Accounting for cryptic population substructure enhances detection of inbreeding depression with genomic inbreeding coefficients: an example from a critically endangered marsupial

Joseph P Zilko[1*], Dan Harley[2], Birgita Hansen[3], Alexandra Pavlova[1], Paul Sunnucks[1]

[1]School of Biological Sciences, Monash University, Clayton VIC, Australia 3800

[2]Zoos Victoria, Parkville VIC, Australia 3052

[3]Centre for eResearch and Digital Innovation, Federation University, Ballarat VIC, Australia 3350

*Corresponding author: Joseph P Zilko (joseph.zilko@monash.edu)

Keywords: population genetics, genetic management, conservation, lethal equivalents, GLMs

Abstract

Characterising inbreeding depression in wildlife populations can be critical to their conservation. Coefficients of individual inbreeding can be estimated from genome-wide marker data. The degree to which sensitivity of inbreeding coefficients to population genetic substructure alters estimates of inbreeding depression in wild populations is not well understood. Using generalised linear models, we tested the power of two frequently used inbreeding coefficients that are calculated from genome-wide SNP markers, F_H and $F\text{^}^{III}$, to predict four fitness traits estimated over two decades in an isolated population of the critically endangered Leadbeater's possum. F_H estimates inbreeding as excess observed homozygotes relative to equilibrium expectations, whereas $F\text{^}^{III}$ quantifies allelic similarity between the gametes that formed an individual, and upweights rare homozygotes. We estimated F_H and $F\text{^}^{III}$ from 1,575 genome-wide SNP loci in individuals with fitness trait data (N=179-237 per trait), and computed revised coefficients, $F_H^{\text{by group}}$ and $F\text{^}^{III\text{by group}}$, adjusted for population

genetic substructure by calculating them separately within two different genetic groups of individuals identified in the population. Using F_H or $F^{\wedge III}$ in the models, inbreeding depression was detected for survival to sexual maturity, longevity, and whether individuals bred during their lifetime. $F^{\wedge III by\ group}$ (but not $F_H^{by\ group}$) additionally revealed significant inbreeding depression for lifetime reproductive output (total offspring assigned to each individual). Estimates of numbers of lethal equivalents indicated substantial inbreeding load, but differing between inbreeding estimators. Inbreeding depression, declining population size, low and declining genetic diversity suggests that genetic rescue may assist in preventing extinction of this unique Leadbeater's possum population.

Introduction

Many threatened species are reduced to small and isolated populations, making them highly susceptible to genetic erosion from genetic drift and an increased likelihood of reproduction among closely-related individuals (Frankham et al., 2017). In diploid species that normally outbreed, inbreeding by one or both of those mechanisms is almost always accompanied by reduced fitness (i.e. inbreeding depression; Hedrick & Garcia-Dorado, 2016). Inbreeding leads to increased homozygosity, which lowers fitness through two main hypothesized mechanisms: homozygosity for partially recessive harmful alleles, and for alleles at loci showing heterozygote advantage (overdominance; Charlesworth & Willis, 2009). Small, isolated populations with inhibited gene flow with other populations additionally lose alleles that contribute to phenotypic diversity, reducing their potential to evolve adaptively to environmental change (Bell & Gonzalez, 2009; Dierks, Baumann, & Fischer, 2012; Hufbauer et al., 2015). If inbreeding in small, isolated populations is not counteracted by gene flow, their extinction becomes increasingly likely due to positive feedback between small population size and reduced fitness (Fagan & Holmes, 2006).

Estimates of inbreeding depression are useful for modelling population viability and predicting the response of populations to different genetic management scenarios (Harrisson et al., 2016; Pierson et al., 2015). The magnitude of potential inbreeding depression in a population may be estimated as the mean number of lethal equivalents, where one lethal equivalent corresponds to a set of deleterious alleles that cause death when homozygous in an individual (Morton, Crow, & Muller, 1956). Determination of the number of lethal equivalents requires estimates of fitness and individual inbreeding (Hedrick & Kalinowski, 2000). Levels of individual inbreeding can now be estimated effectively in wildlife populations using genome-wide genetic markers such as single nucleotide polymorphisms (SNPs; Huisman, Kruuk, Ellis, Clutton-Brock, & Pemberton, 2016; Kardos, Luikart, & Allendorf, 2015; Pemberton, Ellis, Pilkington, & Bérenos, 2017). The proportion of loci that are homozygous in an individual is the simplest such genomic indicator of inbreeding, and may be correlated with fitness estimates, allowing inference of inbreeding depression (Harrisson et al., 2019; Hoffman et al., 2014). Correlations between homozygosity and fitness are attributed to genomic markers being within or close to loci influencing fitness, and/or correlation between average homozygosity across multiple marker loci and loci that affect fitness (Miller & Coltman, 2014). However, alternative inbreeding estimators may more accurately reflect homozygosity at loci associated with fitness. Accordingly, considerable attention has been applied to designing and testing different genomic inbreeding coefficients that emphasize various aspects of homozygosity (Keller, Visscher, & Goddard, 2011; Yang, Lee, Goddard, & Visscher, 2011).

The individual inbreeding coefficient, F, is the probability that two alleles at a locus chosen at random in an individual are identical by descent (autozygous) with respect to a reference population in which all alleles are independent (Keller et al., 2011). One relatively straightforward genomic inbreeding coefficient, F_{H}, estimates inbreeding as the excess

number of observed homozygotes in an individual relative to the number expected under Hardy-Weinberg equilibrium given the allele frequencies in the population sample in which inbreeding was estimated (Purcell et al., 2007). F_{H} does not specifically weight homozygotes differently according to their frequency. However, the harmful alleles thought to be predominantly responsible for inbreeding depression are typically expected to be present in populations at low frequencies owing to mutation–selection balance (Charlesworth & Willis, 2009). Homozygosity at such loci results from related parents transmitting identical genome segments to offspring (Keller et al., 2011). Some inbreeding estimators, for example $\hat{F}^{\mathrm{III}}$, have been designed to capture allelic similarity between parental gametes that formed an individual and upweight rare homozygotes (Yang et al., 2011; written as $F^{\wedge\mathrm{III}}$ herein). Because these various inbreeding coefficients reflect somewhat different properties of genomes, and because drivers of inbreeding depression may vary among populations, different coefficients can be expected to differ in their strength of correlation with homozygosity at fitness-relevant loci and thus, in their power to detect inbreeding depression in a given population (Bérénos, Ellis, Pilkington & Pemberton, 2016). Accordingly, there is considerable value in comparing multiple estimators of inbreeding such as F_{H} and $F^{\wedge\mathrm{III}}$, as Huisman et al. (2016) did using empirical data from a wild mammal, and Nietlisbach, Muff, Reid, Whitlock, & Keller (2019) did via simulations.

Genomic inbreeding coefficients are calculated relative to reference allele frequencies, which can distort inbreeding estimates for recent immigrants and their descendants because of their genetic dissimilarity with the reference population. For example, $F^{\wedge\mathrm{III}}$ of an immigrant is inflated by possession of homozygotes that are more common in the immigrants' source population compared to the focal population (Huisman et al., 2016; Nietlisbach et al., 2019). If immigrants with distorted inbreeding estimates are rare, they may have little effect on detection of inbreeding depression. However, where population substructure occurs,

including unexpected genetic groups that might go undetected, similar distortions may have sufficient impact to alter estimates of inbreeding depression substantially. Such possibilities have not been thoroughly explored to date. Given this knowledge gap, empirical research into the utility of SNP-based estimators of inbreeding coefficients as predictors of inbreeding depression in structured populations is warranted.

Inbreeding can impact fitness to varying degrees at different life stages, and these influences can accumulate across the lifespan (Grueber, Laws, Nakagawa, & Jamieson, 2010; Huisman et al., 2016). This makes it beneficial to examine survival rates across different life history stages (Grueber et al., 2010). Estimating inbreeding depression for adult traits can be particularly challenging because selection exerted during juvenile life stages may elevate mortality for individuals with higher inbreeding, thus reducing variation in fitness and inbreeding at later life stages (Judson, Knapp, & Welch, 2018). Thus, estimating inbreeding depression in wild populations is best achieved by monitoring survival and reproductive output over lifespans (Frankham et al., 2017).

