Pedigree-based estimation of reproductive value

Pedigree-based estimation of reproductive value

Darren C. Hunter[1*], Josephine M. Pemberton[2], Jill G. Pilkington[1] and Michael B. Morrissey[1**]

May 8, 2019

[1] Dyers Brae House
School of Biology
University of St Andrews
St Andrews, Fife
United Kingdom
KY16 9TH

[2] Institute of Evolutionary Biology
School of Biological Sciences
University of Edinburgh
Edinburgh
United Kingdom
EH9 3JT

[*] corresponding author, email: dch5@st-andrews.ac.uk

[**] corresponding author, email: michael.morrissey@st-andrews.ac.uk

Running head: *Hunter et al - Pedigree based estimation of reproductive value*

Keywords: reproductive value, genetic contribution, pedigree, Soay sheep, *Ovis aries*

Abstract

How successful an individual or cohort is, in terms of their genetic contribution to the future population, is encapsulated in the concept of reproductive value, and is crucial for understanding selection and evolution. Long-term studies of pedigreed populations offer the opportunity to estimate reproductive values directly. However, the degree to which genetic contributions, as defined by a pedigree, may converge on their long-run values within the time frames of available datasets, such that they may be interpreted as estimates of reproductive value, is unclear. We develop a system for pedigree-based calculation of the expected genetic representation that both individuals and cohorts make to the population in the years following their birth. We apply this system to inference of individual and cohort reproductive values in Soay sheep (*Ovis aries*) from St Kilda, Outer Hebrides. We observe that these genetic contributions appear to become relatively stable within modest time frames. As such, it may be reasonable to consider pedigree-based calculations of genetic contributions to future generations as estimates of reproductive value. This approach and the knowledge that the estimates can stabilise within decades should offer new opportunities to analyse data from pedigreed wild populations, which will be of value to many fields within evolutionary biology and demography.

Introduction

The concept of fitness is central to the study of natural selection and evolution (Endler, 1986). At a given time, natural selection, in the absence of frequency dependence, is generally expected to lead to the maximisation of mean population fitness (Lande, 1976). However, fluctuating environments greatly complicate this maximisation (Lande, 2007). Overlapping generations and fluctuating environments are frequently present in natural populations, and make it less clear how to effectively estimate the fitness that selection is acting on in these systems. It is possible that insights can be gained using data from long term studies which include pedigree information. Such studies not only offer a chance to look at fitness, realised and expected, many generations in the future, but also provide the opportunity to access what predictions would have been made only a few generations after data collection started. Modern genotyping has further increased the value of these studies by allowing construction of high accuracy pedigrees (Sardell *et al.*, 2010; Huisman, 2017). This means that questions about fitness that were previously limited to theoretical studies can start to benefit from information using empirical pedigree data.

Regardless of the exact definition, estimating individual fitness is about characterising how many descendants individuals have left, usually for the purpose of understanding how environmental conditions, class structure, or phenotypic traits influence the underlying propensities of individuals to leave descendants. This is commonly thought of as the genetic contribution that an individual makes to the following generation, relative to other individuals in that population. It is, therefore, common to the use the number of offspring produced as a fitness estimate (Clutton-Brock, 1988) or the total number of grand-offspring (Hunt *et al.*, 2004; Bolund & Lummaa, 2017), accounting accordingly for the relationships to different descendants. This kind of fitness definition can be problematic when parents and offspring interact (Wolf & Wade, 2001; Hadfield, 2012; Thomson & Hadfield, 2017). Furthermore, fitness estimates of this type are complicated by the presence of age structure and overlapping generations in a population. There is general agreement, at least in theory, that fitness is a measure of genetic representation in the future population (Stearns, 1976; Charlesworth, 1980; Endler, 1986). This appears to naturally lead to an estimate of individual fitness that is determined by a combination of an individual's reproductive output and the survival, and reproduction, of its descendants.

An alternative concept closely related to fitness, which accounts for both age structure and overlapping generations, is the reproductive value. As a concept it has caused much debate and confusion since the idea was first published (Samuelson, 1978; Caswell, 2001; Crow, 2002). The idea is usually attributed to Fisher in 1930 although he actually first discussed it in a paper three years earlier (Fisher, 1927). The discovery of this 1927 paper has gone a long way to explaining much of the confusion that has surrounded his writing in 1930

(Crow, 2002). Despite this, there are still questions around the correct interpretation in its original form and there has been much discussion about what exactly Fisher meant, under what conditions it applies, and to what extent the ideas were already present in the literature (Crow, 2002; Grafen, 2006; Galindo, 2007). However, there is general consensus that the reproductive value is the expected contribution an individual will make to some future population given its current age (Grafen, 2006; Crow, 2002). Theoretically it tracks a gene down an individual's pedigree, seeing how well that gene is represented in the future population with the expectation conditional on the pedigree. It does not, therefore, take into account the effect of the genetic background on which that gene finds itself, and is expected in most situations to differ greatly from the realised genetic contribution (Barton & Etheridge, 2011). As time passes estimations initially considered as expected genetic contributions will converge on the theoretical idea of a reproductive value (Grafen, 2006).

Typically, reproductive values are calculated using survival and fecundity information arranged in matrix projection models, which give a reproductive value for each age or stage included in the Leslie matrix (Caswell, 2001). This approach can be extended to get trait-value specific reproductive values using integral projection models (Merow *et al.*, 2014). By necessity these models use current information to project forward, assuming that the life history of the population remains constant through time (Caswell, 2001). In situations where these assumptions may not hold it is difficult to assess the accuracy of predictions from these models. Theoretical work, using simulated bi-parental pedigrees, has shown that individual reproductive values are expected to stabilise within ten generations. The value at which these, theoretical, individual reproductive values stabilise is largely determined within a few initial generations (Barton & Etheridge, 2011). However, work on how the expected genetic contribution actually behaves over time, in wild populations and in fluctuating environments, is currently lacking.

To investigate the potential use of pedigree-based estimates of genetic representation as empirical proxies for reproductive value, we developed an approach for calculating genetic representation of any individual, or group of individuals, from an additive genetic relatedness matrix. It became apparent at an early stage that these calculations can be greatly affected by pedigree incompleteness, which can be particularly acute when populations are not closed, as is the case for most monitored wild populations. We therefore present a method that accounts for pedigree incompleteness by expressing genetic representation of focal individuals in proportion to all representation attributable to them and their contemporaries. We developed this system in the course of analyses of the pedigree from the Soay sheep (*Ovis aries*) population on St Kilda, in the Outer Hebrides (Clutton-Brock & Pemberton, 2004). We present results using our method, and investigate the time frame required for calculations of expected genetic contributions to stabilise, such that they may be interpretable as estimates of reproductive value.

Algorithm

In this section, we describe a general system for calculation of genetic representation in a population conditional on a pedigree, that would be interpretable as realisations of reproductive value, once stabilised. In the next section, we apply this algorithm to an empirical dataset, in order to investigate the time frames required for this stabilisation to be approached.

In order to generate statistics on genetic representation arising from descent, we developed a system of truncating a pedigree such that relatedness between a set of focal individuals and the rest of the pedigree would represent only relatedness arising from direct descent. For a population which consists of a set of individuals, P_t, at time, t, a focal individual, or set of individuals (e.g., a specific sex and/or cohort), F_t, are identified for which calculation of future genetic contributions are required. A separate set of all the individuals alive in the population, $P_{t+\Delta_t}$, at the future time point of interest, $t + \Delta_t$, is also compiled. For any F_t, pedigree links between these focal individuals and their parents are deleted while links to all known descendants are retained, creating a truncated pedigree. An additive genetic relatedness matrix, $\mathbf{A}$, for P_t consists of elements $A_{ij} = 2\Theta_{ij}$ (Walsh & Lynch 2018, chapter 19), where Θ_{ij} is the coefficient of coancestry between individuals i and j (Lynch & Walsh 1998, chapter 7). An altered genetic relationship matrix, A^*_{ij}, can instead be constructed using the truncated pedigree based on F_t. In a pedigree from which all ancestors of a focal individual (or individuals) i have been removed, coancestry between such individuals and another individual j arises solely from descent. The sum of the relationships in this matrix between each individual, i, in F_t and $P_{t+\Delta_t}$ is calculated, $n_{F_t,\Delta t} = \sum_{i=1}^{N_{F_t}} \sum_{j=1}^{N_{P_{t+\Delta t}}} A^*_{ij}$, where i indexes the N_{F_t} individuals in F_t (which may be a single individual) and j indexes the $N_{P_{t+\Delta t}}$ individuals in $P_{t+\Delta t}$. This values gives an estimated genetic contribution, n_{F_t} for each focal individual, i to the future population (Box 1). For example, if a focal individual was still alive then their existence would contribute a value of 1 to the summed value. Similarly, offspring add 0.5 each. Repeating the process for $P_{t+\Delta_t}$, using increasing values of Δ_t, allows investigation of how the estimated genetic contribution of F_t to the population changes temporally. As $\Delta_t \to \infty$, the $n_{F_t} \to v_{F_t}$, where v_{F_t} is the combined reproductive value for F_t. Calculations of genetic representation may be made separately for different sets of P_t, e.g., for male and female focal individuals, and for specific cohorts, estimated genetic contributions must be calculated using all extant individuals at $t + \Delta_t$, regardless of their sex or year of birth.

