}}$, $\mathbf{Z}^\dagger = \lambda\mathbf{Z}\mathbf{D}_{\boldsymbol{w}}$ and $\mathbb{Z}^\dagger = \mathbf{J_2} \otimes \mathbf{Z}^\dagger = \lambda(\mathbf{J_2} \otimes \mathbf{Z})(\mathbf{I}_{g_m} \otimes \mathbf{D}_{\boldsymbol{w}}) = \lambda\mathbb{Z}\mathbb{D}_{\boldsymbol{w}}$. From the above, we get that $Z^\dagger_{i,j} = \sum_k \left(\frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k})}{F_{j,k}} + F_{i,k} - \delta(i,j)\right) F(j,k) w_k$ therefore $\mathbf{Z}^\dagger$ can be computed directly from

$$Z^\dagger_{i,j} = \sum_k w_k \left(Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k}) + F_{i,k}F_{j,k} - \delta(i,j)F_{j,k}\right), \tag{16}$$

and eq.3 is equivalent to

$$\lambda\mathbb{K}_{\boldsymbol{w}} = \mathbb{M}\mathbb{K}_{\boldsymbol{w}}\mathbb{M}^T + (\mathbb{F}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{S}^T + \mathbb{S}\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\mathbb{F}^T + \mathbb{Z}^\dagger). \tag{4}$$

4 Vectorization of the kinship formula

An interesting property of the $\boldsymbol{vec}$ operator and the tensor product is that, for a given set of three matrices with appropriate sizes $\mathbf{A}$,$\mathbf{B}$ and $\mathbf{C}$, we have (42; 73):

$$\boldsymbol{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\boldsymbol{vec}(\mathbf{B}). \tag{17}$$

Applied to eq.3, this yields: $\boldsymbol{vec}(\mathbb{K}) = \left(\mathbb{P}^T \otimes \mathbb{M}\right)\boldsymbol{vec}(\mathbb{K}) + \boldsymbol{vec}(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z})$, that is:

$$\left(\mathbf{I} - \left(\mathbb{P}^T \otimes \mathbb{M}\right)\right)\boldsymbol{vec}(\mathbb{K}) = \boldsymbol{vec}(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z}), \tag{18}$$

with $\mathbf{I}$ the identity matrix of size $s^2g_m^2 \times s^2g_m^2$. Vector $\boldsymbol{vec}(\mathbb{K})$ contains the exact same information than matrix $\mathbb{K}$, but ordered differently: the $c_i + s(g_i - 1) + (sg_m)(c_j - 1) + s^2g_m(g_j - 1)$ element of $\boldsymbol{vec}(\mathbb{K})$ of length $s^2g_m^2$, corresponds to the $((c_i, g_i), (c_j, g_j))$ entry of $\mathbb{K}$ which can be read in its $(c_i + s(g_i - 1))^{th}$ row and its $(c_j + s(g_j - 1))^{th}$ column and is the expected number of (g_i, g_j)-kin in class c_j for *ego* in class c_j. Applying eq.17 to eq.4, we get: $\lambda\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}}) = (\mathbb{M} \otimes \mathbb{M})\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}}) + \left((\mathbb{S} \otimes \mathbb{F} + \mathbb{F} \otimes \mathbb{S})\boldsymbol{vec}(\mathbb{J}\mathbb{D}_{\boldsymbol{w}}) + \boldsymbol{vec}\left(\mathbb{Z}^\dagger\right)\right)$ which can be written as

$$(\lambda\mathbf{I} - (\mathbb{M} \otimes \mathbb{M}))\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}}) = (\mathbb{S} \otimes \mathbb{F} + \mathbb{F} \otimes \mathbb{S})\boldsymbol{vec}(\mathbb{J}\mathbb{D}_{\boldsymbol{w}}) + \boldsymbol{vec}\left(\mathbb{Z}^\dagger\right). \tag{19}$$

5 Existence, unicity and convergence of the kinship formulas

Equation 18 (which is equivalent to eqs.3,4 and 19) admits a solution, and that solution is unique, if and only if $\left(\mathbf{I} - \left(\mathbb{P}^T \otimes \mathbb{M}\right)\right)$ is invertible. For any (square) real matrix $\mathbf{A}$, its spectral radius (the maximum of the absolute values of its eigenvalues) is such that $\rho_{\mathbf{A}} < 1$ if and only if $\mathbf{A}^t$ converges exponentially towards $\mathbf{0}$ as $t \to \infty$, its Neumann series (that is, $\sum_{t=0}^{\infty}\mathbf{A}^t$) converges and $(\mathbf{I} - \mathbf{A})$ is invertible (see 74):

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{t=0}^{\infty}\mathbf{A}^t \tag{20}$$

If $\mathbf{S}$ is a "regular" survival matrix, that is, that it corresponds to the transitions between transient states of an absorbing Markov chain, which absorbing state is "death", we have $\lim_{t\to+\infty}\mathbf{S}^t = \mathbf{0}$ and $\rho_{\mathbf{S}} < 1$. Its Neumann series is called the fundamental matrix and is used to compute, for instance, the next-generation matrix (see, e.g., 35). If $\mathbf{S}$ is a "regular" survival matrix, then $\mathbb{M}$ also corresponds to the transitions between transient states of an absorbing Markov chain, the absorbing states here, being "death" and generation$>g_m$. This implies that $\rho_{\mathbb{M}} < 1$. Matrix $\mathbf{P}$ is the transition matrix of a Markov chain, therefore $\rho_{\mathbf{P}} < 1$. Its generation structured equivalent models the same process, but for individuals with generation number 1, which ancestors via reproduction are not counted, therefore $\rho_{\mathbb{P}} < \rho_{\mathbf{P}} = 1$. One spectral property of the kronecker product, is that, for two square matrices of same size $\mathbf{A}$ and $\mathbf{B}$, we have $\rho_{\mathbf{AB}} = \rho_{\mathbf{A}}\rho_{\mathbf{B}}$ (75). Therefore $\rho_{(\mathbb{P}^T \otimes \mathbb{M})} = \rho_{\mathbb{P}}\rho_{\mathbb{M}} < 1$, which completes the proof and we have demonstrated that, for a

"proper" survival matrix Eqs4,3 and Eq19 have a (unique) solution, which can be written as:

$$\boldsymbol{vec}\left(\mathbb{K}\right) = \left(\mathbf{I} - \left(\mathbb{P}^T \otimes \mathbb{M}\right)\right)^{-1} \boldsymbol{vec}\left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z}\right) \tag{21}$$

or

$$\boldsymbol{vec}\left(\mathbb{K}_{\boldsymbol{w}}\right) = \left(\lambda\mathbf{I} - \left(\mathbb{M} \otimes \mathbb{M}\right)\right)^{-1} \left(\left(\mathbb{S} \otimes \mathbb{F} + \mathbb{F} \otimes \mathbb{S}\right) \boldsymbol{vec}\left(\mathbb{J}\mathbb{D}_{\boldsymbol{w}}\right) + \boldsymbol{vec}\left(\mathbb{Z}^{\dagger}\right)\right) \tag{5}$$

By applying eq.20 and eq.17, to eq.21, we can write this equivalently as an infinite sum which converges:

$$\mathbb{K} = \sum_{t=0}^{\infty} \mathbb{M}^t \left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z}\right) \mathbb{P}^t. \tag{7}$$

References

[1] Lotka A (1931) Orphanhood in Relation to Demographic Factors. *Metron* 2:37–109.