We investigated inbreeding depression and the effect of population substructure on its estimation in Leadbeater's possum (*Gymnobelideus leadbeateri*), an arboreal marsupial that is restricted to a 70 x 95 km range within Victoria, Australia. Drastic reductions in population size due to recent habitat loss following a major bushfire in 2009 has resulted in this species being listed as Critically Endangered by the IUCN (Woinarski & Burbidge, 2016). Based on bone deposits accumulated by predators, Leadbeater's possums are inferred to have experienced a major decline in abundance and range coincident with European human settlement of SE Australia ~150 years ago (Bilney, Cooke, & White, 2010). The possums were common and widespread prey of owls ~300-400 years ago, being found in subfossil owl

126 pellets at 7 of 10 sites along a 115 km transect east of their current distribution, yet absent
127 from 2,754 contemporary owl pellets at 17 sites in that region (Bilney et al., 2010). Within
128 the species' core range, population monitoring data indicate declines of at least 50% since
129 1996 (Woinarski & Burbidge, 2016).
130
131 Leadbeater's possum populations currently occupy two disjunct ranges (Figure 1):
132 approximately 200,000 Ha of high elevation (600-1500 m asl) montane forests and subalpine
133 woodland within the Central Highlands of Victoria and one remnant, outlying low elevation
134 (110m asl) population that is now confined to ~ 20 Ha of floodplain forest within the 680 Ha
135 Yellingbo Nature Conservation Reserve (Harley, Worley, & Harley, 2005). Before extensive
136 clearing for agriculture early last century (Woodgate and Black 1988), this lowland habitat
137 covered at least 10,000 Ha within Leadbeater's possum's historic range (Yugovic & Mitchell,
138 2006). The lowland Yellingbo population is located ~16 km from the nearest Central
139 Highland population, with suitable intervening habitat being cleared for agriculture in the
140 early 1900s. Consistent with its outlying geographic location in distinct, low-lying habitat,
141 the Yellingbo population is genetically differentiated from the higher-elevation populations,
142 and clusters with historic low elevation populations south of the present range (Hansen &
143 Taylor, 2008; Hansen, Harley, Lindenmayer, & Taylor, 2009). Given typical dispersal
144 distances of ≤1.5 km (Harley 2005), contemporary gene flow between lowland and highland
145 populations is extremely unlikely (Hansen & Taylor, 2008).
146
147 The Yellingbo population, the focus of this study, showed low microsatellite and
148 mitochondrial sequence diversity compared to all of three sampled highland populations
149 including Lake Mountain, which we include here for comparison (Hansen et al., 2009).
150 Yellingbo also shows population substructure comprising two adjacent but geographically

localized genetic groups distributed along different creeks in the reserve (Hansen et al., 2009; Sunnucks & Hansen, 2013). In addition to inbreeding, deteriorating habitat condition has contributed to a decline in population size from 112 in 2003 to only 40 in 2014 (Woinarski & Burbidge, 2016).

In this study, we used genomic data and fitness monitoring information spanning two decades to test for inbreeding depression and estimate the inbreeding load on traits including lifetime reproductive output in the Yellingbo population of Leadbeater's possum. We investigated whether previously-described local genetic substructure within this population could hinder the ability to detect inbreeding depression. Because the maximum observed lifespan at Yellingbo was 11 years (Supporting Data 2; Supporting Figure S14), the period of monitoring was sufficient to yield information on lifetime reproductive output of many individuals, and life history data were obtained for a large majority (>70%) of the population each year.

Methods

Population monitoring and genetic sampling

The entire area occupied by the lowland population of Leadbeater's possum at Yellingbo (Figures 1 and 2) was surveyed for presence of the species between 1994 and 1996, and population size and demography has been monitored intensively from 1996 to present in all years except 2000. Monitoring was undertaken predominantly by mark-recapture of individuals denning in nest boxes. Nest boxes are readily occupied by the possums, with up to 75% of the population regularly denning in nest boxes during the day, which facilitates sampling of possums (Harley, 2006). Groups of possums den together during the day in small colonies typically of 3-4 individuals, usually comprising a dominant breeding pair and their

offspring (Harley, 2005). Colonies use multiple denning sites and move frequently between nest boxes within their territory, with some colonies also occasionally denning in natural tree hollows (Harley, 2004, 2005). Monitoring effort was consistent across the study area throughout this period. In instances where established colonies from previous years were not detected denning in nest boxes, secondary techniques to confirm ongoing site occupancy were employed including additional nest box inspections, trapping, observing tree hollows at dusk to survey for occupancy, call imitation, and/or camera trapping (Harley, 2015; Harley, Holland, Hradsky, & Antrobus, 2014; Smith et al., 1989). Individuals captured from nest boxes were marked by tattooing the pinnae of the ears with a unique letter-number identifier. Weight measurements were obtained for each individual using a spring scale (±1 g) following capture. Sex was determined by visual examination of external genitalia and presence or absence of a pouch (Harley & Lill, 2007). Adults were inspected for signs of reproductive activity and pair-bonding (see below).

Genetic samples (n=287) were collected from individuals at Yellingbo by taking small ear biopsies of skin and cartilage (~2 mm diameter) from the edge of pinnae. These were stored in 95% ethanol prior to DNA extraction. Genetic sampling occurred between 1997-2001 (n=157) and 2010-2017 (n=130). We also genotyped Lake Mountain highland individuals (n=113; Figure 1) sampled in 2006-2007 (Hansen et al., 2009) for comparison of heterozygosity with Yellingbo. All population monitoring and tissue sample collection procedures were approved by the Zoos Victoria Animal Ethics Committee.

DNA extraction, genotyping and data filtering

For 130 Yellingbo samples, total genomic DNA was extracted using Qiagen DNeasy Blood and Tissue Kits; for the remaining 157 Yellingbo and 113 Lake Mountain samples, DNA had

been extracted using a salting-out protocol for a previous study (Hansen & Taylor, 2008). Codominant, genome-wide, biallelic single nucleotide polymorphism markers were generated using DArTseq methodology by a commercial provider, Diversity Arrays Technology Pty. Ltd. DArTseq incorporates complexity-reduction methods with combinations of restriction enzymes that target low-copy genomic regions to detect large numbers of informative SNPs within 69 bp sequences (Kilian et al., 2012). Four complexity-reduction methods comprising different combinations of restriction enzymes were tested to optimise representation and genome coverage. The restriction enzyme combination *Pst*I-*Sph*I was applied for genotyping. DNA samples were processed in digestion/ligation reactions as detailed previously (Kilian et al., 2012), replacing the *Pst*I-compatible adaptor with two adaptors corresponding to the two restriction enzyme overhangs. The *PstI*-compatible adaptor was designed with a flow cell adaptor sequence, a sequencing primer region, and a staggered barcode region of varying length. The *SphI*-compatible adaptor comprised a restriction enzyme overhang sequence and a flow-cell binding site. Fragments were amplified in 30 rounds of PCR with an initial 94 °C denaturation step for 1 min, 30 denaturation steps at 94° C lasting 20 s, 30 s annealing at 58 °C, 45 s extension at 72 °C, and a final extension step at 72 °C. Equimolar amounts of PCR products were applied to an Illumina c-bot for bridge amplification. An Illumina Hiseq 2500 was used to generate single-read sequences with 77 cycles.

The resulting sequences were filtered using the DArT proprietary primary pipeline to remove poor quality sequences, whilst applying more stringent filtering criteria to the barcode region compared to the rest of the sequence, using a minimum Phred score of 30 and minimum pass percentage of 75. This higher stringency allowed for reliable assignment of sequences to the same sample. Identical sequences were collapsed into tags, which served as templates against which low quality bases in singleton tags were corrected. The DArT proprietary pipeline was

then used for SNP calling. This pipeline clusters sequences then parses clusters into separate SNP loci using a range of technical parameters, balancing read counts per allelic pair. Scoring consistency of alleles (repeatability) was determined using technical replicates for 25% of samples.

Starting with the total dataset yielded by the process above, loci were filtered using the R (R Core Team, 2017) package DARTR (Gruber, Unmack, Berry, & Georges, 2018). Sex-linked loci were first identified and removed. X-linked loci were identified as not heterozygous in males but heterozygous in >10% females, allowing for a 2% genotyping error rate using the function *gl.sexlinkage* with t.hom=0.02 and t.het=0.02. The dataset was also screened for XY gametologs always heterozygous in males and always homozygous in females (again with provision for 2% error), but none were detected. From the resulting set of autosomally-behaving loci, those with <100% average reproducibility and <90% call rate were removed. Where multiple SNPs were in the same sequence fragment, only the SNP with the highest repeatability and polymorphism information content was retained, to reduce very close linkage among markers. This filtering resulted in 3,378 high-quality SNP loci, with 1.3% missing genotypes, used for comparing genetic diversity between the representative highland Lake Mountain population and the isolated lowland Yellingbo population. Of these loci, 1,575 that were polymorphic within Yellingbo and had only 1.0% of genotypes missing were used to estimate individual inbreeding coefficients. For parentage assignment and analyses of population genetic structure at Yellingbo, a polymorphic genotype data set with few missing genotypes was created by removing loci with <99% call rate, leaving 1,207 loci with only 0.15% of genotype scores missing.