Calculations of genetic contributions as described to this point represent the absolute representation of individuals or groups in terms of expected genome copies. The relevant quantity for most biological questions will concern not the absolute number of copies, but the proportional representation in the future state of the population. As such, expected number of genome copies must be standardised by dividing by population

size at $P_{t+\Delta_t}$. More generally, the standardisation may be made by the maximum possible representation of an individual or group of individuals in $P_{t+\Delta_t}$, in relation to the representation from all extant individuals in the population at the time of the focal individuals' birth (i.e n_{F_t} when $F_t = P_t$) . This more general standardisation will provide for sensible calculations of reproductive value in open populations (Box 1). In practice, this standardisation requires that the pedigree-based representation of all extant individuals, P_t be calculated, exactly as described above for F_t. All extant non-focal individuals may be treated as one large cohort of unrelated individuals, by deleting all parental links for extant individuals at a given time. Then, as above, their contributions to $P_{t+\Delta_t}$ may be obtained using the relationship matrix derived from this modified pedigree.

Application

Study System

The Soay sheep (*Ovis aries*) population of St Kilda, in the Outer Hebrides, has been the subject of an individual-based, long term study since 1984 (Clutton-Brock & Pemberton, 2004). All sheep that are part of the core population are individually marked. The date of birth of the majority of lambs born within the study area is known through observational data. Adult migrants into the population are also tagged so that they can be individually identified. Detailed population monitoring ensures that the date that most individuals appeared in the population, through birth or migration, and the date that individuals died, can be determined with high precision. Population size can vary greatly between years (Clutton-Brock & Pemberton, 2004). Highly reliable maternities are known from observational data during the spring. Maternities are regularly confirmed, and occasionally corrected (primarily for the attribution of maternity to still-born lambs) using genetic data. Paternities are assigned on the basis of 384 SNP loci, chosen on the basis of high minor allele frequency and an even distribution throughout the genome (Bérénos *et al.*, 2014). Paternity assignment is conducted using MasterBayes (Hadfield *et al.*, 2006) and SEQUOIA (Huisman, 2017). Paternity assignments are made with $> 99\%$ confidence in the majority of cases (6542 of 7014 individuals).

Data Selection

Our analysis includes individuals born alive during or after the spring of 1985 up until the spring of 2015. The majority of mortality occurs over the winter and early spring. We therefore consider an individual to have survived a given year, and transitioned to the next age group, if they were known to have survived until the end of April of the following calendar year. This allowed an estimate of the number of individuals

of each age group alive in any given year, with the first age group consisting of new born individuals. After this point any reference to a year will refer to a census year, rather than a calendar year. Winter survival, and inclusion in the following census year, is defined as survival into May.

In order to acquire annual, age-specific, survival estimates each individual needs to be assigned as currently being either alive or dead and all dead individuals need to have a year of death. For individuals with no recorded death information, the last time the sheep had been (i) assigned as a parent, indicating that it was alive during a specific rut (November) for males, or spring for females (ii) captured, or (iii) observed in one of 30 annual censuses of the study area, was used to estimate a minimum possible date of death. Individuals were assumed to have died during the winter following the last evidence that they were alive. For individuals without a known birth year the date that they first appeared in the dataset was estimated similarly. Ignoring rare cases where individuals did not have a sex recorded (or were castrates) there were 4912 female and 5387 male individuals in the pedigree, of which 3223 females and 3491 males have both a birth and death year recorded. Individuals that were known to be born after 1984 numbered 3354 and 3955. A total of 2247 individuals (1144 males and 546 females) required an estimated death year.

We applied the calculations described in the "*Algorithm*" section, using the `makeA()` function of the `nadiv` package (Wolak, 2012) to generate the modified **A** representing relatedness between focal and descendant individuals, arising only from direct descent. The process was repeated with each cohort separately set as F_t with $t + \Delta_t$ being each consecutive year after birth.

Properties of pedigree-based expected genetic contributions in individuals and cohorts

There is a non-trivial level of immigration of males into the population, with a mean of 9 individuals per year since 1985 (Figure 1A). These males do not generally take up residence in the study area, but rather are individuals that are observed in the study area during the rut, and that sire offspring. Female immigration is much lower, with a mean of 1.4 (Figure 1A) individuals per year. While the numbers of immigrants are modest, immigrants tend to have substantial reproductive success, with immigrant males, in particular, siring on the order of 20% of lambs born in the study area (Figure 1B). Consequently, the denominator in our standardisation for genetic contributions, i.e., the total discernible contribution to the future genetic constitution of the population, decreases with the length of time that elapses between the existence of the focal individual(s) and the time at which subsequent genetic contribution is ascertained (Figure 2). The calculation we make that spans the greatest time period is for the total contribution of males extant in 1985 to the genetic composition of the population in 2015. The proportion of the genomes in the population in

2015 accounted for by the pedigree from extant males in 1985 is 0.078 (out of a possible 0.5 in complete absence of immigration and other causes of pedigree incompleteness). This calculation is slightly more favourable for females, where a proportion of 0.206 of the 2015 population genomes can be attributed to extant females in 1985. This disparity between the sexes will have arisen primarily because genetic sampling was less complete early in the study, such that paternities are somewhat sparser in early years.

Individual genetic contributions stabilise relatively quickly (Figures 3 & 4). Although fluctuations in individual values occur across all time intervals that we can assay, several aspects of stability are evident, even from within the lifespan of a Soay sheep. For males (Figure 3), individuals with very low representation after five years rarely become major contributors of descendants to the population. Among males with non-trivial representation after approximately five years, visual inspection of figure 3 shows that large subsequent changes in the rank order of genetic contributions are rare. Consequently, many of the correlations of individual genetic contributions in the year of birth + T, with individual contributions in the final year of data (2015), are often as high as 0.8 by five years after birth (Figure 5).

Fluctuations through time in the genetic contributions of females to the future genetic composition of the population (Figure 4) are of qualitatively similar magnitude and distribution to males (Figure 3). However, because the variance in fitness in females is lower, these fluctuations occur across a narrower range of values, and so the rank order of contributions changes more through time (Figure 4). Consequently, the correlation of female individual contributions with those in the final year (Figure 5B) does not typically increase as quickly as that for males (Figure 5A).

The total cohort estimated genetic contributions start to stabilise even more quickly than individual contributions, with many of the cohorts appearing to be relatively stable after around five years for both sexes (Figure 6). The largest source of within- and among-cohort variation in genetic contributions occurs in the first year; declines associated primarily with overwinter survival seem to be the main determinants of the long-term genetic contributions of each cohort (Figure 6). While the genetic representation of cohorts can become relatively stable very quickly, there are also notable fluctuations in some cohorts between the ages of approximately three and eight (Figure 6). These are normally reductions in genetic representation, arising because of the death of an unusual number of individuals of a given cohort. Because of the unstable dynamic of the population size and composition, few cohorts experience similar sequences of demographic rates, and so it is hard to formulate possible explanations for these fluctuations. It may be relevant that during this period of the life cycle (i.e., reproductive adults), the total reproductive value of a given cohort is concentrated in the smallest number of individuals. As such, these fluctuations could primarily be manifestations of demographic stochasticity, i.e., drift. The initial increase in the correlations for the cohorts is faster than that seen in the individual correlations (Figures 5 & 7).

Discussion

The concept of reproductive value is central to much theory in population and evolutionary biology. However, reproductive values are rarely used in empirical practice, particularly in evolutionary studies such as those estimating the form of natural selection. The method presented here is both intuitive and easily implemented, and thus has the potential to facilitate empirical studies using estimates of reproductive values for individuals and groups of individuals. The necessary information to implement the approach is collected in many long-term studies in the form of pedigree information, and is the same information necessary for inference of quantitative genetic parameters, which is increasingly common in the wild (Wilson *et al.*, 2010). The type of information that is estimated should be valuable to ecologists, demographers and quantitative geneticists and we expect analysis based on this method will be useful for answering questions in many subject areas. The fact that the estimated genetic contributions appear to settle as quickly as they do, even in a variable environment, might be somewhat unexpected but should help reassure people that their calculation, and use, is worthwhile for the long term datasets currently available.