[2] Pavard S, Coste CF (2020) Goodman, Keyfitz & Pullum (1974) and the population frequencies of kinship relationships. *Theoretical Population Biology* 133:15–16.

[3] Clutton-Brock TH (1989) Female transfer and inbreeding avoidance in social mammals. *Nature* 337(6202):70–72.

[4] Rousset F (2002) Inbreeding and relatedness coefficients: What do they measure? *Heredity* 88(5):371–380.

[5] Gardner A, West SA, Wild G (2011) The genetical theory of kin selection. *Journal of Evolutionary Biology* 24(5):1020–1043.

[6] Hamilton W (1964) The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7(1):1–16.

[7] Queller DC (1994) Genetic relatedness in viscous populations. *Evolutionary Ecology* 8(1):70–73.

[8] Cox D, Fafchamps M (2007) Chapter 58 Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions in *Handbook of Development Economics*. Vol. 4, pp. 3711–3784.

[9] Laitner J (1997) Chapter 5 Intergenerational and interhousehold economic links in *Handbook of Population and Family Economics*. Vol. 1, pp. 189–238.

[10] Koster J, et al. (2019) Kinship ties across the lifespan in human communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374(1780):20180069.

[11] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* 438(7066):355–359.

[12] Althouse BM, et al. (2020) Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLOS Biology* 18(11):e3000897.

[13] Wang L, et al. (2020) Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. *Nature Communications* 11(1):1–6.

[14] Streeck H, et al. (2020) Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. *Nature Communications* 11(1):1–12.

[15] Reid JM, Nietlisbach P, Wolak ME, Keller LF, Arcese P (2019) Individuals' expected genetic contributions to future generations, reproductive value, and short-term metrics of fitness in free-living song sparrows (Melospiza melodia). *Evolution Letters* 3(3):271–285.

[16] Nietlisbach P, et al. (2017) Pedigree-based inbreeding coefficient explains more variation in fitness than heterozygosity at 160 microsatellites in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences* 284(1850).

[17] Heyer E, Sibert A, Austerlitz F (2005) Cultural transmission of fitness: Genes take the fast lane. *Trends in Genetics* 21(4):234–239.

[18] Barrai I, Cavalli-Sforza L, Moroni A (1962) Frequencies of pedigrees of consanguineous marriages and mating structure of the population. *Annals of Human Genetics*.

[19] 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.

[20] Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability* 19(A):27–43.

[21] Kingman JF (1982) The coalescent. *Stochastic Processes and their Applications* 13(3):235–248.

[22] Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.

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

[24] Goodman LA, Keyfitz N, Pullum TW (1974) Family formation and the frequency of various kinship relationships. *Theoretical Population Biology* 5(1):1–27.

[25] Rosenberg NA (2020) Fifty years of Theoretical Population Biology. *Theoretical Population Biology* 12(xxxx):1–12.

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

[27] Hammel EA (2005) Demographic dynamics and kinship in anthropological populations. *Proceedings of the National Academy of Sciences of the United States of America* 102(6):2248–53.

[28] Wachter KW (1997) Kinship resources for the elderly. *Philosophical Transactions of the Royal Society B: Biological Sciences* 352(1363):1811–1817.

[29] Coste CF, et al. (2021) The kinship matrix: inferring the kinship structure of a population from its demography. *Ecology Letters* (June):1–13.

[30] Joffe A, Waugh WAO (1982) Exact distributions of kin numbers in a Galton-Watson process. *Journal of Applied Probability* 19(4):767–775.

[31] Demetrius L (1974) Demographic Parameters and Natural Selection. *Proceedings of the National Academy of Sciences* 71(12):4645–4647.

[32] Demetrius L (1975) Natural selection and age structured populations. *Genetics* 79(3):535–544.

[33] Tuljapurkar SD (1982) Why use population entropy? It determines the rate of convergence. *Journal of Mathematical Biology* 13(3):325–337.

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

[35] Caswell H (2001) *Matrix Population Models*. (Sinauer Associates Inc., Sunderland, MA.), p. 722.

[36] Atkins JR (1974) On the fundamental consanguineal numbers and their structural basis. *American Ethnologist* 1(1):1–31.

[37] Pullum TW (1982) The eventual frequencies of kin in a stable population. *Demography* 19(4):549–565.

[38] Pullum TW, Wolf DA (1991) Correlations between frequencies of kin. *Demography* 28(3):391–409.

[39] Simoncini V (2016) Computational methods for linear matrix equations. *SIAM Review* 58(3):377–441.

[40] Zoran Gajić MTJQ (1990) Introduction. *Lyapunov Matrix Equation in System Stability and Control* 35(8):1–12.

[41] Lancaster P, Lerer L, Tismenetsky M (1984) Factored forms for solutions of AX - XB = C and X - AXB = C in companion matrices. *Linear Algebra and Its Applications* 62(C):19–49.

[42] Henderson HV, Searle SR (1981) The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra* 9(4):271–288.

[43] Luo J, Fox BJ (1990) Life-table comparisons between two ground squirrels. *Journal of Mammalogy* 71(3):364–370.

[44] Coste CF, Austerlitz F, Pavard S (2017) Trait level analysis of multitrait population projection matrices. *Theoretical Population Biology* 116:47–58.

[45] Armitage KB (1991) Social and population dynamics of yellow-bellied marmots: Results from long-term research. *Annual Review of Ecology and Systematics* 22(1):379–407.

[46] Wright S (1921) SYSTEMS OF MATING. I. THE BIOMETRIC RELATIONS BETWEEN PARENT AND OFFSPRING. *Genetics* 6(2):111–123.

[47] Wright S (1922) Coefficients of Inbreeding and Relationship. *The American Naturalist* 56(645):330–338.

[48] Dyble M, Migliano AB, Page AE, Smith D (2021) Relatedness within and between Agta residential groups. *Evolutionary Human Sciences* 3:1–11.

[49] Charlesworth B (1980) *Evolution in age structured populations*. (Cambridge University Press), p. 300.

[50] Hamilton W (1964) The genetical Evolution of Social Behavior II.

[51] Coulson T, Tuljapurkar S (2008) The Dynamics of a Quantitative Trait in an Age-Structured Population Living in a Variable Environment. *The American Naturalist* 172(5):599–612.

[52] Coulson T, Tuljapurkar S, Childs DZ (2010) Using evolutionary demography to link life history theory, quantitative genetics and population ecology. *Journal of Animal Ecology* 79(6):1226–1240.

[53] Tuljapurkar S, Steiner UK, Orzack SH (2009) Dynamic heterogeneity in life histories. *Ecology Letters* 12(1):93–106.

[54] Hamel S, et al. (2018) General conclusion to the special issue Moving forward on individual heterogeneity. *Oikos* 127(5):750–756.

[55] Leslie PH (1945) On the use of matrices in certain population mathematics. *Biometrika* 33:183–212.

[56] Lefkovitch LP (1965) The Study of Population Growth in Organisms Grouped by Stages. *Biometrics* 21(1):1–18.

[57] Plard F, Bonenfant C, Delorme D, Gaillard JM (2012) Modeling reproductive trajectories of roe deer females: Fixed or dynamic heterogeneity? *Theoretical Population Biology* 82(4):317–328.

[58] Keyfitz N (1965) The intrinsic rate of natural increase and the dominant root of the projection matrix. *Population Studies* 18(3):293–308.

[59] Tuljapurkar S, Orzack SH (1980) Population dynamics in variable environments I. Long-run growth rates and extinction. *Theoretical Population Biology* 18(3):314–342.