Genetic diversity and population substructure within Yellingbo

Change in population genetic diversity per year was estimated using genotyped individuals known to be alive each year from presence-absence data collected during population monitoring. Mean observed heterozygosity was estimated by averaging the proportion of loci heterozygous within individuals, PHt, calculated using the R package GENHET (Coulon, 2010). To compare genetic diversity in 287 Yellingbo individuals to that of 113 individuals in the larger, more outbred Lake Mountain population, mean heterozygosity at 3,378 loci (above) was calculated. Statistical significance of changes in mean per-locus heterozygosity at these loci within Yellingbo between 1997 and 2017, between genetic groups at Yellingbo (defined below), and between Yellingbo and Lake Mountain was determined using Wilcoxon signed rank tests (with locus as the pairing factor) implemented in the R package STATS (R Core Team, 2017). To compare inbreeding coefficients between genetic groups at Yellingbo, Welch's t-tests in STATS were used, because of unequal variance between groups.

Population substructure within Yellingbo was examined using 1,207 loci in STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000). We ran three analyses on Yellingbo data: one for samples collected from 1997- 2001, one for 2011- 2017, and all combined. Each analysis involved a 10,000 burn-in and 30,000 MCMC iterations for number of genotypic clusters (K) ranging from 1 to 5, with 20 replicate runs for each value of K and a random seed for each run. Admixture was enabled in the model and allele frequencies were assumed to be correlated across clusters, because occasional dispersal events would be likely across the small geographic area occupied by Yellingbo possums. The admixture parameter (α) was inferred from the data, with a uniform prior distribution between 1 and 10 and the frequency of the Metropolis-Hastings step set to 10. The default allele frequency prior (λ) of 1 was specified. Divergence levels (F_k) were also allowed to vary between clusters, using the default priors. Markers were assumed unlinked. The presence of genetic structure and

number of STRUCTURE clusters during each sampling period and for both periods combined was assessed following the recommendations of Janes et al. (2017), applying and comparing the LnPr(X|K) and ΔK methods (Evanno, Regnaut, & Goudet 2005). Replicate runs were summarized using CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) to identify highly similar independent runs at each value of K tested, group runs that produced distinct clustering solutions into modes, and calculate likelihood values for each clustering mode, as part of the Janes et al. (2017) recommendations. A hierarchical analysis was performed by re-running STRUCTURE on separate groupings of individuals that represented the uppermost level of structure in the population (K=2). The level of differentiation between groups of individuals defined by Q>0.5 in STRUCTURE was estimated using F_{ST} calculated in the R package STAMPP (Pembleton, Cogan, & Forster, 2013) using 10,000 bootstrap iterations.

PCA was used to verify the presence of population genetic substructure and was performed using the R package adegenet (Jombart, 2008). Allele frequencies were first scaled using the *scaleGen* function, replacing missing allele frequencies (NAs) with the mean allele frequency in the dataset. The *dudi.pca* function was then used, without further scaling of data or centering. As for STRUCTURE, separate PCAs were run for each sampling period and for both periods combined.

Fitness traits and parentage analysis

Four fitness traits were estimated: survival to sexual maturity, breeding at least once in a lifetime, longevity, and lifetime reproductive output. Longevity was estimated for all detected individuals. Lifetime reproductive output incorporated measures of survival, lifespan and reproduction and included all known individuals regardless of whether they reached sexual

maturity or bred. Survival to sexual maturity was estimated by determining whether animals survived to three years of age: reproductive maturity occurs at approximately two years of age (Harley & Lill, 2007), to which we added another year to account for sampling being annual. The age of each individual (month and year of birth) was estimated based on weight at first capture. As the growth rate of individuals until adulthood has been previously characterised, birth month could be estimated for juveniles under 12 months (weighing < 100 g) and sub-adults between 12-16 months old (100-120 g; Harley & Lill, 2007). Animals first captured as adults (typically weighing ≥ 120 g) were assumed to be two years old (Harley & Lill, 2007). The assumption that these individuals were no more than two years of age was justified by determining how frequently animals were undetected during surveys then later confirmed to be alive. For 258 presence/absence observations of 89 individuals that lived three or more years, there were only 36 occasions (14% observations) where animals were undetected during the annual monitoring but subsequently confirmed to be alive.

Lifetime reproductive output was estimated by determining the total number of offspring that could be assigned to each genotyped individual. To detect parent-offspring relationships within Yellingbo, CERVUS v3.0 (Marshall, Slate, Kruuk, & Pemberton, 1998) was applied using the dataset of 1,207 high-quality SNPs polymorphic within Yellingbo (above). CERVUS assigns parents to offspring on the basis of likelihood, permitting mistyping of loci, and calculates confidence levels of assignments by simulation of genotypes. Simulation of 100,000 offspring genotypes was undertaken prior to assignment of parent pairs of known sex. Twenty candidate parents of each sex were simulated; simulation settings were a 90% detection rate, with 90% of individuals genotyped for 1,207 loci and 2% genotyping error rate. Confidence levels of parentage assignment were calculated using log-likelihood ratios, specifying relaxed and strict confidence of 90% and 95% respectively. Parentage analysis

was carried out separately for the two sampling periods, 1997-2001 (early) and 2011-2017 (late). In the early period, 70% of animals first detected as adults and 62% first detected as juveniles were genotyped. For the late period, 93% of individuals first detected as adults and 86% first detected as juveniles were genotyped. The lower genotyping rate in the early period and potential consequence of bias in reproductive output estimates was accounted for in models (below) by specifying sampling period as a predictor variable. All individuals for which genetic data were available were included as offspring. All genotyped individuals that had reached sexual maturity before a given offspring was born were included as candidate parents in CERVUS.

As no genetic sampling occurred between 2002 and 2010, animals that were born and died in this period were assigned to parents on the basis of observational criteria validated by cases for which genetic assignments could also be made (Supporting Data 1). This method was also used where candidate parents were not genotyped or where CERVUS could not assign parentage with at least 90% confidence. The adult female in the colony that was lactating and the pair-bonded adult male were assigned as parents of any dependent juveniles in the same nest box. Paternity was determined on the basis of the extent of tail-base fur staining around the paracloacal gland, which arises from mutual grooming between pair-bonded individuals (Harley & Lill, 2007). Where tail-base fur staining was detected on more than one male in a colony, paternity was assigned to the male with the most extensive staining in combination with age, weight, and previous breeding history. Where uncertainty remained following these assessments, paternity was left unassigned. Breeding pairs typically remain together for at least three years or until either mate dies (Harley & Lill, 2007). Mothers could readily be assigned to juveniles based on observations of pouch condition (i.e. presence of pouch young, protruding nipples for suckling young, or pouch entrance stretched and stained indicating

previous breeding), because only a single breeding female is present in each colony at any time (Harley & Lill, 2007). Validity of parentage assignment using the two methods was assessed by determining how often the same parent-offspring pairs were assigned by both methods.

Estimating individual inbreeding within Yellingbo

Two genomic estimators of individual inbreeding with different properties were calculated: F_{H} (called F_{PLINK} in Huisman et al., 2016, and F_{h} in Keller et al., 2011), and $F^{\wedge\mathrm{III}}$ ($\hat{F}^{\mathrm{III}}$ of Yang et al., 2011, also called F_{grm} in Huisman et al., 2016, and F_{alt} in Nietlisbach et al., 2019). An assembled genome was not available to apply estimators based on runs of homozygosity, which can be powerful for estimating inbreeding and detecting inbreeding depression (Keller et al., 2011; Nietlisbach et al., 2019). F_{H} estimates inbreeding as excess homozygosity over all genotyped loci in an individual relative to expected homozygosity under Hardy-Weinberg equilibrium (HWE), without weighting of individual loci (Keller et al., 2011). This was calculated using the *-het* command in PLINK (Purcell et al., 2007). F_{H} theoretically ranges from minus infinity to infinity, with negative values for individuals that have fewer homozygotes than expected, and will be centred near zero in random mating populations with allele frequencies taken from the current sample of individuals (Kardos et al., 2015; Nietlisbach et al., 2019). $F^{\wedge\mathrm{III}}$ has the same theoretical range, and estimates genomic similarity between the gametes that formed an individual's genome, relative to a randomly sampled panmictic population with the same allele frequencies as the sampled individuals, and was calculated with the software GCTA (Yang et al., 2011). $F^{\wedge\mathrm{III}}$ strongly upweights rare homozygotes by $(1\text{-}p_{\mathrm{i}})/p_{\mathrm{i}}$ for minor homozygotes, $p_{\mathrm{i}}/(1\text{-}p_{\mathrm{i}})$ for major homozygotes, and −1 for heterozygotes, where p_{i} is the minor allele frequency (Huisman et al., 2016), with individuals less inbred than average having negative values, and individuals

more inbred than the average having positive values (Yang et al., 2011). Loci were not filtered for linkage disequilibrium, because it had negligible effects on F_H and $F\text{^}^{III}$ values (Supporting Figure S1, following Huisman et al., 2016).