Intuition might initially suggest that the unstable dynamic of the Soay sheep population (Clutton-Brock & Pemberton 2004; Figure 8A) size and structure would act to delay stabilisation of reproductive values. However, any such effect seems to be modest, at least at the cohort level. Figure 8B&C shows initial trajectories of groups of four cohorts born preceding and after years of major reductions in population size. Differences among cohorts in their ultimate genetic contributions are largely determined by the levels of first year mortality that they experience. This first year mortality, and the consequences for the longer term breeding success of that cohort, is highly influenced by the population size that year Coltman *et al.* (1999). Despite the number of the adult females, and consequently lamb production, remaining relatively stable through time (Figure 8A), there is substantial variation in first year survival. This is represented in Figure 8A by the notable fluctuations in the size of yearling component of the population among years. Even outside of years with major reductions in population size, very high first year mortality can occur (consider especially the 1997 and 2010 cohorts in Figure 8B&C, particularly for males). These mortality rates seem to determine genetic representations more than subsequent conditions. For cohorts beyond the lamb stage, experiencing a year with a major reduction in population size has very little effect on demographic contributions. This counter-intuitive property of the estimated demographic contributions is particularly evident in the 1997-2000 group of cohorts, in Figure 8B&C, the 1997 and 1998 cohorts experienced multiple large reductions in population size during their first years of life, but have nonetheless relatively stable genetic contributions (Figure 6) through the large reduction in population size that happened between 2001 and 2002 (Figure 8A). This occurs because, while such cohorts suffer high mortality, so too do the other cohorts of animals of

similar age. As such, high mortality reduces the absolute size of a cohort, but it does not necessarily greatly change the size of the cohort in proportion to the sizes of other extant cohorts.

Pedigree-based estimation of reproductive value could prove particularly useful in the study of cohort effects, which is an area of substantial current interest. The idea that early life conditions play a role in the future success of groups of individuals, often referred to as cohort effects, is well established as an important, but complex, component of population dynamics in wild populations (Albon *et al.*, 1987; Beckerman *et al.*, 2002; Lindstrom & Kokko, 2002). Differences in the vital rates due to conditions experienced by different cohorts have now been recorded in many different wild populations where long term data are available (Nussey *et al.*, 2007; Hamel *et al.*, 2009; Pigeon *et al.*, 2017). These environmental variables have long lasting consequences on the average success of individuals within a given cohort (Forchhammer *et al.*, 2001). Calculations of net reproductive values of individual cohorts could prove a very useful and powerful way of encapsulating how successful an individual or, group of individuals, is for studies of cohort effects, by relating the general notion of success to a quantity that is maximised by natural selection. Most calculations of reproductive value assume a stable age distribution (Caswell, 2001), and so do not return cohort-specific values that represent effects of stochasticity, and thus cannot be related to differences among cohorts. Since all aspects of environmental stochasticity that are relevant to reproductive value will be represented in a population's pedigree, the promise of using pedigrees may overcome a key hurdle to relating the notion of success of a cohort to the firm foundation generated by the concept of reproductive value. It is of note that some progress has been made on algorithms for calculating reproductive values in stochastic environments (Tuljapurkar, 1989; Tuljapurkar & Lee, 1997).

Engen *et al.* (2009, 2011, 2012, 2014) have shown that the effective average value of selection in a stochastic environment, accounting for both survival and reproduction, within and/or across age classes, is given if selection coefficients (gradients and/or differentials) are calculated using contribution of individuals to the population's total reproductive value at the next time step as a measure of fitness. Applications of this so far (e.g. Kvalnes *et al.*, 2016) have made these calculations using mathematics based on the stable age distribution to obtain age- or class-specific values for reproductive value. However, in a stochastic environment, the age distribution will not be stable, and furthermore, age- and class-specific reproductive values may vary in time. Calculating estimates of reproductive value based on expected genetic contributions given by a pedigree may provide further opportunities for assaying fitness and inferring selection, especially in studies of environmental stochasticity.

Pedigree incompleteness, particularly due to high levels of migration, may be the biggest challenge in allowing estimated genetic contributions to be assessed in other wild populations. The suitability of this approach would have to be considered for each study system separately and appropriate adjustments made.

For some purposes, such as understanding immigration itself, immigration may not be a problem, but rather a process that pedigree-based calculation of reproductive value (or similar approaches, see Chen *et al.* 2019) is well-suited to tackle. Pedigrees may also be incomplete if insufficient molecular are a available to make parentage assignments. Insufficient molecular data may also lead to non-trivial rates of erroneous parentage assignments. These kinds of complexities of empirical pedigrees are probably not a major issue in the Soay sheep study system, where the majority of paternities are determined by overwhelming molecular data (Bérénos *et al.*, 2014), but they could present important considerations in other systems.

Despite the existence of important theoretical work indicating that reproductive value is the quantity most directly maximised by natural selection (Grafen, 2006), it does not necessarily follow that empirical estimates of reproductive value, for example based on genetic contributions calculated from a pedigree, are sensible measures of fitness for studies of natural selection. For example, when individuals interact, measures of reproductive success that conflate the direct fitness (e.g., production of fertilised zygotes) of the interacting individuals will not generate measures of selection that correctly predict evolution (Hadfield, 2012; Thomson & Hadfield, 2017). Another situation where pedigree-based estimates of reproductive value could be very misleading would be in estimation of the genetic variance of fitness, for example as motivated by the fundamental theorem of natural selection (Fisher, 1930). All contributions of demographic stochasticity to a pedigree will generate covariance in estimated genetic contributions between progenitors and descendants, in proportion to their relatedness. This same relatedness between individuals in a pedigree is precisely the information that mixed model-based approaches (Wilson *et al.*, 2010) use to estimate genetic parameters. Consequently, demographic stochasticity would almost certainly generate upwardly biased estimates of genetic variance for fitness, if such analyses were conducted on pedigree-based estimates of reproductive value.

It is worth noting that although we have discussed the calculation of expected genetic contributions as potential estimators of reproductive values, which are the target of selection (Grafen, 2006), these are not the same as the actual genetic contributions that an individual, or cohort, will ultimately make to the future population. This is discussed in detail by Barton & Etheridge (2011). The pedigree-based genetic contributions we calculated are based on expected genetic relationships between individuals using descent. While parents always contribute equally to offspring (autosomal) genomes, it does not follow that all grandparents contribute exactly one quarter of the genetic compliment of their grandoffspring. Rather, because of segregation and recombination in parents, grandparental contributions to their grandoffspring follow a random distribution (Hill, 1993). Theory predicts that these actual genetic contributions will take far longer to equilibrate than will expected contributions (Barton & Etheridge, 2011). This does not invalidate the concept of reproductive value as estimated using a pedigree for studying natural selection

and demography, since the deviations between expected and realised contributions represent contributions of segregation to genetic drift, which will generally be unrelated to other, more deterministic processes.

Calculation of reproductive values allows quantification of success in a way that is intuitive as an ecological concept, while at the same time directly relates to the target of selection, the reproductive value, as understood at a theoretical level (Grafen, 2006). The ability to calculate the same metric at both the cohort and individual level offers new opportunities to compare these values. Additionally, tracking how estimated genetic contributions change through time allows important dynamics within a population to be identified. The presence of large fluctuations, or points of stabilisation, within a given system, offers the opportunity to link key moments in the determination of realised genetic contribution to causal ecological factors. Wider application of the developed approach may be limited due to the required completeness of the pedigree and life-history data. Nonetheless, the timescales of the stabilisation observed within the Soay sheep data demonstrate the potential for pedigree-based reproductive values to provide valuable information, potentially even in studies spanning shorter time periods.

Acknowledgements

We thank the National Trust for Scotland and Scottish Natural Heritage for permission to work on St Kilda and NYS, QinetiQ and Eurest for providing logistics and other support on the island. We thank all the project members and volunteers who have helped with field work on the island and everyone who has contributed to keeping the project going over many years, including S. Albon, T. Clutton-Brock, T. Coulson, M. Crawley, L. Kruuk, D. Nussey, J. Slate and A. Wilson. The long term project on St Kilda has been largely funded by the UK Natural Environment Research Council. M. B. Morrissey is supported by a University Research Fellowship from the Royal Society (London). D. C. Hunter is funded by a Ph.D Scholarship from the University of St. Andrews.

References

Albon, S.D., Clutton-Brock, T.H. & Guinness, F.E. 1987. Early Development and Population Dynamics in Red Deer. II. Density-Independent Effects and Cohort Variation. *The Journal of Animal Ecology* **56**: 69. doi:10.2307/4800. URL `https://www.jstor.org/stable/4800?origin=crossref`.

Barton, N.H. & Etheridge, A.M. 2011. The relation between reproductive value and genetic contribution. *Genetics* **188**: 953–973. doi:10.1534/genetics.111.127555.

337 Beckerman, A., Benton, T.G., Ranta, E., Kaitala, V. & Lundberg, P. 2002. Population dynamic conse-
338 quences of delayed life-history effects. *Trends in Ecology and Evolution* **17**: 263–269. doi:10.1016/S0169-
339 5347(02)02469-2.

340 Bérénos, C., Ellis, P.A., Pilkington, J.G. & Pemberton, J.M. 2014. Estimating quantitative genetic pa-
341 rameters in wild populations: A comparison of pedigree and genomic approaches. *Molecular Ecology* **23**:
342 3434–3451. doi:10.1111/mec.12827.

343 Bolund, E. & Lummaa, V. 2017. The effects of resource availability and the demographic transition on the
344 genetic correlation between number of children and grandchildren in humans. *Heredity* **118**: 186–192.
345 doi:10.1038/hdy.2016.81. URL `http://dx.doi.org/10.1038/hdy.2016.81`.

346 Caswell, H. 2001. *Matrix population models., Second ed.*. Wiley Online Library. doi:
347 10.1002/9780470057339.vam006m.