[60] Engen S, Bakke O, Islam A (1998) Demographic and Environmental Stochasticity-Concepts and Definitions. *Biometrics* 54(3):840–846.

[61] Pollak RA (1986) A reformulation of the two-sex problem. *Demography* 23(2):247–259.

[62] Goodman LA (1971) On the sensitivity of the intrinsic growth rate to changes in the age-specific birth and death rates. *Theoretical Population Biology* 2(3):339–354.

[63] Johnstone Ra, Cant Ma (2010) The evolution of menopause in cetaceans and humans: the role of demography. *Proceedings. Biological sciences / The Royal Society* 277(1701):3765–3771.

[64] Wright BM, Stredulinsky EH, Ellis GM, Ford JK (2016) Kin-directed food sharing promotes lifetime natal philopatry of both sexes in a population of fish-eating killer whales, Orcinus orca. *Animal Behaviour* 115:81–95.

[65] Pavard S, Gagnon A, Desjardins B, Heyer E (2005) Mother's death and child survival: The case of early Quebec. *Journal of Biosocial Science* 37(2):209–227.

[66] Alburez-Gutierrez D, Mason C, Zagheni E (2021) The "Sandwich Generation" Revisited: Global Demographic Drivers of Care Time Demands. *Population and Development Review* 47(4):997–1023.

[67] Pavard S, Branger F (2012) Effect of maternal and grandmaternal care on population dynamics and human life-history evolution: A matrix projection model. *Theoretical Population Biology* 82(4):364–376.

[68] Pavard S, Coste CFD (2021) Evolutionary demographic models reveal the strength of purifying selection on susceptibility alleles to late-onset diseases. *Nature Ecology & Evolution* 5(3):392–400.

[69] Cyrus Chu C, Lee RD, Chu CY, Lee RD (2006) The co-evolution of intergenerational transfers and longevity: An optimal life history approach. *Theoretical Population Biology* 69(2):193–201.

[70] May RM (1976) Simple mathematical models with very complicated dynamics. *Nature* 261(5560):459–467.

[71] Tuljapurkar SD (1982) Population Dynamics in Variable Environments IV Weak ergodicity in the Lotka Equation. *Journal of mathematical biology* 14:221–230.

[72] Lewis EG (1942) On the generation and growth of a population. *Sankyha* 6, part 1(1942):93–96.

[73] Loan CF (2000) The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics* 123(1-2):85–100.

[74] Varga RS (2000) *Matrix Iterative Analysis*, Springer Series in Computational Mathematics ed. Prentice-Hall, Englewood Cliffs NJ. (Springer Berlin Heidelberg, Berlin, Heidelberg) Vol. 27.

[75] Neudecker H (1969) A Note on Kronecker Matrix Products and Matrix Equation Systems. *SIAM Journal on Applied Mathematics* 17(3):603–606.

Supplementary Material

I Concept of trait

In this manuscript, we call projective traits, or simply *traits*, the categorisation of the *states* of individuals in a structured population. In population ecology, for instance, such traits can be *age*, *stage*,*body mass*, *location*,*genotype* etc. In such a population, the population vector is made of several states. The traits categorise these states. In other words, they are the structure of the structure: *traits* are decomposed into *classes* and each individual *state* is a combination of these *classes* . For instance, in a stage × patch model (a 2-trait projection model therefore, also known as a meta-population model following Lebreton (1)), such as the one in (2), each individual *state* can be represented by a pair of *classes*, such as *state* (juvenile, A) which describes the juveniles (individuals in the class "juveniles" of trait stage) in patch A (individuals in class "A" of trait patch).

The concept of projective traits has been rarely used (in population ecology) until recently as most models embedded only one *trait*: such as *age* or *stage*. The concept of multi-trait projection models arises from the increased quantity and quality of data in all fields of population dynamics, which allows an individual in the population to be described by different traits influencing its fate. In (3), the authors develop a theory of such models where they define the "trait structure" of the r-trait model as $ts = \{t_1, t_2, .., t_r\}$, where t_i is the number of *classes* of the ith trait t_i. The related projection model is a block-matrix, that can be project over time, like any matrix model (4), the r-dimensional population vector, or tensor, providing it is vectorized (5). The result is a $q \times q$ matrix model with $q = t_1 t_2 ... t_r$, where the ith *state* corresponds to the combination of *classes* $(c_1, c_2, ..., c_r)$ such that $i = c_1 + \sum_{i=2}^{r}(c_r - 1)\prod_{j=1}^{i-1} t_j$. If there are 3 stages (newborn, juvenile and adults) and 2 patches (A and B) in the aforementioned stage × patch model, with $t_1 = 3$ and $t_2 = 2$, then state (newborn, B)=$(c_1 = 1, c_2 = 2)$ corresponds to entry $i = 1 + (2 - 1) * 3 = 4$ of the population vector of size $q = t_1 t_2 = 6$.

Such models allow to analyse the effect of one specific trait, and the process(es) it embeds, on the demography of a population by comparing the model with the trait to "equivalent" models from which the trait is absent. The procedure to obtain the "equivalent" model, called folding by Coste et al. (3), consists in computing, at the demographic stable state, the transition rates between states of the model without the trait under scrutiny by merging all states with the same classes for the other traits. This merger, yielding from projection model $\mathbf{M}$, its folded version $\mathbf{M}^{fold}_{trait_r}$, is simply performed by summing all incoming flows and summing all outgoing flows, weighted by the relative abundances of the states to be merged into one.Models $\mathbf{M}^{fold}_{trait_r}$ and $\mathbf{M}$ correspond to the same population: if trait r is the colour of the wings of a butterfly, then two entomologists, one colour-blind one not, will obtain, from the exact same data and analysis, respectively models $\mathbf{M}^{fold}_{trait_r}$ and $\mathbf{M}$ (6). These have therefore the same asymptotic growth rate and relative abundances but will differ in reproductive values and therefore sensitivities. The analysis of these differences, called Trait-Level Analysis in (3)will describe the effects of trait r on the dynamics (and evolution) of the studied population. We illustrate the interest of Trait-Level Analysis when applied to kinship, via the ground squirrel illustration of S.M.V in S.M.V.

II Discrete time and state space

In many models, such as those of population ecology, the space structure cannot be made continuous (e.g. for a population structured by life history morphs or instars, see 7). At the possible cost of discretisation, the structure consists of s classes, and the population can be described by a vector $\boldsymbol{n}(t)$ of size s. Structure discretisation is commonly used in population ecology, when accounting for continuous traits such as body mass, via, for instance, integral projection models (8; 9; 10). For many populations also, the time structure cannot be structured continuously. In population ecology again, for instance, many species, such as most birds or ungulates, reproduce at a specific period (or at specific periods) of the year. In such cases, the models are built from several "seasonal" projectors that are periodic matrices (in the sense of 11) combined (that is, multiplied, if in matrix form) to get an annual (or another time period) projector that does not depend on the year (4).

pre- vs post-breeding If, for instance, reproduction occurs at a specific time-period of the year (modeled by $\mathbf{F}'$), and the rest of the year individual "only" survive (modeled by $\mathbf{S}$), one can build an annual model by combining these, either by having reproduction first: $\mathbf{M}_{pre} = \mathbf{S}(\mathbf{I}_s + \mathbf{F}') = \mathbf{S} + \mathbf{S}\mathbf{F}' = \mathbf{S} + \mathbf{F}_{pre}$

or last $\mathbf{M}_{post} = (\mathbf{I}_s + \mathbf{F}')\mathbf{S} = \mathbf{S} + \mathbf{F}'\mathbf{S} = \mathbf{S} + \mathbf{F}_{post}$. $\mathbf{M}_{pre}$ is called a pre-breeding model and $\mathbf{M}_{post}$ a post-breeding model, they share the same survival ($\mathbf{S}$) component but differ in their reproduction components ($\mathbf{F} = \mathbf{F}_{pre}$ or $\mathbf{F}_{post}$). As they model the same populations these two models have the same asymptotic properties (same asymptotic growth rate) (see, for instance 12; 4).

