OXFORD

Genetics and population analysis

purgeR: inbreeding and purging in pedigreed populations

Eugenio López-Cortegano

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK

Abstract

Summary: Inbreeding depression and genetic purging are important processes shaping the survivability and evolution of small populations. However, detecting purging is challenging in practice, in part because there are limited tools dedicated to it. I present a new R package to assist population analyses on detection and quantification of the inbreeding depression and genetic purging of biological fitness in pedigreed populations. It includes a collection of methods to estimate different measurements of inbreeding (Wright's, partial and ancestral inbreeding coefficients) as well as purging parameters (purged inbreeding, and opportunity of purging coefficients). Additional functions are also included to estimate population parameters, allowing to contextualize inbreeding and purging these results in terms of the population demographic history. purgeR is a valuable tool to gain insight into processes related to inbreeding and purging, and to better understand fitness and inbreeding load evolution in small populations.

Availability and implementation: purgeR is an R package available at CRAN, and can be installed via `install.packages("purgeR")`. Source code is maintained at a GitLab repository (https://gitlab.com/elcortegano/purgeR).

Contact: e.lopez-cortegano@ed.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Inbreeding may result in the decline of biological fitness due to the increase in the frequency of homozygote genotypes for deleterious recessive alleles, causing the so-called 'inbreeding depression' (Charlesworth and Charlesworth, 1987). However, as deleterious alleles become more exposed under inbreeding, selection also becomes more effective removing, or 'purging', them (Charlesworth, 2018; Glémin, 2003). Consequently, the expectation of population fitness evolution under inbreeding changes with purging, possibly allowing for the survival of small populations (Hedrick and García-Dorado, 2016).

Pedigree information has attracted the attention of genetic purging models, as these allow for direct inferences on single individuals based on their genealogical history, and many conservation programs maintain studbooks with pedigree records. Boakes and Wang (2005) used ancestral inbreeding coefficients (F_a, Ballou, 1997) to measure the consequences of purging on fitness, given the expectation that individuals born from more inbred ancestors are expected to be more fit than individuals with the same level of inbreeding but less inbred ancestors. Gulisija and Crow (2007) developed a method to evaluate the potential reduction in the individual inbreeding load (B) using the probability of transmission of highly deleterious recessive alleles under inbreeding. García-Dorado (2012) defined a purged inbreeding coefficient (g) that measures the expected frequency of recessive deleterious loci in homozygosity, as a function of a purging coefficient (d) that relates to the recessive component of deleterious effects.

While many software packages have been developed to compute inbreeding, resources for purging analysis are more limited. To my knowledge, F_a is only computed by a few software packages (Baumung *et al.*, 2015; Doekes *et al.*, 2020; García-Dorado *et al.*, 2016), and only PURGd estimates g (García-Dorado *et al.*, 2016). No informatic tool is available to compute Gulisija and Crow's parameters for the opportunity of purging. purgeR computes all these parameters and others, including parameters related to population diversity and demography (e.g. effective population size, N_e), all functions being introduced in tutorials accessible via `browseVignettes("purgeR")`.

2 Input data

For illustrative purposes here, a population with known fitness and $B = 4.4$ was simulated with size $N = 10^3$ for 10^3 generations and then bottlenecked to $N = 25$ for 50 generations using SLiM 3.5 (Haller and Messer, 2019), under conditions favorable to the detection of purging, similarly as in García-Dorado *et al.* (2016). Details on the mutational model used and code to reproduce the simulation are given in Supplementary File S1. The simulated pedigree is included as Supplementary Table S1. Input pedigrees are required to be 'data.frame' objects in R, and to include individual, maternal and paternal identities.
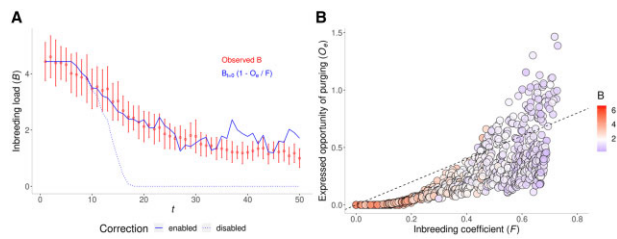
Fig. 1. Inbreeding load and opportunity of purging. (**A**) Observed decline of *B* over generations (in read, mean values as points and error bars spanning one standard deviation). Expected *B* is given in blue lines, calculated as $B_{t=0}(1 - O_e/F)$, where O_e/F represents the normalized opportunity of purging (Gulisija and Crow, 2007). The solid line corresponds to corrected O_e estimates, and dotted line to raw O_e values. (**B**) Relationship between O_e and *F*, colored by value of *B* (white is the median pedigree value *B* = 1.8; increasingly red and blue coloration indicate higher and lower *B* values, respectively). A dashed line indicates the value $O_e = F$. Code can be found in Supplementary File S4

3 Opportunity of purging

Here, the computation of opportunity of purging measures is presented, since this is the major novelty in purgeR compared to its predecessor PURGd (from which many functions are reimplemented), apart from other improvements related to portability, performance and usability. Some assays on the performance of purgeR are given in Supplementary File S2.