348 Charlesworth, B. 1980. *Evolution in age-structured populations.* Cambridge University Press, Cambridge,
349 UK.

350 Chen, N., Juric, I., Cosgrove, E.J., Bowman, R., Fitzpatrick, J.W., Schoech, S.J., Clark, A.G.
351 & Coop, G. 2019. Allele frequency dynamics in a pedigreed natural population. *Proceed-*
352 *ings of the National Academy of Sciences* **116**: 2158–2164. doi:10.1073/pnas.1813852116. URL
353 `http://www.pnas.org/lookup/doi/10.1073/pnas.1813852116`.

354 Clutton-Brock, T.H. 1988. *Reproductive success: studies of individual variation in contrasting breeding*
355 *systems*. University of Chicago Press.

356 Clutton-Brock, T.H. & Pemberton, J.M. (eds) 2004. *Soay Sheep: Dynamics and Selection in an Island*
357 *Population*. Cambridge University Press, Cambridge, UK.

358 Coltman, D.W., Smith, J.A., Bancroft, D.R., Pilkington, J., MacColl, A.D.C., Clutton-Brock, T.H. &
359 Pemberton, J.M. 1999. Density-Dependent Variation in Lifetime Breeding Success and Natural and
360 Sexual Selection in Soay Rams. *The American Naturalist* **154**: 730–746. doi:10.1086/303274. URL
361 `https://www.journals.uchicago.edu/doi/10.1086/303274`.

362 Crow, J.F. 2002. Perspective: Here's to Fisher, additive genetic variance, and the fundamental the-
363 orem of natural selection. *Evolution; international journal of organic evolution* **56**: 1313–6. URL
364 `http://www.ncbi.nlm.nih.gov/pubmed/12206233`.

Endler, J.A. 1986. *Natural Selection in the Wild*. Monographs in population biology. Princeton University Press. URL http://books.google.co.uk/books?id=MYk1XbelDssC.

Engen, S., Kvalnes, T. & Sæther, B.E. 2014. Estimating phenotypic selection in age-structured populations by removing transient fluctuations. *Evolution* **68**: 2509–2523. doi:10.1111/evo.12456.

Engen, S., Lande, R. & Sæther, B.E. 2011. Evolution of a plastic quantitative trait in an age-structured population in a fluctuating environment. *Evolution* **65**: 2893–2906. doi:10.1111/j.1558-5646.2011.01342.x.

Engen, S., Lande, R., Saether, B.E. & Dobson, F.S. 2009. Reproductive value and the stochastic demography of age-structured populations. *The American Naturalist* **174**: 795–804. doi:10.1086/647930.

Engen, S., Sæther, B.E., Kvalnes, T. & Jensen, H. 2012. Estimating fluctuating selection in age-structured populations. *Journal of Evolutionary Biology* **25**: 1487–1499. doi:10.1111/j.1420-9101.2012.02530.x.

Fisher, R.A. 1927. The actuarial treatment of official birth records. *The Eugenics review* **19**: 103–8. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2987494/.

Fisher, R.A. 1930. *The genetical theory of natural selection*. The Clarendon Press.

Forchhammer, M.C., Clutton-Brock, T.H., Lindström, J. & Albon, S.D. 2001. Climate and population density induce long-term cohort variation in a northern ungulate. *Journal of Animal Ecology* **70**: 721–729. doi:10.1046/j.0021-8790.2001.00532.x.

Galindo, C. 2007. On Fisher's reproductive value and Lotka's stable population. In: *Annual Meeting of the Population Association of America*, pp. 1–26. New York. URL http://paa2007.princeton.edu/download.aspx.

Grafen, A. 2006. A theory of Fisher's reproductive value. *Journal of Mathematical Biology* **53**: 15–60. doi:10.1007/s00285-006-0376-4. URL http://link.springer.com/10.1007/s00285-006-0376-4.

Hadfield, J. 2012. The quantitative genetic theory of parental effects. *The evolution of parental care. Oxford Univ. Press, Oxford, UK* pp. 267–284.

Hadfield, J.D., Richardson, D.S. & Burke, T. 2006. Towards unbiased parentage assignment: Combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology* **15**: 3715–3730. doi: 10.1111/j.1365-294X.2006.03050.x.

Hamel, S., Gaillard, J.M., Festa-Bianchet, M. & Côté, S.D. 2009. Individual quality, early-life conditions, and reproductive success in contrasted populations of large herbivores. *Ecology* **90**: 1981–1995. doi: 10.1890/08-0596.1. URL http://doi.wiley.com/10.1890/08-0596.1.

Hill, W.G. 1993. Variation in genetic identity within kinships. *Heredity* **71**: 652–653.

Huisman, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources* **17**: 1009–1024. doi:10.1111/1755-0998.12665. URL http://doi.wiley.com/10.1111/1755-0998.12665.

Hunt, J., Bussière, L.F., Jennions, M.D. & Brooks, R. 2004. What is genetic quality? *Trends in Ecology and Evolution* **19**: 329–333. doi:10.1016/j.tree.2004.03.035.

Kvalnes, T., Saether, B.E., Haanes, H., Røed, K.H., Engen, S. & Solberg, E.J. 2016. Harvest-induced phenotypic selection in an island population of moose, Alces alces. *Evolution* **70**: 1486–1500. doi:10.1111/evo.12952. URL http://www.ncbi.nlm.nih.gov/pubmed/27174031.

Lande, R. 1976. Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* **30**: 314. doi:10.2307/2407703. URL http://www.jstor.org/stable/2407703?origin=crossref.

Lande, R. 2007. Expected Relative Fitness and the Adaptive Topography of Fluctuating Selection. *Evolution* **61**: 1835–1846. doi:10.1111/j.1558-5646.2007.00170.x. URL http://doi.wiley.com/10.1111/j.1558-5646.2007.00170.x.

Lindstrom, J. & Kokko, H. 2002. Cohort effects and population dynamics. *Ecology Letters* **5**: 338–344. doi:10.1046/j.1461-0248.2002.00317.x. URL http://doi.wiley.com/10.1046/j.1461-0248.2002.00317.x.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer.

Merow, C., Dahlgren, J.P., Metcalf, C.J.E., Childs, D.Z., Evans, M.E.K., Jongejans, E., Record, S., Rees, M., Salguero-Gómez, R. & McMahon, S.M. 2014. Advancing population ecology with integral projection models: A practical guide. *Methods in Ecology and Evolution* **5**: 99–110. doi:10.1111/2041-210X.12146.

Nussey, D.H., Kruuk, L.E., Morris, A. & Clutton-Brock, T.H. 2007. Environmental conditions in early life influence ageing rates in a wild population of red deer. *Current Biology* **17**: R1000–R1001. doi:10.1016/j.cub.2007.10.005. URL http://linkinghub.elsevier.com/retrieve/pii/S0960982207020222.

Pigeon, G., Festa-Bianchet, M. & Pelletier, F. 2017. Long-term fitness consequences of early environment in a long-lived ungulate. *Proceedings of the Royal Society B: Biological Sciences* **284**: 20170222. doi:10.1098/rspb.2017.0222. URL http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2017.0222.

Samuelson, P.A. 1978. Generalizing Fisher's "reproductive value": "Incipient" and "penultimate" reproductive-value functions when environment limits growth; linear approximants for nonlinear Mendelian mating models. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 6327–31. URL `http://www.jstor.org/stable/68957 http://www.ncbi.nlm.nih.gov/pubmed/16592600 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC393175`.

Sardell, R.J., Keller, L.F., Arcese, P., Bucher, T. & Reid, J.M. 2010. Comprehensive paternity assignment: Genotype, spatial location and social status in song sparrows, Melospiza Melodia. *Molecular Ecology* **19**: 4352–4364. doi:10.1111/j.1365-294X.2010.04805.x.

Stearns, S.C. 1976. Life-history tactics: a review of the ideas. *The Quarterly Review of Biology* **51**: 3–47. URL `http://www.ncbi.nlm.nih.gov/pubmed/778893`.

Thomson, C.E. & Hadfield, J.D. 2017. Measuring selection when parents and offspring interact. *Methods in Ecology and Evolution* **8**: 678–687. doi:10.1111/2041-210X.12795.

Tuljapurkar, S. 1989. An uncertain life: Demography in random environments. doi:10.1016/0040-5809(89)90001-4.

Tuljapurkar, S. & Lee, R. 1997. Demographic uncertainty and the stable equivalent population. *Mathematical and Computer Modelling* **26**: 39–56. doi:10.1016/S0895-7177(97)00168-4.

Walsh, B. & Lynch, M. 2018. *Evolution and selection of quantitative traits*. Oxford University Press.

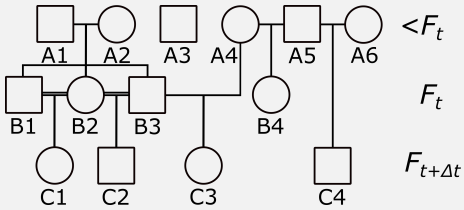
Wilson, A.J., Reale, D., Clements, M.N., Morrissey, M.B., E. Postma, C.A.W., Kruuk, L.E.B. & Nussey, D.H. 2010. An ecologist's guide to the animal model. *Journal of Animal Ecology* **79**: 13–26.