III Independent and identically distributed processes of survival and reproduction

We consider that the process $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{S}}$ are i.i.d. In other words, we consider that (1) all individuals in class c_i follow the same processes of reproduction ($\tilde{\mathbf{F}}_{c_i}$) and survival ($\tilde{\mathbf{S}}_{c_i}$), (2) that the individual realisations of survival and fertility are independent from that of other individuals in that same ($c_i = c_j$) or other ($c_i \neq c_j$) classes and (3) that for a given individual in c_i, the realisation of survival and reproduction are independent: $Var(\tilde{\mathbf{F}}_{c_i}, \tilde{\mathbf{S}}_{c_i}) = \mathbf{0}$. The first assumption of identity and the second assumption of independence are classic in structured population models: the probabilistic fate of an individual is entirely determined by its state and therefore individuals in the same class are considered identical, their fate only differing because of stochasticity. The third assumption, also classic. does not prevent (as it may seem at first sight), to prevent the use of the Kinship Formula for population models that include, for instance, survival/fertility trade-offs. However, it can be easily shown that any model can be made to obey assumption (3), by adapting its structure (13). We illustrate it here for two cases that are common in population ecology.

Consider the case of "seasonal" matrices described in S.M.II. As is obvious from the formulas of $\mathbf{F} = \mathbf{F}_{pre} = \mathbf{S}\mathbf{F}'$ and $\mathbf{F} = \mathbf{F}_{post} = \mathbf{F}'\mathbf{S}$, the reproductive process at the annual level $\mathbf{F}$ depends, in general, on the survival process. However, while the post-breeding reproduction, at the annual level, via $\tilde{\mathbf{F}}_{post}$, of an individual in class c_j depends on the realisation of $\tilde{\mathbf{S}}_{c_j}$ for that same individual, it is not the case for the pre-breeding reproduction. Via $\tilde{\mathbf{F}}_{pre}$, the first seasonal stochastic matrix encountered is $\mathbf{I}_s + \tilde{\mathbf{F}}'$, where the first component is constant and the second one yields different individuals by definition. Therefore, in the case of seasonal models, using pre-breeding formulation, allows to respects the third assumption. On a side note, the extension of the kinship formula to account for varying environments (in prep), and in particular to compute the kinship of a population modelled via periodic matrices (in the sense of 11), such as $\mathbf{M}_{post} = (\mathbf{I}_s + \mathbf{F}')\mathbf{S}$ will only require independence of survival and fertility at the level of each periodic matrix. Here survival and fertility are clearly independent for both $\mathbf{S}$ and $(\mathbf{I}_s + \mathbf{F}')$ as each contain only one such process, and therefore the kinship of the population modelled by $\mathbf{M}_{post}$ will be possible and provide the kinship both after the fertility event and after the survival event.

The second case seemingly preventing assumption (3) is the case of trade-offs between *realisations* of survival and reproduction. In (14), such trade-offs are called "individual" trade-offs (which differ from "population" trade-offs which occur between *expectations* of survival and reproduction, across classes). Those can be modelled by having, for instance, for an individual in class c_j in a given year, survival $\tilde{\mathbf{S}}_{c_j}$ that year be a function of reproduction $\tilde{\mathbf{F}}_{c_j}$ that year, which violates the third assumption above. Here again, a careful choice of the structure allows to circumvent the problem, by adding trait *realised reproduction* to the structure. This is commonly done in bird population ecology, where , by adding classes "breeders" and "non-breeders" in the structure, one segregates individuals producing offspring or not during the current year. The trade-off is then incorporated by having "breeders" and "non-breeders" have different survival processes. In such a model, the stochasticity in the realisation of fertility – will I breed or not this year? – is therefore embedded in the survival of the previous time-step –will I breed or not next year ? that is, this current year, in case I survive, am I transitioning towards the "breeders" or the "non-breeders" class ? (in a structured model, survival involves both survival proper, and the transition of survivors to the various stages). In that case, there is independence, for a given individual in class c_j at a given time-step, between the realisation of its fertility (which can be constant, or depend on offspring survival or breeding success, but does not influence survival) and its survival, and assumption (3) is respected (see 15, for an example of such a model).

IV Same Litter Fertility

In order to count the number of kin, we count the descendants of the ancestors of *ego*. We can reach the descendants stemming from the older and younger sisters of the daughter of the ancestor by respectively. applying to the ancestor one backward survival projection (via $\mathbf{P_S}$ or $\mathbb{P}_{\mathbb{S}}$ to consider the ancestor *before* it produced its daughter in the genealogy of *ego*) followed by general forward projection (via $\mathbf{M}$ or $\mathbb{M}$) or one forward survival projection (via $\mathbf{S}$ or $\mathbb{S}$ to consider the ancestor *after* it produced its daughter in the

genealogy of *ego*) followed by general forward projection. In other words, the expected number of kin via older and younger sisters of the ancestor's daughter only depend on the deterministic projection model $\mathbf{M} = \mathbf{F} + \mathbf{S}$.

However when counting the descendants via sisters of the ancestor's daughter, it proves necessary, in the general case, to consider a further building-block, the same-litter-sister matrix $\mathbf{Z}$, which element $Z_{i,j}$ is the number of sisters of the same-litter in class j of an individual in class i that has just been born. $\mathbf{Z}$ can be inferred from $\tilde{\mathbf{F}}$ via the following steps:

$$Z_{i,j} = \sum_k F_{i,k}^{(j)} P_F(k,j), \tag{22}$$

where $F_{i,k}^{(j)}$ is the number of same-litter daughters, that is, the expected number of supplemental offspring produced by a mother knowing she has produced one, structured by the class i of the supplemental offspring, the class k of the mother and the class j of the offspring known to have been produced. Equation 22 simply equates the number of same-litter sisters of *ego* in class j to the number of same-litter daughters of the mother of *ego* weighted by the probability distribution of the class of the latter. The number of same-litter daughters depends, from its definition, on $\tilde{F}_{i,k}^{\dagger(j)}$ which has the distribution of $\tilde{F}(i,k)$ biased by $\tilde{F}(j,k)$ and on whether *ego* and the kin are in the same class:

$$F_{i,k}^{(j)} = (F_{i,k}^{\dagger(j)} - \delta(i,j)), \tag{23}$$

where $\delta(i,j) = 1$ if $i = j$ and 0 otherwise, and where

$$P(\tilde{\mathbf{F}}_k^{\dagger(j)} = \boldsymbol{n}) = \frac{n_j}{F_{j,k}} P(\tilde{\mathbf{F}}_k = \boldsymbol{n}), \tag{24}$$

where for a matrix $\mathbf{A}$, $\mathbf{A}_k$ is the vector corresponding to the kth column of $\mathbf{A}$, that is to $\{A_{i,k}\}_i$.