Earlier work led us to expect two geographically localized genetic groups of Leadbeater's possums within Yellingbo (Hansen et al., 2009; Sunnucks & Hansen, 2013). Such genetic groups within the population would elevate homozygosity of some individuals relative to a single randomly mating population, yet this 'inbreeding by population subdivision' need not necessarily have fitness consequences (Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016; Keller & Waller, 2002). Further, members of a group with higher mean homozygosity would tend to have higher homozygosity because of genetic affiliation to that group, irrespective of their immediate ancestors' relatedness and this elevated homozygosity may or may not have fitness consequences. We considered all Yellingbo samples together, effectively ignoring spatial population genetic substructure. We then reduced the effects of population substructure on estimates F_H and $F\text{^}^{III}$ by calculating them separately within each of two genetic groups of individuals defined by shared majority membership (Q>0.5) to one of two spatially coherent genetic clusters identified in STRUCTURE and supported by PCA. For brevity, we refer to these adjusted inbreeding coefficients as $F_H^{\text{by group}}$ and $F\text{^}^{\text{IIIby group}}$. F_H is calculated as (O-E)/(L-E) where O is the number of homozygous loci observed in an individual, E is the expected number of homozygous loci under HWE, and L is the number of loci typed (Keller et al., 2011). The intent of $F_H^{\text{by group}}$ was to provide allele frequency estimates (and thus E) from the group to which an individual belongs, and hence reduce the contribution of inbreeding by population subdivision to an individual's inbreeding coefficient. Similarly, $F\text{^}^{\text{IIIby group}}$ was intended to supply reference allele frequencies that more closely represent the genetic group in which an individual's genome arose, since the

401 weighting scheme (above) is dependent on these frequencies and by definition, such groups
402 will differ in frequencies of some rare alleles. To understand differences in performance
403 between coefficients at detecting inbreeding depression, we explored the patterns and
404 magnitude of the adjustments for population substructure to F_H and $F^{\wedge III}$ using Pearson r
405 correlations calculated in the STATS package in R.
406
407 To determine whether there was sufficient variance in inbreeding in the population to detect
408 inbreeding depression, we calculated identity disequilibrium (g_2), for all samples pooled
409 together and separately for each genetic group, using the R package INBREEDR (Stoffel et
410 al., 2016) with 1,000 permutations and 1,000 bootstrap iterations. Identity disequilibrium
411 summarizes variance in identity by descent in a population, based on the covariance of
412 heterozygosity among loci within individuals. Meta-analysis indicates that higher values of g_2
413 are associated with larger effect sizes in general-effect heterozygosity fitness correlations,
414 and closer correlations between heterozygosity and inbreeding (Miller & Coltman, 2014).
415
416 *Modelling inbreeding depression*
417 Because inbreeding coefficients can differ in their power to detect inbreeding depression, it
418 was of interest to compare the performance of F_H (which does not upweight rare
419 homozygotes) with $F^{\wedge III}$ (which does) as predictors of inbreeding depression, and the impact
420 of population substructure on their predictive power. Values for the inbreeding coefficients
421 F_H, $F^{\wedge III}$, $F_H^{\text{by group}}$ and $F^{\wedge III\text{by group}}$ (defined above) were assessed for their relationship with
422 fitness traits using generalized linear models (GLMs). Relationships were also assessed for
423 each of these predictors calculated separately in each sampling period and heterozygosity
424 (PHt) as a predictor in place of inbreeding coefficients. GLMs were fitted for the four fitness
425 traits, with inbreeding coefficients, genetic group (based on membership to two genotypic

426 clusters, above) and sampling period as predictors of fitness, thus accounting for
427 spatial/temporal heterogeneity in genetic variation and detectability of parent-offspring
428 relationships. Not all fitness variables could be scored for all individuals, for example
429 because some were detected after they reached reproductive maturity, or they were still alive
430 in the final sampling year. A binomial distribution with logit-link function was specified for
431 models of survival to sexual maturity and whether individuals bred during their lifetime. Data
432 for longevity and lifetime reproductive output were over-dispersed, so a negative binomial
433 distribution with log-link function was used for these fitness traits. To enable estimation of
434 relative strengths of predictors where models included interactions between predictors,
435 continuous predictors were standardised by subtracting the mean and dividing by two
436 standard deviations for each predictor (Gelman et al., 2009). Binary predictors were
437 standardised by subtracting the mean, without dividing by the standard deviation. Interactions
438 were specified between F_{H} or $F_{\mathrm{\Lambda III}}$ and the remaining predictors to determine whether the
439 relationship between inbreeding differed between genetic groups and sampling periods.
440
441 An information-theoretic approach with model-averaging was used to estimate relationships
442 between predictors and fitness traits, following the procedure of Grueber, Nakagawa, Laws,
443 & Jamieson (2011). This approach allows several plausible hypotheses (models) to be tested
444 simultaneously, with inferences based on weighted support from multiple models with
445 similar likelihoods. For each inbreeding coefficient, global models containing all predictors
446 and interactions between the inbreeding coefficient and each of the remaining predictors were
447 first fitted for fitness traits using the *glm* function in the R package STATS. A subset of
448 models comprising all possible combinations of predictor variables and interactions included
449 in each global model was generated using the *dredge* function in the R package MUMIN
450 (Barton, 2009). Goodness-of-fit of models in this subset was measured by maximum

likelihood, using Akaike Information Criterion values corrected for small sample size (AICc). Final parameter estimates of relationships between predictors and responses were obtained by averaging the models falling within two AICc of the best fit model (ΔAICc <2) within each subset, using MUMIN. When a predictor was not included in one of these top models, its value was set to zero for that model. Predicted values for each fitness trait were generated for each individual in the dataset from each of the ΔAICc <2 top models and averaged. Correlation of predicted fitness responses against $F\text{^}^{\text{IIIby group}}$ was used to estimate percent change in fitness with increasing inbreeding, with values for fitness traits back-transformed from the link function-scale to the response-scale; this particular inbreeding coefficient was chosen for this purpose because it was the only one tested that significantly predicted all fitness indicators (see Results). The effect sizes of F_{H}, $F\text{^}^{\text{III}}$, $F_{\text{H}}^{\text{by group}}$ and $F\text{^}^{\text{IIIby group}}$ in predicting fitness were compared to assess their relative predictive power. These effect sizes were given by the slopes of averaged models, because all predictors were standardised to a common scale (Grueber et al., 2010).

GLMs with the sole fitness predictors of each of F_{H}, $F\text{^}^{\text{III}}$, $F_{\text{H}}^{\text{by group}}$ and $F\text{^}^{\text{IIIby group}}$ were used to estimate inbreeding load, represented by the number of haploid lethal equivalents. Unlike averaged models with multiple predictors, these single predictor models did not require standardization of variables, which would yield incorrect inbreeding load estimates. In order to obtain unbiased estimates of inbreeding load, we followed the method of Nietlisbach et al. (2019), which requires fitting GLMs with a log-link function, rather than a logit-link function normally used to model binary response variables. Accordingly, models to estimate numbers of lethal equivalents for fitness traits were fitted using a Poisson distribution and log-link function. Although this method yields reliable point estimates, it also produces large standard

errors for slope estimates, and so 95% confidence intervals were computed using parametric bootstrapping, refitting each model 100,000 times and resampling with replacement.