Total (*O*) and expressed (O_e) opportunity of purging can be computed for every individual to provide an estimate of the expected reduction in *B*. While *O* measures the potential reduction of *B* as a consequence of having inbred ancestors, O_e relates to the reduction in expressed *B* as a consequence of having inbred ancestors, and being homozygous for alleles derived from them (Gulisija and Crow, 2007). They can be computed in simple pedigrees for an individual *i* as:

$$O_{(i)} = \sum_j \sum_k (1/2)^{n-1} F_j$$

$$O_{e(i)} = \sum_j 2F_{i(j)} F_j$$

where the summation *j* is over all inbred ancestors, and the summation *k* is over all paths from *i* to ancestor *j* (each involving a number of individuals *n*, *i* and *j* included). $F_{i(j)}$ is the partial inbreeding of *i* referred to ancestor *j*, indicating *i*'s probability to be inbred by descent for alleles derived from *j*. *O* and O_e are computed via the function `purgeR::ip_op()`, and notes on the validation for this function and others in the package are provided in Supplementary File S3.

For complex pedigrees involving several autozygotes in the same path, these parameters need to be corrected by discounting from close ancestors' contributions the contribution made by far ancestors (Gulisija and Crow, 2007). However, equations for complex pedigrees derived by Gulisija and Crow involve highly nested complex loops and recursivity, and are not scalable. To make this method more usable, a heuristic approach was developed to correct *O* and O_e measurements. This approach skips far ancestors, that contribute little to *O* and O_e, and also to its correction terms. Specifically, given an individual *i* of interest, contributions to O_i and $O_{e(i)}$ from far ancestors *k* are ignored as long as $F_{j(k)} > 0$, where *j* is an intermediate ancestor. As a drawback, this method can result in estimates of $O > 1$ and $O_e > F$, particularly in cases involving many recent ancestors such that $F_{j(k)} = 0$ (e.g. from consecutive generations) undergoing selfing or breeding very close relatives. These situations are unexpected under the original model, and in these cases $O = 1$ and $O_e = F$ should be considered.

Figure 1A shows the observed decline of *B* in the bottlenecked population, together with expectations based on normalized O_e estimated from the pedigree (i.e. O_e/F), and also based on purged

inbreeding. Figure 1B shows the substantial variation in O_e for individuals with the same level of inbreeding. It also shows the problem of obtaining $O_e > F$ estimates. However, as illustrated in Figure 1A, assuming $O_e = F$ in such cases allows to estimate *B* reasonably well. Results for alternative mutational models, and an example on a real population, are given in Supplementary File S3, and show that using both corrected and uncorrected O_e estimates might be useful setting, respectively, lower and upper bounds to *B* decline.

It must be noted that Guliisja and Crow's (2007) model assumes highly deleterious and recessive mutations, thus relating to the most deleterious component of the inbreeding load, and ignoring the possible purging of variants with low effect sizes. Evidence from genomic studies however suggest that purging is only relevant for highly deleterious variants (Grossen *et al.*, 2020). In small populations affected by drift, purging is also expected to be efficient only for strongly deleterious alleles (Glémin, 2003). Therefore, assumptions of the model should hold in most practical cases. In addition, estimation of inbreeding load decline using *O* and O_e provide a way to estimate inbreeding load decline that is not dependent on accurate measures of fitness and other factors, which can be troublesome or be incomplete in many real scenarios.

4 Concluding remarks

purgeR is a versatile toolkit to measure inbreeding and purging parameters in pedigreed populations. The inclusion of opportunity of purging parameters is a valuable contribution to the field, since it allows for the inference of purging without requiring more information than the pedigree structure.

References

Ballou,J.D. (1997) Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.*, **88**, 169–178.

Baumung,R. *et al.* (2015) GRAIN: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *J. Anim. Breed. Genet.*, **132**, 100–108.

Boakes,E.H. and Wang,J. (2005) A simulation study on detecting purging of inbreeding depression in captive populations. *Genet. Res.*, **86**, 139–148.

Charlesworth,B. (2018) Mutational load, inbreeding depression and heterosis in subdivided populations. *Mol. Ecol.*, **27**, 4991–5003.

Charlesworth,D. and Charlesworth,B. (1987) Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.*, **18**, 237–268.

Doekes,H.P. *et al.* (2020) Revised calculation of Kalinowski's ancestral and new inbreeding coefficients. *Diversity*, **12**, 155.

García-Dorado,A. (2012) Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics*, **190**, 1461–1476.

García-Dorado,A. *et al.* (2016) Predictive model and software for inbreeding-purging analysis of pedigreed populations. *G3*, **6**, 3593–3601.

Glémin,S. (2003) How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution*, **57**, 2678–2687.

Grossen,C. *et al.* (2020) Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat. Commun.*, **11**, 1001.

Gulisija,D. and Crow,J.F. (2007) Inferring purging from pedigree data. *Evolution*, **61**, 1043–1051.

Haller,B.C. and Messer,P.W. (2019) SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Mol. Biol. Evol.*, **36**, 632–637.

Hedrick,P.W. and García-Dorado,A. (2016) Understanding inbreeding depression, purging, and genetic rescue. *Trends Ecol. Evol.*, **31**, 1109–1144.