From this we can compute the expectation,

$$F_{i,k}^{\dagger(j)} = E(\tilde{F}_{i,k}^{\dagger(j)}) = \sum_{n_i} n_i P\big(\tilde{F}_{i,k}^{\dagger(j)} = n_i\big), \tag{25}$$

by noting that $P\big(\tilde{F}_{i,k}^{\dagger(j)} = n_i^*\big) = \sum_{\boldsymbol{n}=n_1,n_2,\ldots,n_i^*,\ldots,n_s} P\big(\tilde{\mathbf{F}}_k^{\dagger(j)} = \boldsymbol{n}\big)$. From eq. 24, we get therefore

$$P\big(\tilde{F}_{i,k}^{\dagger(j)} = n_i^*\big) = \sum_{\boldsymbol{n}=n_1,n_2,\ldots,n_i^*,\ldots,n_s} \frac{n_j}{F_{j,k}} P\big(\tilde{\mathbf{F}}_k = \boldsymbol{n}\big) = \sum_{n_j} \frac{n_j}{F_{j,k}} P\big(\tilde{F}_{i,k} = n_i^*, \tilde{F}_{j,k} = n_j\big) \tag{26}$$

From Eqs 26 and 25, we get therefore

$$F_{i,k}^{\dagger(j)} = \sum_{n_i,n_j} \frac{n_i n_j}{F_{j,k}} P(\tilde{F}_{i,k} = n_i, \tilde{F}_{j,k} = n_j) = \frac{1}{F_{j,k}} E\big(\tilde{F}_{i,k}\tilde{F}_{j,k}\big).$$

That is,

$$F_{i,k}^{\dagger(j)} = \frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k}) + F_{i,k}F_{j,k}}{F_{j,k}} \tag{27}$$

From Eqs 27 and 23, we then have

$$F_{i,k}^{(j)} = \frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k})}{F_{j,k}} + F_{i,k} - \delta(i,j), \tag{14}$$

Note that this formula differs from and corrects Eq(9) of (16) which was erroneous despite the calculations being "elementary". From Eqs 14 and 22, we get finally

$$Z_{i,j} = \sum_k \left(\frac{Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k})}{F_{j,k}} + F_{i,k} - \delta(i,j) \right) P_F(k,j) \tag{15}$$

Particular reproductive cases

In three particular (but very common in the literature) cases, the number of same-litter daughters does not depend on the class of the daughter ancestor of *ego*: $F_{i,k}^{(j)} = F_{i,k}^*$. The same-litter daughters is now

2-dimensional and can be represented as a matrix $\mathbf{F}^* = \{F^*_{i,j}\}_{i,j}$. eq. 22 becomes $Z_{i,j} = \sum_* F^{(j)}_{i,k} P_F(k,j)$, which can be written in matrix form as $\mathbf{Z} = \mathbf{F}^* \mathbf{P_F}$, which relates these particular common cases with the general case. These three cases are:

Bernoulli Reproduction In that case, individuals can produce at most 1 offspring per time step: there is by definition no same-litter daughters, then we simply have

$$\mathbf{F}^* = \mathbf{0}. \tag{28}$$

Independent Poissons reproduction Here, $\tilde{F}_{i,j}$ is a Poisson independent from $\tilde{F}_{k,j}$, therefore $\|\tilde{\mathbf{F}}_j\| = \sum_i \tilde{F}_{i,j}$ is a Poisson as the sum of s independent Poisson. A known result is that, in that framework, the number and state of supplemental offspring is independent from the number and states of offspring produced and therefore

$$\mathbf{F}^* = \mathbf{F} \tag{29}$$

This can be checked directly. From Eq14 we have, $F^{(i)}_{i,k} = \frac{Var(\tilde{F}_{i,k}) + F_{i,k}F_{i,k}}{F_{i,k}} - \delta(i,i) = F_{i,k}$, as $Var(\tilde{F}_{i,k}) = F_{i,k}$ and $\delta(i,i) = 1$. For $i \neq j$, the same equation gives $F^{(j)}_{i,k} = F_{i,k}$ (as $Cov(\tilde{F}_{i,k}, \tilde{F}_{j,k}) = 0$ and $\delta(i,j) = 0$) which is indeed independent from j. Therefore, in all cases, $F^*_{i,k} = F_{i,k}$.

One offspring class If there is only one class of offspring (say class 1), then $\mathbf{F}^*$ is non-zero only for $F^*_{1,i}$. $F^{(j)}_{i,k}$ is indeed only non-zero for $i = 1$ and $j = 1$, we have $F^{(1)}_{1,k} = \frac{Cov(\tilde{F}_{1,k}, \tilde{F}_{1,k})}{F_{1,k}} + F_{1,k} - \delta(1,1) = \frac{Var(\tilde{F}_{1,k})}{F_{1,k}} + F_{1,k} - 1$. Therefore $\mathbf{F}^*$ is such that

$$F^*_{1,j} = \frac{Var(\tilde{F}_{1,j})}{F_{1,j}} + F_{1,j} - 1, \tag{30}$$

and $\forall i \neq 1, F^*_{i,j} = 0$.

V Illustration

Here we illustrate our method on a population of ground squirrels *Spermophilus dauricus*, structured by stages. It is an extended Leslie model, i.e. an age-structured model with a class that regroups all individuals older than a certain age. All the computations can be reproduced from the main text equations or directly from the implementation of the equations of the main text (in R and in Matlab) to be found in S.M.VIII.

Here, classes 1 and 2 correspond to individuals that are about to turn 1 and 2 year(s) old, respectively, and class 3 corresponds to older individuals. We want to compute the expected number of all $(g_i - g_j)$-kin for all g_j and $g_i \leq g_m = 3$; there are $g_m^2 - 1 = 8$ such kin: the mother (1,2), grand-mother (1,3), the daughters (2,1), grand-daughters (3,1), the sisters (2,2), maternal aunts (2,3), nieces (3,2) and cousins (3,3) (see fig.1). From the life table for that population from (17), we get the following projection matrix: $\mathbf{M} = \mathbf{S} + \mathbf{F} = \begin{bmatrix} 0.23 & 0.51 & 0.84 \\ 0.57 & 0 & 0 \\ 0 & 0.54 & 0.46 \end{bmatrix}$.

Generation-structured matrices

From these and via eqs 1 and 2, we can construct $\mathbb{S} = \mathbf{I}_{g_m} \otimes \mathbf{S}$, $\mathbb{F} = \mathbf{J_1} \otimes \mathbf{F}$, $\mathbb{P}_\mathbb{S} = \mathbf{I} \otimes \mathbf{P_S}$ and $\mathbb{P}_\mathbb{F} = \mathbf{J_1}^T \otimes \mathbf{F}$ for $g_m = 3$. That is, $\mathbb{M} = \mathbb{S} + \mathbb{F} = \left[\begin{array}{c|c|c} \mathbf{S} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{F} & \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{F} & \mathbf{S} \end{array}\right]$ (see below for $\mathbb{P}_\mathbb{S}$ and $\mathbb{P}_\mathbb{F}$).