Results

Population substructure within Yellingbo

STRUCTURE and PCA analyses using 1,207 loci indicated the presence of two main genetic groups of Leadbeater's possums: each group was distributed mainly along one creek within the Yellingbo population, persisting from the early to the late period (i.e. from 1997-2001 to 2011-2017; Figure 2). In STRUCTURE analyses, the existence of only a single genotypic cluster was not supported: the likelihood of K (LnPr(X|K)) was markedly the lowest for K=1 in each time period and for all data pooled (Supporting Figure S2). Instead, the presence of two clusters was supported by the ΔK method (Supporting Figure S3), and by the procedures of Janes et al. (2017) as follows. With K =2, one of the two genotypic clusters was prevalent in individuals occurring along either northern or southern creek. With K>2, each of the two original clusters were subdivided further, but with only incremental support and patchy spatial distribution of animals sharing the same genotypic cluster (Supporting Figures S4-6). Exploration of likelihood values of multiple clustering modes identified for each value of K and conducting hierarchical analyses within each of the main clusters did not reveal clear alternative substructure to the two creek-associated clusters; K=2 showed a single mode with greater likelihood than for K=1, whereas multiple modes were observed for each K>2 (Supporting Table S1). The two genetic groups became somewhat less distinct in the later sampling period, coincident with a population decline in the northern group (Figure 2). Nonetheless, allele frequency differences between the two groups, as quantified by $F_{ST,}$ were large and significant in both periods (early: F_{ST}=0.19, 95% CI=0.17-0.20, p<0.001; late: F_{ST}=0.16, 95% CI=0.15-0.19, p<0.001). The genetic group to which each individual was assigned based on Q> 0.5 did not differ when assignments were undertaken for all samples

500 together or for each period separately (Supporting Figures S7-8). PCA divided individuals
501 into the same two genetic groups as did STRUCTURE-based assignment, in the total data
502 and in both sampling periods (Figure 3, Supporting Figure S9).
503
504 *Genetic diversity at Yellingbo*
505 Mean heterozygosity (PHt) of 3,378 loci at Yellingbo declined approximately monotonically
506 from 2005, and fell significantly by 12% from 0.083 ± SD 0.01 in 1997 to 0.073 ± SD 0.01 in
507 2017 (Wilcoxon signed rank test, V=413700, $p<0.01$; Figure 4). Mean PHt at Yellingbo
508 across all years assessed was 0.08 ± SD 0.01, which was significantly lower and
509 approximately half that of the outbred highland Lake Mountain population sample, calculated
510 using the same 3,378 loci (0.17 ± SD 0.01; Wilcoxon V=727400, $p<0.001$). Mean
511 heterozygosity was slightly but significantly lower in the southern genetic group than the
512 northern one (southern PHt=0.076, ± SD 0.01; northern PHt=0.080 ± SD 0.01; Wilcoxon
513 signed rank test, V = 689610, $p<0.01$).
514
515 *Outcomes of parentage assignment at Yellingbo*
516 Using a combination of genetic- and observation-based assignments, 95% of juveniles across
517 all sampling periods had mothers assigned and 91% had fathers assigned (Table 1). High
518 reliability of the observational parentage assignments based on pair-bonding was indicated by
519 CERVUS assigning the same parent in ~90% of cases where genetic data were available
520 (89% fathers, 90% mothers; Supporting Data 2). The proportion of parentages assigned
521 genetically rather than by observation was lowest for 2002-2010 (when genetic sampling was
522 not conducted), and highest for 1995-2001.
523
524 *Effects of population substructure on inbreeding coefficients within Yellingbo*

525 The adjustment of F_H to account for population substructure had modest impacts: the range of
526 $F_H^{by\ group}$ was only slightly lower than that of F_H and highly correlated with it (Supporting
527 Figure S10). This downward adjustment was very similar in the individuals within each
528 genetic group: by mathematical necessity, this effect was slightly greater for the more
529 homozygous southern group, and because the means of $F_H^{by\ group}$ must be close to zero within
530 each group, the adjustment eliminated the significant difference seen in F_H between groups
531 (Supporting Figure S10).
532
533 The adjustment of $F\wedge^{III}$ for population substructure had greater and more individual-based
534 effects than for F_H: $F\wedge^{IIIby\ group}$ was generally lower than $F\wedge^{III}$ and highly correlated with it, but
535 the adjustment was not universally downwards (Supporting Figure S11). Individuals with
536 high probability of membership to either genotypic cluster had lower $F\wedge^{IIIby\ group}$ than $F\wedge^{III}$,
537 whereas in more-admixed individuals, $F\wedge^{III}$ was the lower; the effects were very similar in
538 each group (Supporting Figure S11). Neither coefficient was significantly different in the two
539 population substructure groups (Supporting Figure S11). Frequency distributions of the four
540 F estimators are given in Supporting Figure S12, the relationships between F_H vs $F\wedge^{III}$, F_H^{by}
541 ^{group} vs $F\wedge^{IIIby\ group}$ are shown in Supporting Figure S13 with correlation coefficients in
542 Supporting Table S2).
543
544 Significant covariance of heterozygosity among loci within individuals (thus variance in
545 inbreeding) was detected in the population, as estimated by identity disequilibrium (g_2=0.022,
546 95% CI: 0.018 to 0.027, p<0.01). Identity disequilibrium calculated separately for the
547 northern group (g_2=0.014, 95% CI: 0.010 to 0.020, p<0.01) was lower than this overall value,
548 but higher for the southern group (g_2=0.03, 95% CI: 0.023 to 0.036, p<0.01). Thus, there was
549 likely sufficient variation in inbreeding to detect inbreeding depression if present.

Inbreeding depression detected using different genomic predictors

The model-averaged GLMs using F_H as the measure of inbreeding revealed significant inbreeding depression for three fitness traits: survival to sexual maturity, longevity, and whether individuals bred during their lifetime (Figure 5, Supporting Table S3, with additional details of models in Supporting Tables S4-11). Adjusting F_H for population substructure caused little change in inferences of inbreeding depression: $F_H^{\text{by group}}$ was not significantly associated with any more fitness traits and showed little or no increase in effect sizes compared to F_H (similar effect sizes for survival to sexual maturity and longevity, smaller for whether individuals bred, and larger for lifetime reproductive output; Figure 5, Supporting Table S3). In contrast, adjusting $F\text{^}^{\text{III}}$ for population substructure was associated with consistent increases in detection of inbreeding depression: effect sizes for $F\text{^}^{\text{IIIby group}}$ were larger than those of $F\text{^}^{\text{III}}$ for all four fitness indicators, and lifetime reproductive output became significant (Figure 5, Supporting Table S3).

Overall, $F\text{^}^{\text{III}}$ and $F\text{^}^{\text{IIIby group}}$ were more strongly correlated with fitness than were F_H and $F_H^{\text{by group}}$. Effect sizes of $F\text{^}^{\text{IIIby group}}$ were larger than those of F_H and $F_H^{\text{by group}}$ for all fitness indicators, and the same was true for $F\text{^}^{\text{III}}$ with the exception that $F\text{^}^{\text{III}}$ was not significantly correlated with lifetime reproductive output, having a weaker relationship with this fitness trait than did $F_H^{\text{by group}}$ (Figure 5, Supporting Table S3). Of the inbreeding estimators tested, only $F\text{^}^{\text{IIIby group}}$ was significantly associated with all four fitness traits.

Quantifying the strength of inbreeding depression

$F\text{^}^{\text{IIIby group}}$ was the only coefficient tested that was significantly associated with all fitness traits examined and showed the strongest effect sizes. It was thus used to estimate the strength of inbreeding depression. Correlation of $F\text{^}^{\text{IIIby group}}$ against averaged point-

predictions from fitted GLMs (Figure 6) indicated that for every 10% increase in this coefficient relative to the range of observed values, probability of survival to sexual maturity decreased by an average of 0.08, longevity by 0.03 years, the probability of breeding at least once during a lifetime by 0.09, and lifetime reproductive output by 0.26 offspring.

Regression of inbreeding coefficients against values of each fitness trait enabled estimation of number of haploid lethal equivalents (Table 2). For traits that showed significant inbreeding depression using all four coefficients, the largest estimated number of haploid lethal equivalents was obtained using $F\text{^}^{\text{IIIby group}}$ (means ranging from 8.06 to 11.46), followed by $F\text{^}^{\text{III}}$ (5.08 to 8.60). Estimates were substantially lower using F_{H} (1.17 to 1.35) and $F_{\text{H}}^{\text{by group}}$ yielded the lowest estimates of all, for all traits (1.02-1.28). Thus, although the adjustment of F_{H} and $F\text{^}^{\text{III}}$ for population substructure generally reduced individual inbreeding coefficients, the effect on estimation of lethal equivalents was opposite for the two metrics: estimates decreased when adjusting F_{H} to $F_{\text{H}}^{\text{by group}}$ but increased from $F\text{^}^{\text{III}}$ to $F\text{^}^{\text{IIIby group}}$.