Wolak, M.E. 2012. Nadiv: An R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution* **3**: 792–796. doi:10.1111/j.2041-210X.2012.00213.x.

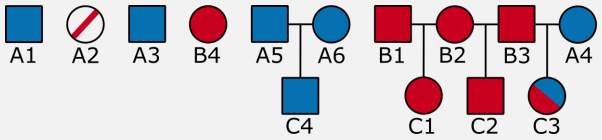
Wolf, J.B. & Wade, M.J. 2001. On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *Journal of Evolutionary Biology* **14**: 347–356. doi:10.1046/j.1420-9101.2001.00277.x.

Box 1: Example calculation of cohort expected genetic contribution

Pedigree (A) is an imaginary pedigree of what might actually be observed, looking at two cohorts (F_t & $F_{t+\Delta t}$), born at times t and $t+\Delta t$ respectively, along with their ancestors. Calculating the contribution that individuals in this fictitious F_t cohort made to the population at $t+\Delta t$ would use an altered pedigree as shown in B. Here all the information regarding the ancestors of the F_t cohort is removed and, as a consequence, there is no information about the relatedness between the individuals in this cohort. In pedigree (B) individuals in the population at $t+\Delta t$ that arise from the F_t cohort are shaded red (including these individuals themselves) while contributions from individuals that were alive before t are shaded in blue. Matrix (C) is an additive relationship matrix, A^*_{ij}, constructed based on the relationships shown in pedigree (B), with all individuals from the F_t cohort, along with all individuals present in the population before this, considered to be unrelated founders. In matrix (C), as in pedigree (B), blue cells relate to contributions to the population at $t+\Delta t$ by individuals from $< F_t$ while red cells are the contributions from the F_t cohort. The sum of each red column is the expected genetic contribution, n_{F_t}, of a F_t individual in $t+\Delta t$ while the sum of all red cells is this value for the cohort as a whole. The sum of all shaded cells is used to correct these values for fluctuations in population size and for immigration.



(A) An example pedigree showing two cohorts (F_t & $F_{t+\Delta t}$) along with their ancestors.



(B) An altered pedigree based on pedigree (A) where the relationships between the F_t cohort and their ancestors have been removed. Individuals shaded red are those which contribute to the F_t cohorts' expected genetic contribution. Those with a line through them died prior to $t+\Delta t$.

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	C1	C2	C3	C4
A1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	1	0	0	0	0	0	0	0	0	½	0
A5	0	0	0	0	1	0	0	0	0	0	0	0	0	½
A6	0	0	0	0	0	1	0	0	0	0	0	0	0	½
B1	0	0	0	0	0	0	1	0	0	0	½	0	0	0
B2	0	0	0	0	0	0	0	1	0	0	½	½	0	0
B3	0	0	0	0	0	0	0	0	1	0	0	½	½	0
B4	0	0	0	0	0	0	0	0	0	1	0	0	0	0
C1	0	0	0	0	0	0	½	½	0	0	1	¼	0	0
C2	0	0	0	0	0	0	0	½	½	0	¼	1	¼	0
C3	0	0	0	½	0	0	0	0	½	0	0	¼	1	0
C4	0	0	0	0	½	½	0	0	0	0	0	0	0	1

(C) An additive relationship matrix, A^*_{ij}, constructed using pedigree (B). The sum of the cells shaded red is the uncorrected F_t total cohort expected genetic contribution, while each column of red shading is each F_t individual's uncorrected expected genetic contribution. The sum of all the shaded cells is total genetic contribution made by all extant individuals to the $t+\Delta t$ population and this value is used to correct the expected genetic contribution for the F_t cohort for changes in population size and migration.

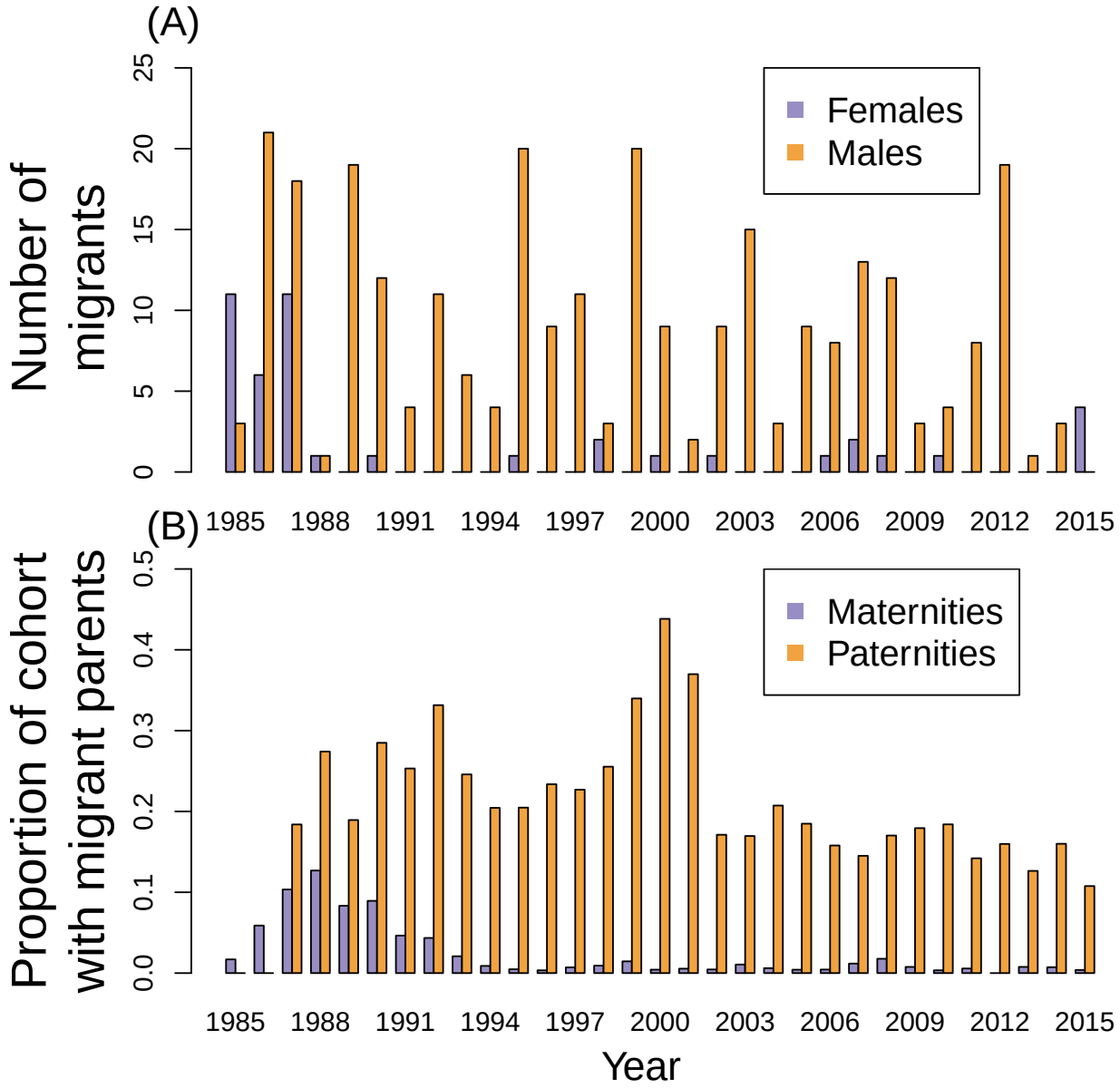


Figure 1: Rates of migration into the study system (A) in terms of numbers of individuals that ultimately bred in the study area, by sex, and (B) proportions of maternities and paternities attributed to individuals that are not known to have been born in the study area.