Kinship Matrix

Without further information, we assume that the fertility is Poisson and therefore $\mathbf{F}^* = \mathbf{F}$ and $\mathbb{F}^* = \mathbb{F}$. Via a simple eigen-analysis, we obtain the asymptotic growth rate $\lambda = 1.00$ (we round all numbers to 2 digits, note however that this asymptotic growth rate is < 1) and the stable class distribution $\mathbf{w} = (0.47, 0.27, 0.27)$. From the latter, we can construct $\mathbf{D}_{\boldsymbol{w}}$ the matrix with zeros everywhere but for its diagonal made up of the elements of $\boldsymbol{w}$, and from it $\mathbb{D}_{\boldsymbol{w}} = \mathbf{I}_{g_m} \otimes \mathbf{D}_{\boldsymbol{w}}$. We now have all the elements to

generate, from eq. 8, $\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}}) = (\lambda\mathbf{I} - (\mathbb{M}\otimes\mathbb{M}))^{-1}(\mathbb{S}\otimes\mathbb{F} + \mathbb{F}\otimes\mathbb{S} + \mathbb{F}\otimes\mathbb{F}^*)\,\boldsymbol{vec}(\mathbb{J}\mathbb{D}_{\boldsymbol{w}})$, where $\mathbb{J} = \mathbf{J_3}\otimes\mathbf{I}_s$, with $\mathbf{J_3}$, the zero $g_m \times g_m$ zero-matrix with 1 in the (1,1) position). From the solution to this equation we easily get (see main text for interpretation)

$$\mathbb{K} = \boldsymbol{vec}^{-1}_{sg_m,sg_m}\left[\boldsymbol{vec}(\mathbb{K}_{\boldsymbol{w}})\right]\mathbb{D}^{-1}_{\boldsymbol{w}} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.13 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.38 & 0.24 & 0.08 & 0.11 & 0.05 & 0.02 \\ \hline 0 & 0.23 & 0.66 & 0.60 & 0.38 & 0.14 & 0.18 & 0.09 & 0.03 \\ 0 & 0 & 0.24 & 0.22 & 0.34 & 0.16 & 0.18 & 0.10 & 0.03 \\ 0 & 0 & 0.08 & 0.08 & 0.16 & 0.22 & 0.20 & 0.19 & 0.11 \\ \hline 0 & 0 & 0.19 & 0.18 & 0.32 & 0.35 & 0.33 & 0.30 & 0.18 \\ 0 & 0 & 0.05 & 0.05 & 0.10 & 0.19 & 0.17 & 0.19 & 0.14 \\ 0 & 0 & 0.02 & 0.02 & 0.03 & 0.11 & 0.10 & 0.14 & 0.16 \end{array}\right] \tag{9}$$

Unstructured kinship

From equation 10, we get the unstructured kinship matrix, the symetric matrix which gives the expected number of any kin for *ego* taken at random in the stable state population:

$$\mathbb{K}^u = \begin{bmatrix} 0 & 0.32 & 0.07 \\ 0.32 & 0.80 & 0.41 \\ 0.07 & 0.41 & 0.58 \end{bmatrix} \tag{11}$$

Reading $\mathbf{K_u}$, we know that *ego* is expected to have 0.80 sisters, 0.58 cousins, 0.41 nieces and aunts and 0.07 grand-daughters and grandmothers (that is, the probability that its grandmother is alive is 0.07)

Kinship of unstructured population In order to illustrate one of the uses of these calculations in ecology and evolution, we generate $\mathbf{M}^u = \mathbf{F}^u + \mathbf{S}^u$ the unstructured model equivalent to $\mathbf{M} = \mathbf{F} + \mathbf{S}$ (see S.M.I). $\mathbf{M}^u$ is a scalar, that is a 1×1 matrix which says that all individuals in the population have the same probability to survive $\mathbf{S}^u = 0.53$ and the same expected number of offspring $\mathbf{F}^u = 0.47$. This model has the same asymptotic growth rate than $\mathbf{M}$: $\lambda = \mathbf{M}^u = 1.00$, as they model the same population. However, the kinship matrix of $\mathbf{M}^u$, generated assuming similarly Poisson reproduction yields a different output:

$$\mathbb{K}(\mathbf{M}^u) = \begin{bmatrix} 0 & 0.35 & 0.12 \\ 0.35 & 0.55 & 0.34 \\ 0.12 & 0.34 & 0.40 \end{bmatrix} \tag{31}$$

By comparing the kinship matrix of the unstructured population (eq.31) and the unstructured kinship of the structured population (eq.11), we can draw conclusions about the effect of the processes embedded in the classes of the model on kinship. Here the classes correspond to age classes (broadly speaking, the last one being open ended and therefore a "stage"), and they embed mostly, an increase in fertility with age and a decline in survival with age (senescence). The combination of these effects implies that the probabilities that direct ancestors (mother, grandmother) will be alive and the expected number of direct descendants (daughters, grand-daughters) will be lower than for the equivalent model without individual heterogeneity, where all individuals have the same vital rates (S.M.I). To the contrary, it implies that the number of other, indirect, kin, the descendants of ancestors (sisters, nieces, cousins and aunts) are more numerous. In other words, there is a trade-off between the number of direct and indirect kin as a response at the kinship level of the trade-off embedded in the life history of the population (increased allocation with age towards reproduction at the cost of survival).

Genealogical Markov chains

These are intermediate tools that are not necessary for the computation of the Kinship Matrix via eq 8 or eq 5. However they appear in the Kinship Formula itself (eq 3), we therefore provide their expression for the illustration, here. From λ and $\mathbf{w}$, we can construct $\mathbf{P}$ from eq13, $\mathbf{P_F} = \begin{bmatrix} 0.23 & 0 & 0 \\ 0.29 & 0 & 0 \\ 0.48 & 0 & 0 \end{bmatrix}$ and $\mathbf{P_S} = \begin{bmatrix} 0 & 1.00 & 0 \\ 0 & 0 & 0.54 \\ 0 & 0 & 0.46 \end{bmatrix}$. From these, it is easy to build $\mathbb{P} = \mathbb{P}_{\mathbb{S}} + \mathbb{P}_{\mathbb{F}} = \left[\begin{array}{c|c|c} \mathbf{P_S} & \mathbf{P_F} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P_S} & \mathbf{P_F} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{P_S} \end{array}\right]$

VI The kinship formula as an infinite sum

Eq.3 can also be written as

$$\mathbb{K} = \sum_{t=0}^{\infty} \mathbb{M}^t\left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{Z}\right)\mathbb{P}^t, \tag{7}$$

This can be seen by by applying eq.20 and eq.17, to eq.21, or, and it is easier, by multiplying both sides of eq.7 on the left by $\mathbb{M}$ and on the right by $\mathbb{P}$ and via the $T = t + 1$ change of index. The sum in eq.7 converges iff $\mathbf{S}$

does. This formulation as an infinite sum is the generation structured equivalent of that in (16) and corresponds to the natural approach to Kinship Demography when one considers the ancestor of *ego* at all time-steps in the past (18; 19). Contrary to these approaches however,eq.7 accounts for all kin relationships (up to the arbitrary distance g_m) simultaneously. Matrix $\mathbb{P}^t$ provides the probability distribution of the g_j ancestor of *ego* t time-steps before present. The term in parenthesis then considers younger sisters, older sisters and same-litter sisters of the $g_j - 1$ ancestor, and $\mathbb{M}^t$ accounts for their descendants today. Kinship Matrix $\mathbb{K}$ differs from and extends the matrix of the same name in (16), where it corresponds to a specific kin relationship. Equation 7, like eq. 5, and contrary to eq.3, provides a direct expression for $\mathbb{K}$.