Discussion

In this study, we revealed substantial inbreeding depression for a suite of important fitness traits in the last known lowland population of the critically endangered Leadbeater's possum. The combination of intensive population monitoring over decades and application of genome-wide SNP markers allowed us to estimate inbreeding depression in lifetime reproductive output in a free-living organism with a maximum observed lifespan of 11 years (Supporting Data 2; Supporting Figure S14). Such assessments are demanding and not commonly achieved (Huisman et al., 2016). Using approaches that enhance their comparability among studies (Nietlisbach et al., 2019) we estimated inbreeding load for lifetime reproductive output as haploid lethal equivalents. We also quantified fine-scale

599 within-population genetic substructure, and showed how accounting for its presence can
600 enhance the detection of inbreeding depression and substantially change estimates of
601 numbers of lethal equivalents. In our data, the inbreeding coefficient $F_{\wedge}^{III}$ was more powerful
602 at detecting inbreeding depression than F_H, and even more so after being adjusted for
603 population substructure.
604
605 *Substantial inbreeding depression in lowland Leadbeater's possums*
606 Relationships between inbreeding and fitness showed that the small, isolated lowland
607 population of Leadbeater's possum at Yellingbo experienced substantial inbreeding
608 depression for four fitness traits: survival to sexual maturity, longevity, whether individuals
609 bred during their lifetime, and lifetime reproductive output (Figures 5 and 6). Reduced fitness
610 is a near-universal consequence of inbreeding in small and isolated populations of diploid
611 species that usually outbreed (Hedrick & Garcia-Dorado, 2016). However, it was possible
612 that low variation in contemporary levels of inbreeding and fitness in the population could
613 have hindered the ability to detect inbreeding depression, and purging can sometimes
614 ameliorate inbreeding depression (Hedrick & Garcia-Dorado, 2016). Despite these possible
615 effects, more-inbred Yellingbo possums had measurably lower fitness. The reproductive
616 biology of Leadbeater's possum may make it prone to inbreeding. This species is socially
617 monogamous, with access to limited breeding vacancies controlled by a strong mating
618 hierarchy, leading to great variance in reproductive output that reduces effective population
619 size (Harley & Lill, 2007; Nomura, 2002). Whether individuals avoid inbreeding by choosing
620 less-inbred or less-related mates warrants investigation.
621
622 *The influence of population genetic substructure on estimating inbreeding and inbreeding*
623 *depression*

Our analyses examined the effect of two spatially-coherent genetic groups at Yellingbo on estimation of inbreeding and inbreeding depression. The occurrence of two genetic groups within a small geographical area (<150 Ha hectares of habitat) was unexpected when they were originally detected with a modestly powerful microsatellite assay (Hansen et al., 2009). Here, we clarified spatial genetic patterning using 1,207 high quality SNP. The population occurs along two adjoining creeks, currently totaling ~4 km of roughly linear flood plain forest. Our genetic data show some minor population genetic substructure along each creek, likely to reflect resistance to dispersal caused by territorial aggression, as commonly observed in territorial mammals (Supporting Figure S4; Armansin et al., 2020). Leadbeater's possums live in strongly territorial colonies that maintain exclusive, non-overlapping home ranges, with the distance between the dens of neighboring groups within intact habitat at Yellingbo being ~165 m (Harley, 2005). At Yellingbo, mean dispersal distance is 407 m (range 125 – 1080 m) for females and 95 m (range 105 – 1460 m) for males (Harley, 2005). Longer dispersal movements of up to 3.2 km have been observed during translocations of the species (D. Harley, unpubl. data). Thus, the two genotypic clusters fall within the dispersal capabilities of the species and genetic substructure cannot be explained solely by geographical distance. Most of the genetic differentiation in the population occurred between the two creeks and persisted across the 20-year study period, which represents more than three generations given the generation time of 6 years (Figure 2, Supporting Figures S2-9; Woinarski & Burbidge, 2006). The habitat at the confluence of the two creeks is of marginal quality, which may restrict dispersal and gene flow between the creeks relative to that along each creek. The allele frequencies characterizing the two genetic groups may be transitory over several generations. Nonetheless, accounting for this substructure altered our analyses of inbreeding and inbreeding depression. The kind of substructure we detected is common in

territorial animals (Armansin et al., 2020), hence our findings are likely to apply to other species.

In our study, adjusting estimates of inbreeding coefficients by calculating them separately for each genetic group within Yellingbo had different impacts on inferences of inbreeding depression derived from the two coefficients we tested. Owing to inbreeding by population subdivision, the adjustment from F_H to $F_H^{\text{by group}}$ slightly reduced the expected number of homozygotes (E) in a group that acted as the reference for individual observed homozygosity. This applied equally to all individuals in a group irrespective of their genotype, e.g. bearing alleles that are particularly common or rare in their group, and was slightly lower for the more homozygous southern group. The outcomes of the adjustment –slight reduction in estimated inbreeding of similar magnitude for all individuals in a group, and centering the values for each group around zero – did not uniformly increase effect sizes in models of inbreeding depression, nor did it predict additional fitness components significantly. Thus, the adjustment of F_H to $F_H^{\text{by group}}$ apparently did not usefully increase the correlation of inbreeding values with homozygosity at loci contributing to inbreeding depression.

Compared to F_H (and $F_H^{\text{by group}}$), $F\text{^}^{III}$ (and $F\text{^}^{III\text{by group}}$) was more predictive of fitness traits in terms of effect sizes and numbers of significant traits. $F\text{^}^{III}$ strongly upweights rare homozygous genotypes (Yang et al., 2011), and if these are more correlated with the loci that cause inbreeding depression than is multilocus homozygosity, this weighting should improve performance in estimating inbreeding depression. This is consistent with the apparently greater ability of $F\text{^}^{III}$ than F_H in our study to register inbreeding depression. In addition, this outcome accords with the lower error variance of $F\text{^}^{III}$ than F_H, notwithstanding different indicators performing better or worse in different studies due to technical considerations and

673 underlying biology (Bérénos et al., 2016; Keller et al., 2011). Our study was conducted in a
674 parameter space that is suboptimal for performance of F_H: simulations of populations
675 showing rapid decline in N_e comparable to Yellingbo and using similar numbers of SNPs
676 (~1,000) produced a correlation with genome-wide identity by descent of r^2 ~0.5, whereas
677 this increased to ~0.9 with 10,000 SNPs (Kardos et al., 2015). Furthermore, high
678 homozygosity (as observed in Yellingbo) causes imprecise F_H values when using only a few
679 thousand loci, and F_H also greatly overestimates the difference in inbreeding among
680 individuals with as few loci as we used (Kardos et al., 2015), which will be unhelpful in
681 estimating inbreeding depression. Kardos et al. (2015) did not make a direct comparison of
682 the extent to which these limitations apply to $F^{\wedge III}$.
683
684 The adjustment from $F^{\wedge III}$ to $F^{\wedge IIIby\ group}$ to more accurately capture the signal of rare
685 homozygotes possessed by an individual resulted in larger inbreeding depression effect sizes
686 for all traits, considerably higher estimates of numbers of lethal equivalents, and the key
687 additional inference of inbreeding depression on lifetime reproductive success. The latter was
688 unlikely to be a false positive, because this trait encompasses the other three fitness traits
689 examined, each of which was significantly associated with $F^{\wedge III}$ and F_H. The adjustment of
690 $F^{\wedge III}$ had individual-specific effects (unlike that of F_H) because of the emphasis of $F^{\wedge III}$ on rare
691 homozygous genotypes. This emphasis causes the previously described over-estimation of
692 $F^{\wedge III}$ of immigrants (Huisman et al., 2016; Nietlisbach et al. 2019). However, in our dataset
693 this over-estimation was greatest for individuals with a strong genetic affiliation with their
694 local genotypic cluster. Such individuals appeared somewhat differentiated relative to the
695 whole population, because it contained the other cluster (Supporting Figure S11). This effect
696 was removed by adjustment to $F^{\wedge IIIby\ group}$, for which the reference allele frequencies were
697 local. The considerable increase in inferred inbreeding depression through this correction