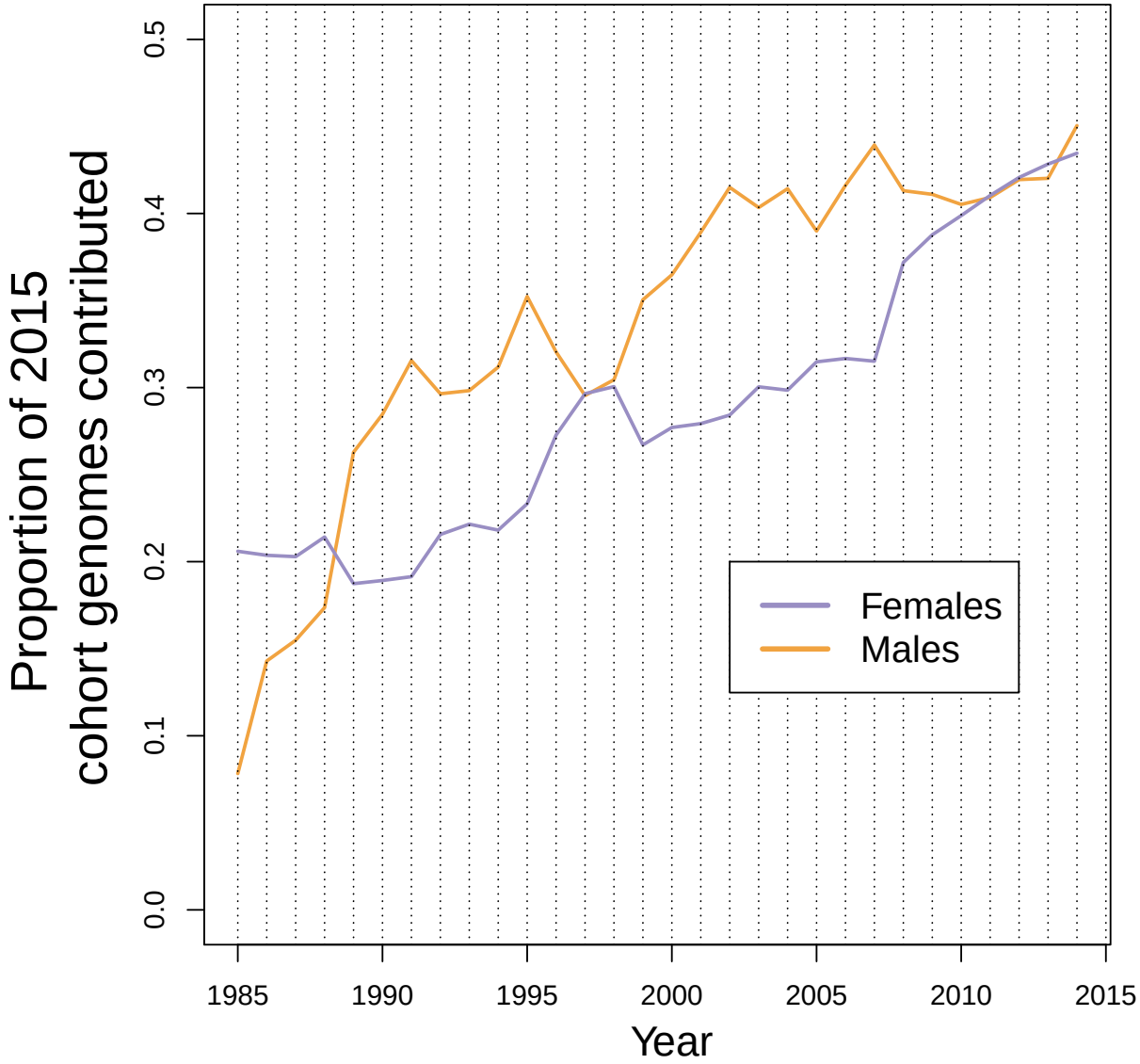


Figure 2: Proportion of the genomes attributable to the past population. The proportion of genomes present in the 2015 cohort which are attributable to extant individuals in all previous years for both females (purple) and males (orange). In the absence of any migration this contribution would constantly be 0.5 for each sex.

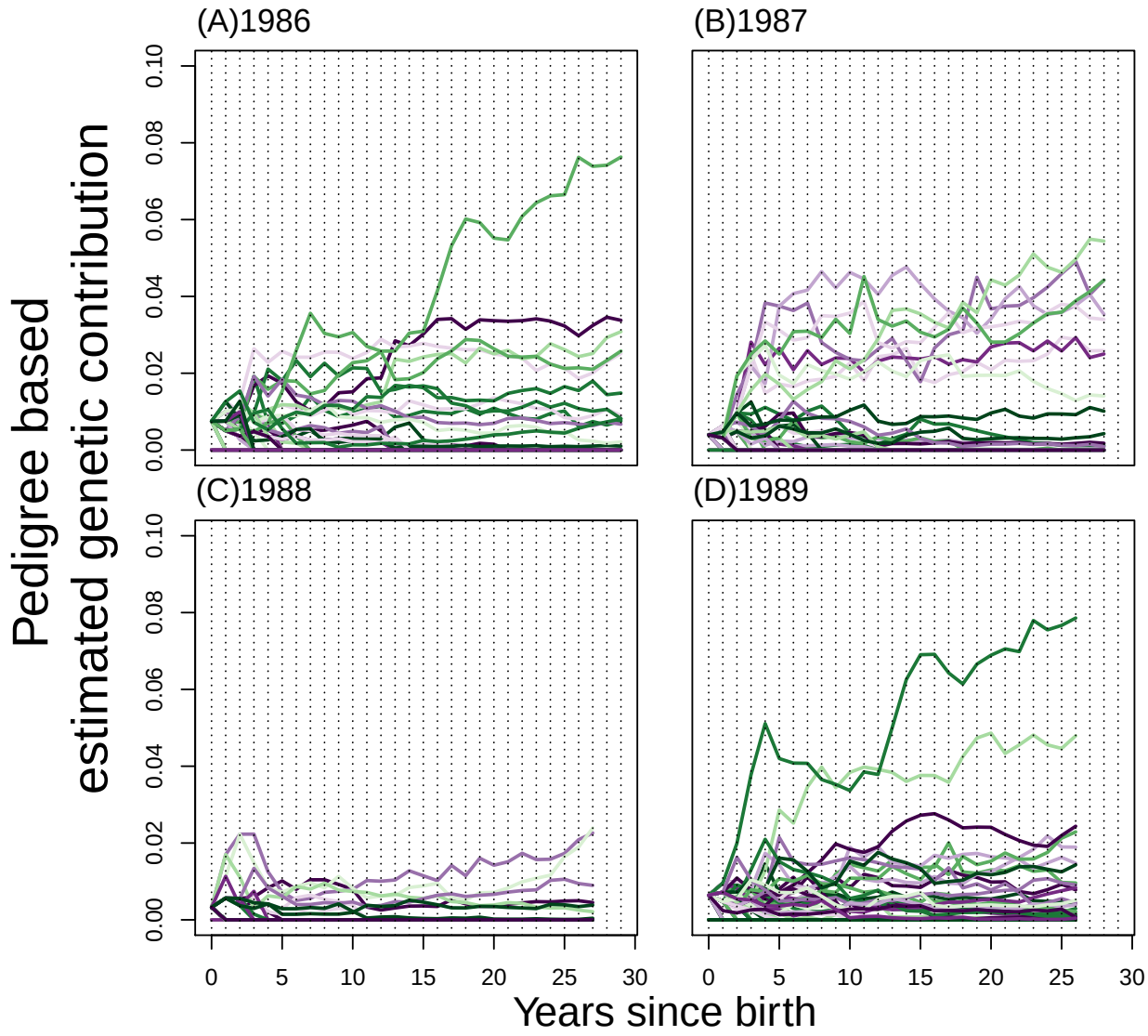


Figure 3: Individual, pedigree based, estimated genetic contributions of males from some of the earliest cohorts. The 1986 to 1989 cohorts are plotted with each line representing a separate individual born that year, showing their pedigree based genetic contributions estimated each year after their birth.