Embedded maximum age

Consider the case where there is a maximum age embedded in the model (i.e., it is age-structured), that is there exist ω whereby $\mathbf{S}^\omega = \mathbf{0}$ (but $\mathbf{S}^{\omega-1} \neq \mathbf{0}$).In that case $\mathbf{P_S}^\omega = \mathbf{0}$ as well (if one cannot live more than ω time-steps, one cannot have been alive more than ω time-steps ago). This is also true in the generation structured model: $\mathbb{S}^\omega = \mathbb{P_S}^\omega = \mathbf{0}$. To the contrary, $\mathbf{F}$ and $\mathbb{F}$ do not behave similarly, and contrary to $\mathbf{F}$, we have for the generation structured model $\mathbb{F}^{g_m} = \mathbb{P_F}^{g_m} = \mathbf{0}$. Any product of a series of $\mathbb{F}$ and $\mathbb{S}$ such as $\mathbb{SSFFFSF}$, of any size, that contains more than g_m $\mathbb{F}$s is worth $\mathbf{0}$ since $\mathbb{F}$ increases the generation number by 1 and $\mathbb{S}$ preserves it. Therefore, as it contains either a run of $\mathbb{S}$s longer than $\omega - 1$ or more than $g_m - 1$ $\mathbb{F}$s, we have $\mathbb{M}^{g_m\omega} = \mathbf{0}$ and similarly $\mathbb{P}^{g_m\omega} = \mathbf{0}$. Each element of $(\mathbb{FJP_S} + \mathbb{SJP_F} + \mathbb{F}^*\mathbb{JP_F})$ contains either one $\mathbb{F}$ or one $\mathbb{P_F}$ and one $\mathbb{S}$ or one $\mathbb{P_S}$ and therefore

$$\mathbb{K} = \sum_{t=0}^{\omega(g_m-1)-1} \mathbb{M}^t \left(\mathbb{FJP_S} + \mathbb{SJP_F} + \mathbb{F}^*\mathbb{JP_F}\right)\mathbb{P}^t \tag{32}$$

VII Eventual number of kin

The Kinship Formula, as formulated in eq 7, provides the *current* number of kin of *ego*, that is , the number of kin that are currenlty alive. It may be interesting to know also the *eventual* number of kin, that is, the total number of kin that *ego* will eventually have, including those which died before *ego* was born and those who will be born after *ego*'s death. To achieve this, one natural first step is to consider the non-generation-overlapping version of the projection model $\mathbf{M} = \mathbf{F} + \mathbf{S}$. It corresponds to the next-generation matrix $\mathbf{R} = \mathbf{F}(\mathbf{I} - \mathbf{S})^{-1}$ (4). For simplicity, let us consider the case where there is only one class of offspring. Then, over a generation, we have $\mathbf{F} = R_0$, a scalar, the net reproductive rate, corresponding the $(1,1)$ entry of $\mathbf{R}$. We also have simply that $\mathbf{S} = 0$. From these, we get $\mathbf{P_F} = 1$, $\mathbf{P_S} = 0$ and $\omega = 1$. We are in the case here with only one offspring class (as the model has only one class) and therefore we can compute the number of same-litter daughters/sisters (these are the same here as $\mathbf{P_F} = 1$) which is a scalar: $\mathbf{Z} = \mathbf{F}^* = Z = \frac{Var(L\tilde{R}S)}{R_0} + R_0 - 1$, where $L\tilde{R}S$ is the Lifetime Reproductive Success, the expectation of which is R_0 (this can be computed from $\tilde{\mathbf{F}}$ and $\mathbf{S}$ via, for instance, Markov chains with rewards (see e.g. 20)). From eq 32, we can compute the Kinship Matrix for this model: $\mathbb{K} = \sum_{t=0}^{g_m-2} \mathbb{M}^t \left(\mathbb{F}^*\mathbb{JP_F}\right)\mathbb{P}^t$, which yields a diagonal matrix such that $\mathbb{K}_{g_i,g_i} = Z{R_0}^{g_i-2}$ and $\mathbb{K}_{1,1} = 0$. In other words, *ego* is expected to have $\mathbb{K}_{2,2} = Z$ sisters, $\mathbb{K}_{3,3} = ZR_0$ (first cousins), $\mathbb{K}_{4,4} = Z{R_0}^2$etc. By construction, apart from these kin of the same generation than *ego*, and therefore on the diagonal of $\mathbb{K}$, there is no other kin alive as generations do not overlap: *ego*'s mother and aunts are dead already and its daughters and nieces are not born yet.

eventual number of kin in the general case

To remedy this, and be able to compute the eventual number of kin, we add class "death" to the model. Obviously, in this case, the "survival" matrix of the model includes dead individual as well, and therefore $\mathbf{S}^t$ does not converge as $t \to \infty$. In the general case, this could imply that the kinship formula does not converge, however, we are here in a particular case where the individuals that are allowed to live forever cannot reproduce (that is, here, individuals in the class "death"). The kinship formula then converges and the kinship matrix is made easier to read by ensuring that all individuals to be counted are in the "death" class. This is done by summing the summands of eq.7 from $t = 0$ up to $t = \omega(g_m - 1) - 1 + \omega$ (instead of $t = \omega(g_m - 1) - 1$ as per eq32). Let us illustrate this for our non-generation overlapping, unstructured example.

illustration: eventual number of kin in a non-generation-overlapping population with one class of offspring

In this case, the components of the projection matrix become $\mathbf{S} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} R_0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\mathbf{Z} = \begin{bmatrix} Z & 0 \\ 0 & 0 \end{bmatrix}$. Similarly, we have $\mathbf{P_S} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, $\mathbf{P_F} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. And remembering that here, $\omega = 1$, we get

$$\mathbb{K}_{eventual} = \sum_{t=0}^{\omega(g_m-1)-1+\omega} \mathbb{M}^t \left(\mathbb{FJP_S} + \mathbb{SJP_F} + \mathbb{F}^*\mathbb{JP_F}\right)\mathbb{P}^t = \sum_{t=0}^{g_m-1} \mathbb{M}^t \left(\mathbb{FJP_S} + \mathbb{SJP_F} + \mathbb{F}^*\mathbb{JP_F}\right)\mathbb{P}^t$$

which, for $g_m = 4$ yields

$$\mathbb{K}_{eventual} = \sum_{t=0}^{3} \mathbb{M}^t \left(\mathbb{F}\mathbb{J}\mathbb{P}_{\mathbb{S}} + \mathbb{S}\mathbb{J}\mathbb{P}_{\mathbb{F}} + \mathbb{F}^{*}\mathbb{J}\mathbb{P}_{\mathbb{F}}\right)\mathbb{P}^t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_0 & Z & Z & Z & Z & Z & Z \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_0^2 & ZR_0 & ZR_0 & ZR_0 & ZR_0 & ZR_0 & ZR_0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_0^3 & ZR_0^2 & ZR_0^2 & ZR_0^2 & ZR_0^2 & ZR_0^2 & ZR_0^2 \end{bmatrix} \tag{33}$$

In this kinship matrix, we are only interested in the "death" class for both the kin and *ego* (those are highlighted in blue in eq 33), that is:

$$\mathbb{K}_e = \mathbb{K}_{eventual}[(*,2),(*,2)] = \begin{bmatrix} 0 & 1 & 1 & 1 \\ R_0 & Z & Z & Z \\ R_0^2 & ZR_0 & ZR_0 & ZR_0 \\ R_0^3 & ZR_0^2 & ZR_0^2 & ZR_0^2 \end{bmatrix} \tag{34}$$