698 suggests that rare homozygotes are correlated with the loci causing inbreeding depression in
699 this population.
700
701 The apparently superior performance here of $F\text{^}^{III}$ over F_H need not apply to all studies. For
702 example, these two measures gave very similar estimates of inbreeding depression effect
703 sizes in red deer (*Cervus elaphus*) because there was very little contribution of rare
704 homozygotes to $F\text{^}^{III}$ (Huisman et al., 2016). In that study, PHt was highly correlated to $F\text{^}^{III}$
705 (r=0.94) compared to only 0.59 for $F\text{^}^{III}$ and 0.62 for $F\text{^}^{IIIby\ group}$ in our study (Supporting
706 Table S2).
707
708 *Biological effects of inbreeding depression and numbers of lethal equivalents*
709 Estimates of lethal equivalents are challenging to obtain in wild-living species, but are
710 important for modelling management actions to counteract inbreeding depression (Frankham
711 et al., 2017). Simulations suggest that F_H underestimates and $F\text{^}^{III}$ overestimates numbers of
712 lethal equivalents, with estimates based on pedigree inbreeding falling somewhere in-
713 between (Nietlisbach et al., 2019). In accordance with that assessment, in our data, ~5-10-
714 fold fewer lethal equivalents were inferred based on F_H (and $F_H^{by\ group}$) than by $F\text{^}^{III}$ (and
715 $F\text{^}^{IIIby\ group}$; Table 2). Our values were obtained following the recommendations of Nietlisbach
716 et al. (2019) for estimating comparable lethal equivalents, and using two of the commoner
717 estimators, F_H and $F\text{^}^{III}$ (although as suitable genomic data become more available,
718 estimators based on runs of homozygosity may be preferable at least under the right technical
719 and biological conditions; Kardos et al., 2015; Keller et al., 2011; Nietlisbach et al., 2019).
720 Therefore, we took $F_H^{by\ group}$ and $F\text{^}^{IIIby\ group}$ to be the most appropriate estimators in our data
721 because they account for known notable deviations from panmixia in the population. Using
722 these, the inbreeding load we estimated for survival to sexual maturity in Yellingbo

Leadbeater's possums lies somewhere in the order of 1.3 to 11.5 haploid lethal equivalents, which is potentially much higher than the mean of 2.3 reported for other mammals and 3.5 for vertebrates generally (Nietlisbach et al., 2019 based on pedigree inbreeding; Ralls, Ballou, & Templeton, 1988). Our only estimate for lifetime reproductive output (based on $F_{\wedge^{III}\text{by group}}$) of ~ 13.4 compares with 4.6 reported for the only other mammal population for which we could find a published estimate that accords with the Nietlisbach et al. (2019) criteria for comparability (banner-tailed kangaroo rats, *Dipodomys spectabilis*; Willoughby et al., 2019). Nietlisbach et al. (2019) report that with 1,000 simulated loci similar to what we used here, F_H underestimates inbreeding load quite considerably whereas $F_{\wedge^{III}}$ produces estimates more similar to the true value, in which case our relatively high estimates may be realistic. Irrespective of the absolute numbers of lethal equivalents, the Yellingbo Leadbeater's possum population has detectable inbreeding load with biologically non-trivial fitness consequences, suggesting that inbreeding depression strongly contributed to population decline over the past two decades.

Implications of this study for species management

Genetic rescue via assisted gene flow from an outbred or differentiated population is demonstrated to be a powerful tool for counteracting inbreeding depression (Frankham et al., 2017). Augmenting genetic diversity of the Yellingbo population from one or more highland populations offers potential fitness benefits to the Yellingbo population (detailed justification in Supporting Materials Appendix 1). Gene flow can reduce the frequencies of segregating harmful alleles responsible for the fitness differences among individuals in the population, ameliorate the effects of fixed harmful alleles in Yellingbo population relative to an outbred population, and reduce the number of individuals that are sufficiently inbred to experience fitness loss. In addition, increased genetic variation arising from gene flow should enhance

748 evolutionary potential (Harrisson et al., 2014). Application of risk-assessment approaches to
749 the conservation of Leadbeater's possum suggests that risk of extinction from inbreeding
750 depression currently outweighs known risks associated with gene flow (Appendix 1;
751 Frankham et al., 2017; Liddell, Cook, & Sunnucks, 2020; Ralls et al., 2018).
752
753 **Acknowledgements**
754 This research was funded by an ARC Linkage Grant LP160100482 with Partner
755 Organizations Victorian Department of Environment, Land, Water and Planning (DELWP),
756 Diversity Arrays Technology, Zoos Victoria, Australian Capital Territory Environment,
757 Planning & Sustainable Development Directorate, and Western Australian Department of
758 Biodiversity, Conservation and Attractions. Joanne Antrobus from Parks Victoria has
759 provided substantial support to the long-term monitoring program for Leadbeater's possum at
760 Yellingbo. BH, JZ, and DH were supported by Monash University internal funds and
761 Research and Training Program Scholarships (formerly Australian Postgraduate Award), and
762 awards from the Holsworth Wildlife Research Endowment. Computationally intensive
763 analyses were run with technical and infrastructure support from the Monash e-Research and
764 Bioinformatics Platforms. This research drew on materials from an Australian Academy of
765 Sciences Research Award for the Conservation of Endangered Australian Vertebrate Species
766 to Dr Andrea Taylor. We thank Dick Frankham, Bob Lacy and Kathy Ralls for helpful
767 discussions. We are grateful to Prof. David Coltman and at least four anonymous reviewers
768 for their comments on earlier versions of the manuscript that substantially improved it.
769
770 **Conflict of interest**
771 The authors have no conflict of interest to declare.
772

Author Contributions

JZ, PS, AP and DH designed the study. DH and BH performed the population monitoring and collection of genetic samples. JZ and BH performed DNA extractions. JZ collated genetic and population monitoring datasets and performed analyses. The manuscript was written by JZ, with contributions from all authors including reviewing the manuscript.

Data accessibility

Genotype and population monitoring data are publicly available on Bridges Monash University research repository (doi: 10.26180/5ef6ec4550f3b).

ORCID

Paul Sunnucks orcid.org/0000-0002-8139-7059

Alexandra Pavlova orcid.org/0000-0001-9455-4124

Joseph Zilko orcid.org/0000-0002-0732-2681

Birgita Hansen orcid.org/0000-0003-3608-7622

References

Armansin, N. C., Stow, A. J., Cantor, M., Leu, S. T., Klarevas-Irby, J. A., Chariton, A. A., & Farine, D. R. (2020) Social barriers in ecological landscapes: the social resistance hypothesis. *Trends in Ecology & Evolution*, **35**, 137-148.

Barton, K. (2009). MUMIN: multi-model inference, R package version 0.12. 0. Retrieved from http://r-forge/.*r-project.org/projects/mumin/*.

Bell, G., & Gonzalez, A. (2009). Evolutionary rescue can prevent extinction following environmental change. *Ecology Letters*, **12**, 942-948.

Bérénos, C., Ellis, P.A., Pilkington, J.G., & Pemberton, J.M. (2016). Genomic analysis reveals inbreeding depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular Ecology*, **25**, 3152-3168.

Bilney, R. J., Cooke, R., & White, J.G. (2010). Underestimated and severe: small mammal decline from the forests of south-eastern Australia since European settlement, as revealed by a top-order predator. *Biological Conservation*, **143**, 52-59.

Charlesworth, D., & Willis, J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783-796.

Coulon, A. (2010). genhet: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources*, **10**, 167-169.

Dierks, A., Baumann, B., & Fischer, K. (2012). Response to selection on cold tolerance is constrained by inbreeding. *Evolution*, **66**, 2384-2398.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Fagan, W. F., & Holmes, E. (2006). Quantifying the extinction vortex. *Ecology Letters*, **9**, 51-60.

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., . . . Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford, UK: Oxford University Press.

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., . . . Sunnucks, P. (2019) *A practical guide for genetic management of fragmented animal and plant populations*. Oxford, UK: Oxford University Press.

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., . . . Dorie, V. (2009). ARM: Data analysis using regression and multilevel/hierarchical models. R package version 9.01.Retrieved from https://CRAN.R-project.org/package=arm

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). DARTR: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, **18**, 691-699.

Grueber, C. E., Laws, R. J., Nakagawa, S., & Jamieson, I. G. (2010). Inbreeding depression accumulation across life-history stages of the endangered takahe. *Conservation Biology*, **24**, 1617-1625.