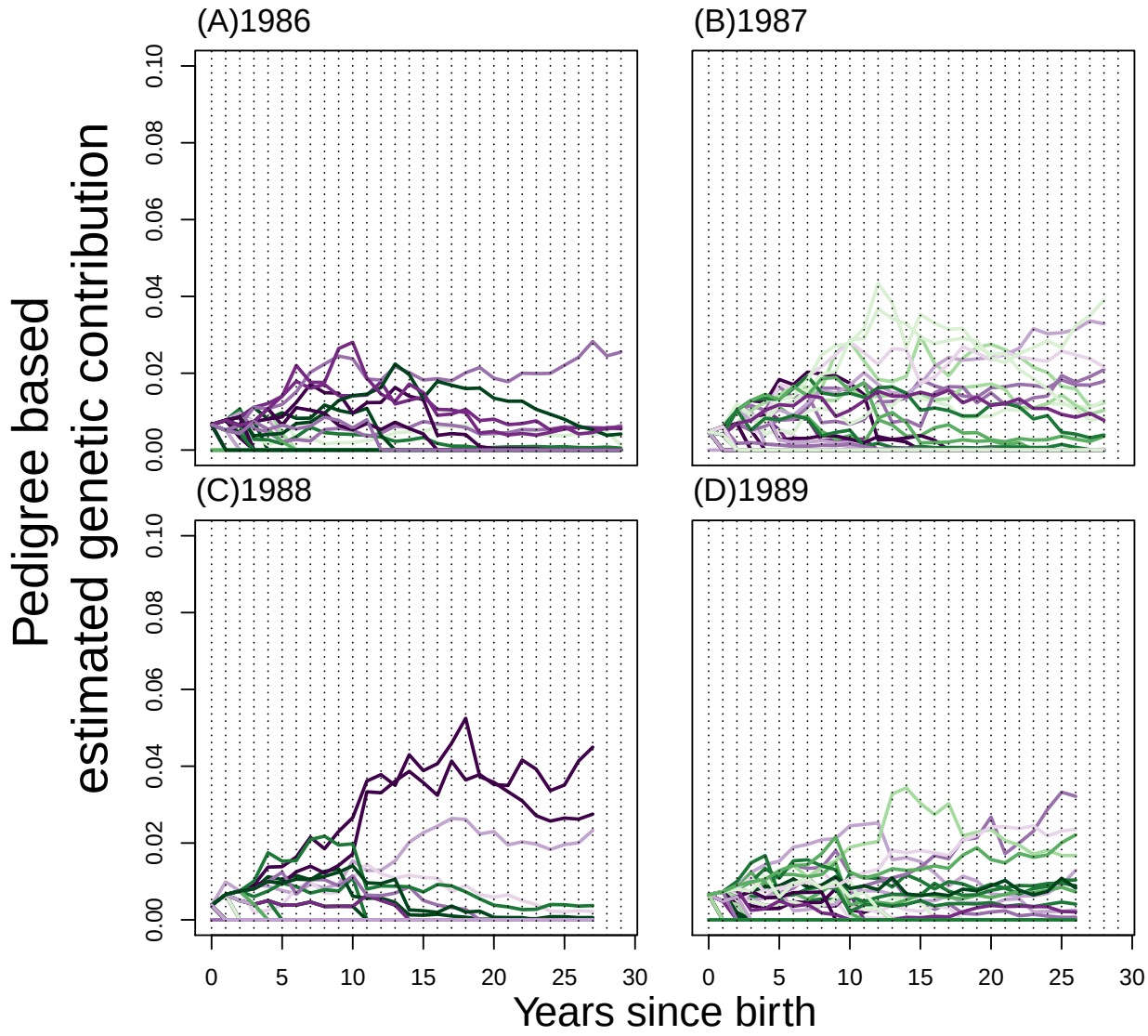


Figure 4: Individual, pedigree based, estimated genetic contributions of females from some of the earliest cohorts. The 1986 to 1989 cohorts are plotted with each line representing a separate individual born that year, showing their pedigree based genetic contributions estimated each year after their birth.

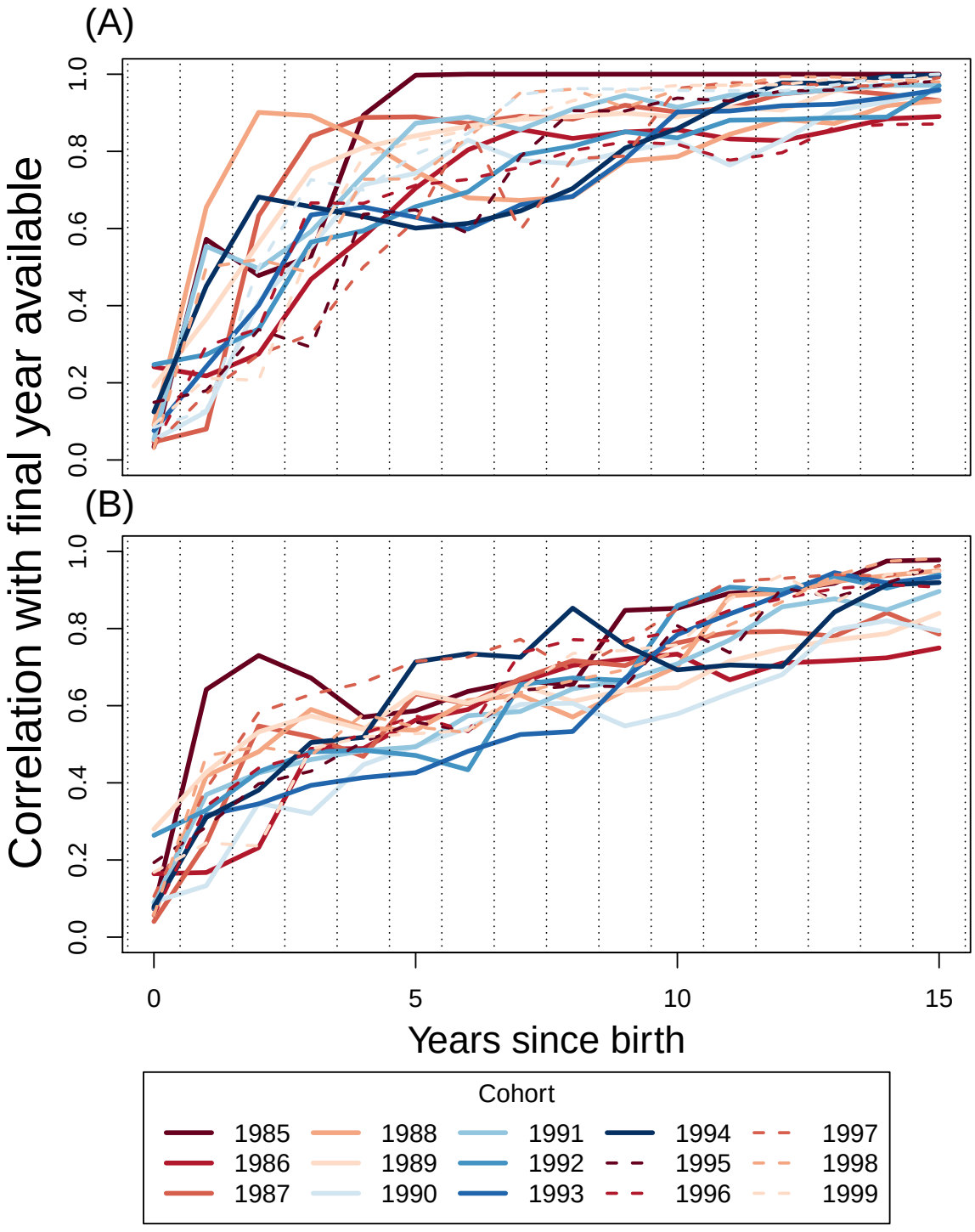


Figure 5: Correlations between pedigree-based estimated genetic contributions for the 1985 through 1999 cohorts, for (A) males and (B) females, with the last year available (2015). Each line represents the correlation of individuals genetic contributions in a given year after birth, with those in 2015, for separate cohorts.

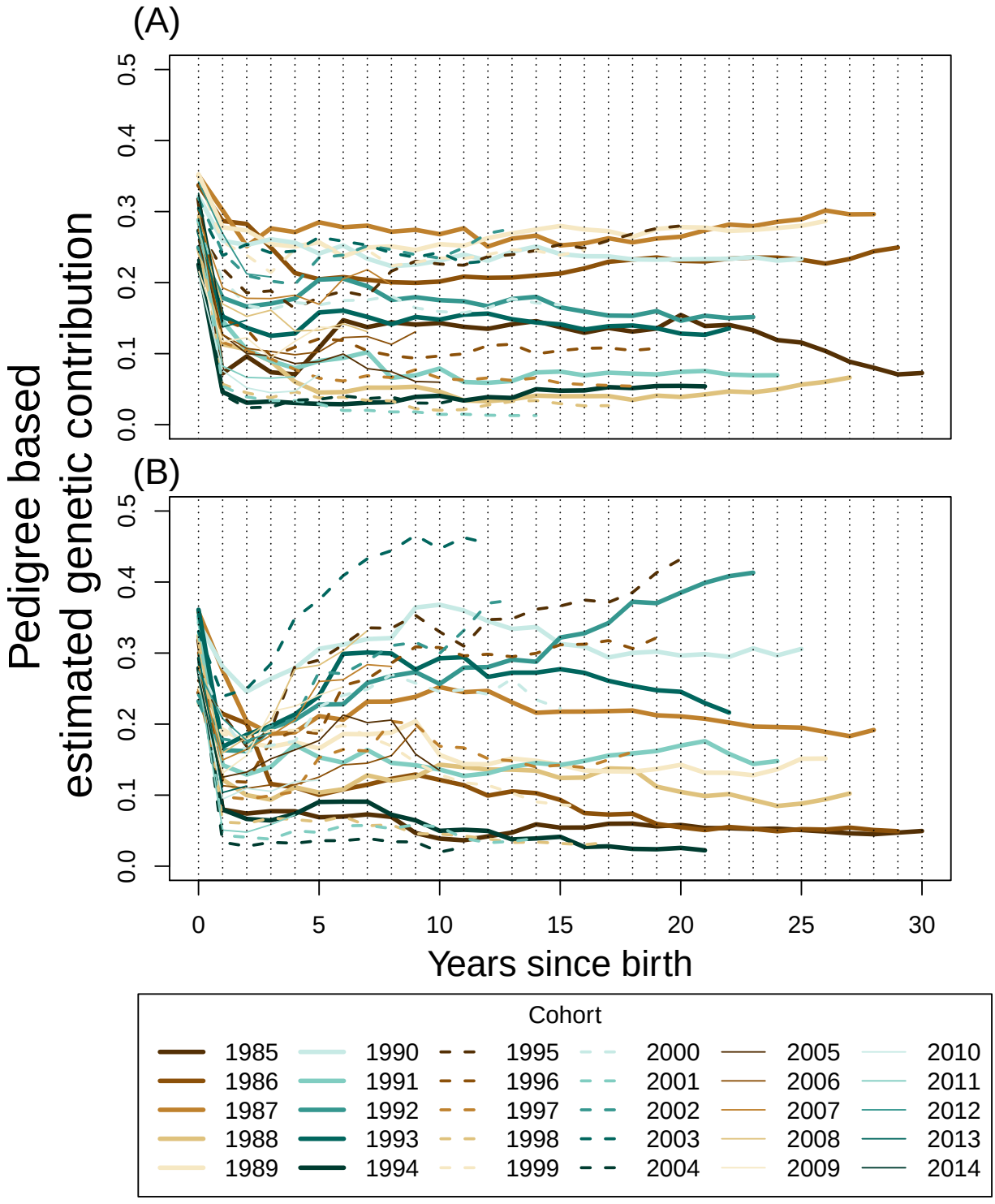


Figure 6: Total pedigree based estimated genetic contributions for (A) male and (B) female cohorts. Each line represents a separate cohort from 1985 to 2014.