VIII Code

R Code

```
# Inputs
F = matrix(c(0.23, 0.51, 0.84, 0, 0, 0, 0 ,0 ,0),nrow=3,byrow=T) #fertility component of matrix
S = matrix(c(0, 0, 0, .57 ,0 ,0,0, .54 ,.46),nrow=3,byrow=T)   #survival component of matrix
gmax=3  #number of generations investigated

#Z= ?   #same-litter-sister matrix

# eigen properties of the population projection matrix
M=F+S #projection matrix
ev <- eigen(M);lam <- max(Re(ev$values));
w_raw <- abs(Re(ev$vector))[, which.max(Re(ev$values))]
w <- w_raw/sum(w_raw)


#S<-myDiag(matrix(0, omega, omega),  rep(1, omega-1), -1) %x% S

# genealogical matrices
# w[w<0]<-0
#P=t((1/lam)*A*(w^(-1)%*%t(w)))
#P[is.nan(P)]<-0
#PF= (t(F)/t(A))*P
#PF[is.nan(PF)]<-0
#PS=P-PF

# Matrix Z
#POISSON reproduction
Fstar=F

#BERNOULLI reproduction
#Fstar=matrix(0, dim(F)[1],dim(F)[1])

#1 CLASS OF OFFSPRING
#Fstar=

#Z=Fstar%*%PF

#  Generation structured matrices
#s=dim(F)[1]
myDiag <- function(x, vec, k) {  x[row(x) == col(x) - k] <- vec
  x } #turns a vector (vec) into a zero matrix matrix but with vec on the kth diagonal
J1<-myDiag(matrix(0, gmax, gmax),  rep(1, gmax-1), -1)

bigF=J1%x%F
```

```
bigS=diag(gmax)%x%S
bigM=bigF+bigS

bigFstar=J1%x%Fstar

#temp=matrix(0, gmax, gmax);temp[2,2]=1;
#bigZ=temp%x%Z


# kinship matrix
temp=matrix(0, gmax, gmax);temp[1,1]=1;s=dim(F)[1]
J=temp%x%diag(s)

Dw <- diag(w)
bigDw <- diag(gmax)%x% Dw

vec <- function(x) {  y <- as.vector(x)
  return(y) }
vecKw=solve(diag(s*s*gmax*gmax)-bigM%x%bigM)%*%(bigS%x% bigF + bigF%x% bigS+ bigF%x% bigFstar)%*%vec(J*bigDw)
Kw <- matrix(vecKw, nrow = gmax * s, ncol = gmax * s)
K <- matrix(vecKw, nrow = gmax * s, ncol = gmax * s) %*% solve(bigDw)

Jy <- diag(gmax)%x%matrix(1, nrow = 1, ncol = s)
Ku <- matrix((Jy%x%Jy)%*%vecKw, nrow = gmax, ncol = gmax)
```

Matlab Code

```
%% inputs
%components of the projection matrix
F =[ 0.2300    0.5100    0.8400;  0   0   0;   0 0  0]
S = [ 0  0  0;  0.5700  0   0;   0    0.5400    0.4600]
%kinship distance investigated
gmax=3

%% Basic eigen-analysis
M=F+S %projection matrix
[w,lam]=eigs(M,1); %eigen analysis
lam % asymptotic growth rate
w=w/sum(w) %asymptotic relative abundances

%same-litter daughters
Fstar=F;

%%  Generation structured matrices/tensors
ss=size(F,1);
J1=diag(ones(1,gmax-1),-1);

bigF=kron(J1,F);
bigS=kron(eye(gmax),S);
bigM=bigF+bigS

bigFstar=kron(J1,Fstar);

temp=zeros(gmax,gmax);temp(1)=1;ss=length(F);
J=kron(temp,eye(ss));

Jy=kron(eye(gmax),ones(1,ss));

%% Direct Computation of expectation (of kinship matrix)
Dw=diag(w);
bigDw=kron(eye(gmax),Dw);

vecKw=((lam*eye(ss*ss*gmax*gmax)- kron(bigM,bigM))^(-1)) *
(kron(bigS,bigF)+kron(bigF,bigS)+kron(bigF,bigFstar))*vec(J*bigDw);

Kw=reshape(vecKw,[gmax*ss,gmax*ss]);
```

```
K=reshape(vecKw,[gmax*ss,gmax*ss])*(bigDw^(-1));
Ku=reshape(kron(Jy,Jy)*vecKw,[gmax,gmax]);

format shortg
round(K,2)
round(Ku,2)

%% vec operator
function [y] = vec(x)
    y=x(:);
end
```

References

[1] Lebreton JD, Gonzáles-Dávila G (1993) An introduction to models of subdivided populations. *Journal of Biological Systems* 01(04):389–423.

[2] Layton-Matthews K, Griesser M, Coste CF, Ozgul A (2021) Forest management affects seasonal source-sink dynamics in a territorial, group-living bird. *Oecologia* 196(2):399–412.

[3] Coste CF, Austerlitz F, Pavard S (2017) Trait level analysis of multitrait population projection matrices. *Theoretical Population Biology* 116:47–58.

[4] Caswell H (2001) *Matrix Population Models.* (Sinauer Associates Inc., Sunderland, MA.), p. 722.

[5] Henderson HV, Searle SR (1981) The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra* 9(4):271–288.

[6] Vindenes Y, Langangen Ø (2015) Individual heterogeneity in life histories and eco-evolutionary dynamics. *Ecology Letters* 18(5):417–432.

[7] Nisbet RM, Gurney WS (1983) The systematic formulation of population models for insects with dynamically varying instar duration. *Theoretical Population Biology* 23(1):114–135.

[8] Rees M, Ellner SP (2009) Integral projection models for populations in temporally varying environments. *Ecological Monographs* 79(4):575–594.

[9] Rees M, Childs DZ, Ellner SP (2014) Building integral projection models: A user's guide. *Journal of Animal Ecology* 83(3):528–545.

[10] Ellner SP, Rees M (2006) Integral Projection Models for Species with Complex Demography. *The American Naturalist* 167(3):410–428.

[11] Bacaër N (2009) Periodic Matrix Population Models: Growth Rate, Basic Reproduction Number, and Entropy. *Bulletin of Mathematical Biology* 71(7):1781–1792.

[12] Beissinger S, Westphal M (1998) On the Use of Demographic Models of Population Viability in Endangered Species Management. *The Journal of Wildlife Management* 62(3):821–841.

[13] Coste CFD (2024) Projecting structured populations I: components of individual heterogeneity and individual stochasticity. *in preparation.*

[14] Coste CF, 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.

[15] Jenouvrier S, Aubry LM, Barbraud C, Weimerskirch H, Caswell H (2017) Interacting effects of unobserved heterogeneity and individual stochasticity in the life-history of the Southern fulmar. *Journal of Animal Ecology* 38(1):42–49.

[16] Coste CF, et al. (2021) The kinship matrix: inferring the kinship structure of a population from its demography. *Ecology Letters* (June):1–13.

[17] Luo J, Fox BJ (1990) Life-table comparisons between two ground squirrels. *Journal of Mammalogy* 71(3):364–370.

[18] Goodman LA, Keyfitz N, Pullum TW (1974) Family formation and the frequency of various kinship relationships. *Theoretical Population Biology* 5(1):1–27.

[19] Caswell H (2021) The formal demography of kinship III : Kinship dynamics with time-varying demographic rates.

[20] Caswell H (2011) Beyond R0: Demographic Models for Variability of Lifetime Reproductive Output. *PLoS ONE* 6(6):e20809.