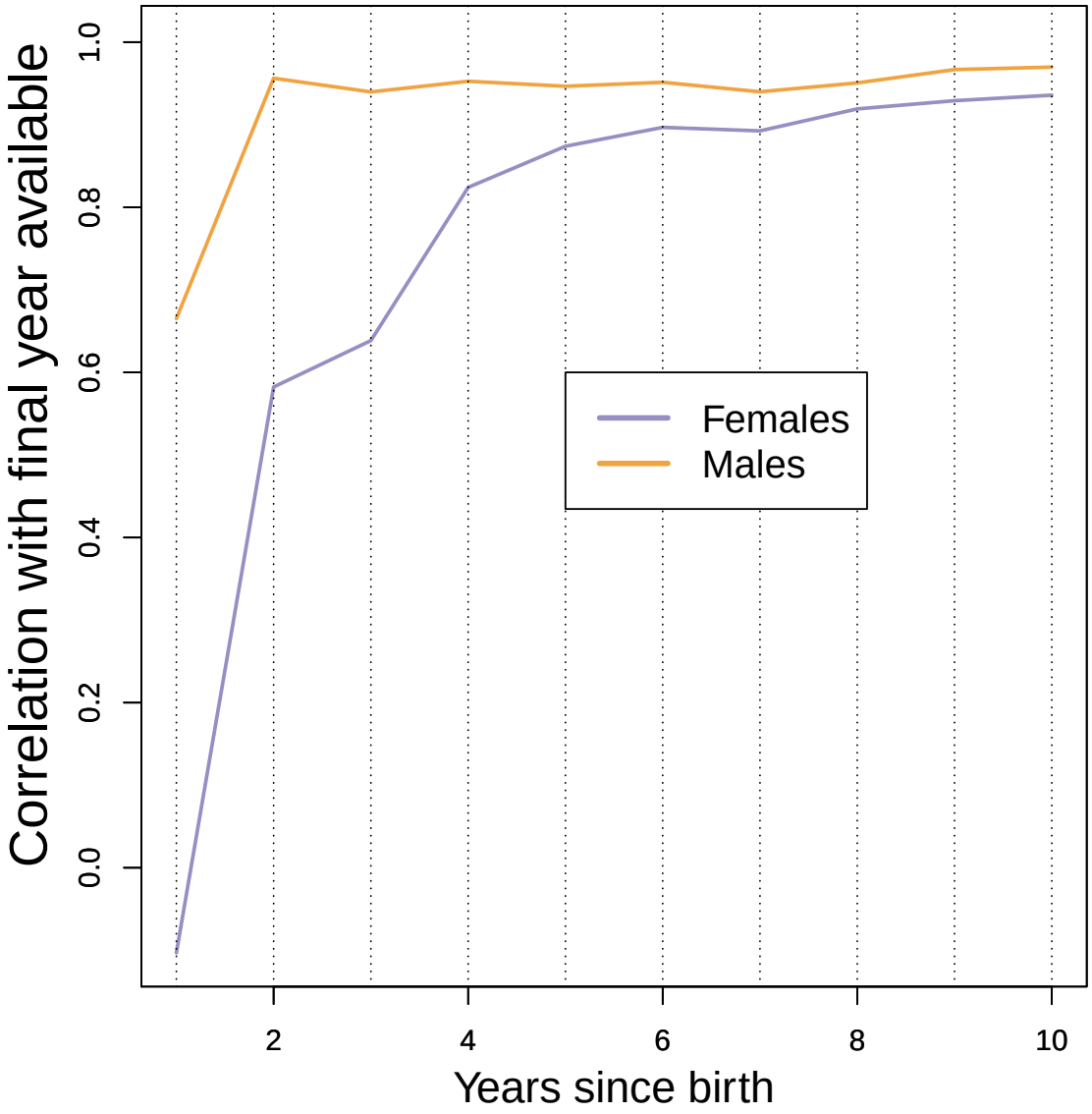


Figure 7: Correlations between the pedigree based, estimated genetic contributions for each cohort from 1985 to 2004, in years after their birth, with the last year available (2015). It can be seen that at time progresses the correlation with the pedigree based estimated genetic contributions for the cohorts in 2015 quickly becomes very high for both males and females.

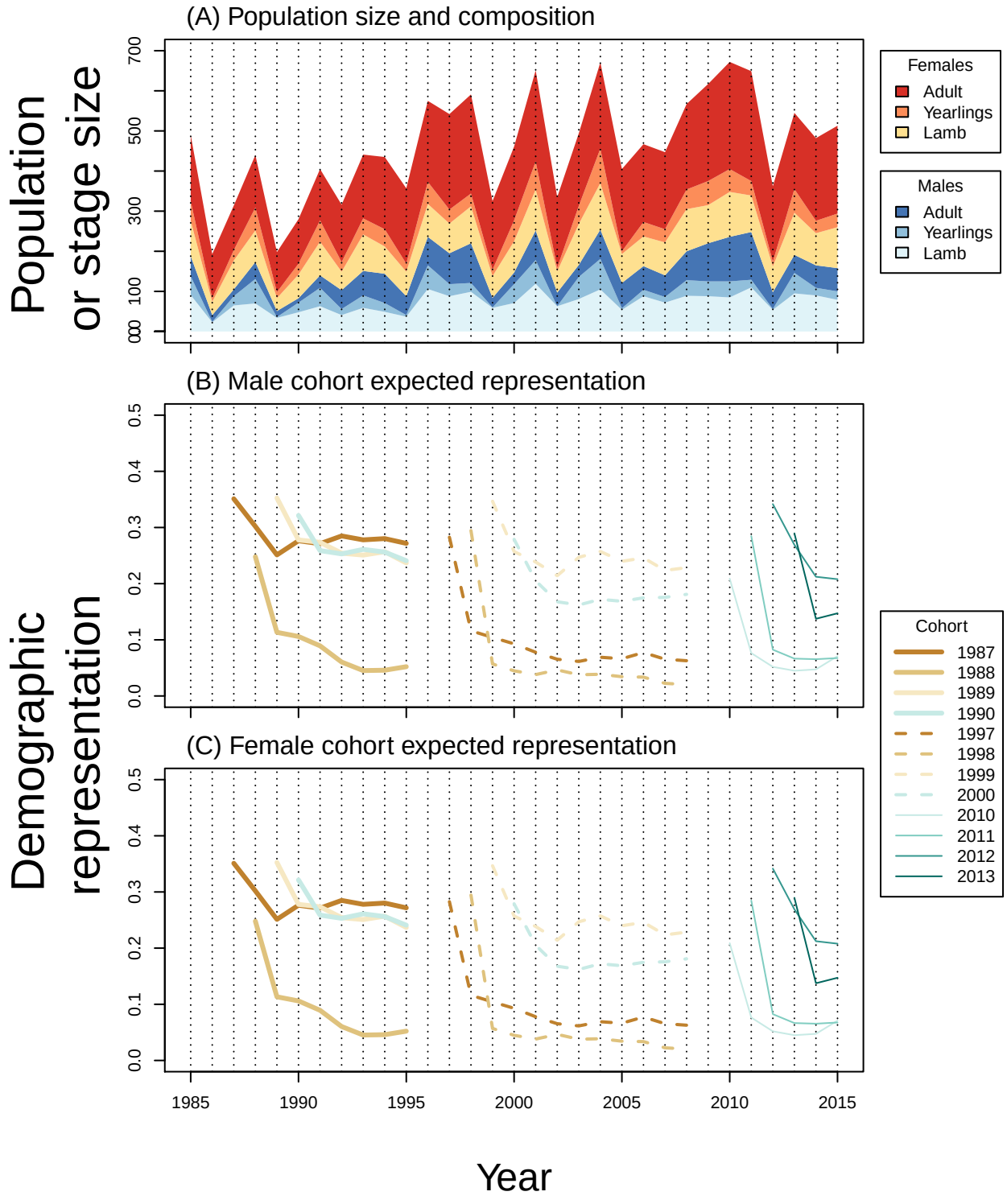


Figure 8: The dynamic of the size and composition of the population of Soay sheep using the study area in Village Bay on St Kilda (A). The class sizes in part (A) represent individuals resident in the study area, and are therefore slightly different from the data from which we calculate expected genetic contributions. (B) and (C) show estimated genetic contributions of selected male and female cohorts respectively to the population in future time intervals. Focal cohorts are identifiable from the year in which each respective line begins in parts (B) and (C); line colours and styles correspond to those used to identify cohorts in figures 5 and 6.