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, **24**, 699-711. doi:10.1111/j.1420-9101.2010.02210.x

Hansen, B. D., & Taylor, A. (2008). Isolated remnant or recent introduction? Estimating the provenance of Yellingbo Leadbeater's possums by genetic analysis and bottleneck simulation. *Molecular Ecology*, **17**, 4039-4052. doi:10.1111/j.1365-294X.2008.03900.x

Hansen, B. D., Harley, D. K. P., Lindenmayer, D. B., & Taylor, A. C. (2009). Population genetic analysis reveals a long-term decline of a threatened endemic Australian marsupial. *Molecular Ecology*, **18**, 3346-3362. doi:10.1111/j.1365-294X.2009.04269.x

Harley, D. (2015). The use of call imitation to establish territory occupancy by Leadbeater's possum (*Gymnobelideus leadbeateri*). *Australian Mammalogy*, **37**, 116-119. doi:10.1071/AM14014

Harley, D., Worley, M., & Harley, T. (2005). The distribution and abundance of Leadbeater's possum Gymnobelideus leadbeateri in lowland swamp forest at Yellingbo Nature Conservation Reserve. *Australian Mammalogy*, **27**, 7-15. doi:https://doi.org/10.1071/AM05007

Harley, D. K. P. (2004). Patterns of nest box use by Leadbeater's possum Gymnobelideus leadbeateri: Applications to research and conservation. In S. M. Jackson & R.L. Goldingay (Eds.), *The Biology of Australian Possums and Gliders* (pp. 12). Chipping Norton, NSW: Surrey Beatty & Sons.

Harley, D. K. P. (2005). *The life history and conservation of Leadbeater's possum (Gymnobelideus leadbeateri) in lowland swamp forest.* (Unpublished doctoral dissertation). Monash University, Melbourne.

Harley, D. K. P. (2006). A role for nest boxes in the conservation of Leadbeater's possum (Gymnobelideus leadbeateri). *Wildlife Research*, **33**, 385. doi:10.1071/WR04038

Harley, D.K.P., Holland, G., Hradsky, B., & S. Antrobus, J. (2014). The use of camera traps

to detect arboreal mammals: lessons from targeted surveys for the cryptic Leadbeater's possum (Gymnobelideus leadbeateri). In P.Meek, P. Fkeming, G. Ballard, P. Banks, A. Claridge, J. Sanderson & D. Swann (Eds.) *Camera Trapping: Wildlife Management and Research* (pp. 233-243). Collingwood, VIC: CSIRO Publishing.

Harley, D. K. P., & Lill, A. (2007). Reproduction in a population of the endangered Leadbeater's possum inhabiting lowland swamp forest. *Journal of Zoology*, **272**, 451-457. doi:10.1111/j.1469-7998.2007.00288.x

Harrisson, K.A., Magrath, M.J., Yen, J.D., Pavlova, A., Murray, N., Quin, B., Menkhorst, P., Miller, K.A., Cartwright, K., & Sunnucks, P. (2019). Lifetime fitness costs of inbreeding and being inbred in a critically endangered bird. *Current Biology*, **29**, 2711-2717.

Harrisson, K. A., Pavlova, A., Gonçalves da Silva, A., Rose, R., Bull, J. K., Lancaster, M. L., . . . Sunnucks, P. (2016). Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. *Molecular Ecology*, **25**, 1242-1258. doi:10.1111/mec.13547

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding Inbreeding Depression, Purging, and Genetic Rescue. *Trends in Ecology & Evolution*, **31**, 940-952. doi:https://doi.org/10.1016/j.tree.2016.09.005

Hedrick, P.W., & Kalinowski, S.T. (2000). Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics*, **31,** 139-162.

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., . . . Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, **111**, 3775-3780. doi:10.1073/pnas.1318945111

Hufbauer, R. A., Szűcs, M., Kasyon, E., Youngberg, C., Koontz, M. J., Richards, C., . . . Melbourne, B.A. (2015). Three types of rescue can avert extinction in a changing environment. *Proceedings of the National Academy of Sciences*, **112**, 10557-10562. doi:10.1073/pnas.1504732112

Huisman, J., Kruuk, L. E. B., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, **113**, 3585-3590. doi:10.1073/pnas.1518046113

Janes J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., & Andrew, R.L. (2017). The K = 2 conundrum. *Molecular Ecology*, **26**, 3594-3602.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405

Judson, J. L. M., Knapp, C. R., & Welch, M. E. (2018). Age-dependent, negative heterozygosity–fitness correlations and local effects in an endangered Caribbean reptile, Iguana delicatissima. *Ecology and Evolution*, **8**, 2088-2096.

Kardos, M., Luikart, G., & Allendorf, F.W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity*, **115**, 63-72

Kardos, M., Taylor, H.R., Ellegren, H., Luikart, G., & Allendorf, F.W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, **9**, 1205-1218.

Keller, L.F., & Waller, D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, **17**, 230-241.

Keller, C.K., Visscher, P.M., & Goddard, M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorpphism data. *Genetics*, **189**, 237-249.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., . . . Hopper, C. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. In F. Pompanon & A. Bonin (Eds.), *Methods in Molecular Biology* (pp. 67-89). New York, NY: Springer.

Kopelman, N.M., Mayzel. J., Jakobsson, M., Rosenberg, N.A., & Mayrose I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources,* **15**, 1179–1191. doi:10.1111/1755-0998.12387

Liddell, E., Cook, C.N., & Sunnucks, P. (2020) Evaluating the use of risk assessment frameworks in the identification of population units for biodiversity conservation. *Wildlife Research*, doi.org/10.1071/WR18170.

Marshall, T. C., Slate, J., Kruuk, L., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655. doi:10.1046/j.1365-294x.1998.00374.x

Miller, J.M. & Coltman, D.W. (2014). Assessment of identity disequilibrium and its relation to empirical heterozygosity-fitness correlations: a meta-analysis. *Molecular Ecology*, **23**, 1899-1090.

Morton, N.E., Crow, J.F. & Muller, H.J. (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences of the United States of America*, **42**, 855-863

Nietlisbach, P., Muff, S., Reid, J. M., Whitlock, M. C., & Keller, L. F. (2019). Nonequivalent lethal equivalents: Models and inbreeding metrics for unbiased estimation of inbreeding load. *Evolutionary Applications*, **12**. doi:doi:10.1111/eva.12713

Nomura, T. (2002). Effective size of populations with unequal sex ratio and variation in mating success. *Journal of Animal Breeding and Genetics*, **119**, 297-310. doi:10.1046/j.1439-0388.2002.00347.x

Pemberton, J. M., Ellis, P. E., Pilkington, J. G., & Bérénos, C. (2017). Inbreeding depression by environment interactions in a free-living mammal population. *Heredity*, **118**, 64-77. doi:10.1038/hdy.2016.100

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, **13**, 946-952.

Pierson, J. C., Beissinger, S. R., Bragg, J. G., Coates, D. J., Oostermeijer, J. G. B., Sunnucks, P., . . . Young, A. G. (2015). Incorporating evolutionary processes into population viability models. *Conservation Biology*, **29**, 755-764. doi:10.1111/cobi.12431

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., . . . Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**, 559-575

Ralls, K., Ballou, J. D., & Templeton, A. (1988). Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology*, **2**, 185-193.

Ralls K., Ballou J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., . . . Frankham, R. (2018). Call for a paradigm shift in the genetic management of fragmented populations. *Conservation Letters*, **11**, 1-6.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Smith, A., Lindenmayer, D., Begg, R., Macfarlane, M., Seebeck, J., &, Suckling, G. (1989). Evaluation of the stag-watching technique for census of possums and gliders in tall open forest. *Wildlife Research,* **16**, 575-580.

Stoffel, M.A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J.I. (2016). InbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, **7,** 1331-1339.

Sunnucks, P., & Hansen, B. (2013). Null alleles and Bonferroni 'abuse': Treasure your exceptions (and so get it right for Leadbeater's possum). In F. Allendorf, G. Luikart, & S. Aitken (Eds.), *Conservation and the Genetics of Populations* (2nd ed., pp. 93-95). Oxford, UK: Wiley-Blackwell.

Willoughby, J. R., Waser, P. M., Brüniche-Olsen, A., & Christie, M. R. (2019). Inbreeding load and inbreeding depression estimated from lifetime reproductive success in a small, dispersal-limited population. *Heredity,* **123**, 192-201. doi:10.1038/s41437-019-0197-z

Woinarski, J., & Burbidge, A.A. (2016). *Gymnobelideus leadbeateri*. *The IUCN Red List of Threatened Species*: e.T9564A21959976. http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T9564A21959976.en. Downloaded on 25 July 2019.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics,* **88**, 76-82.

Yugovic, J., & Mitchell, S. (2006). Ecological review of the Koo-Wee-Rup swamp and assocaited grasslands. *The Victorian Naturalist*, **123**, 223-234.

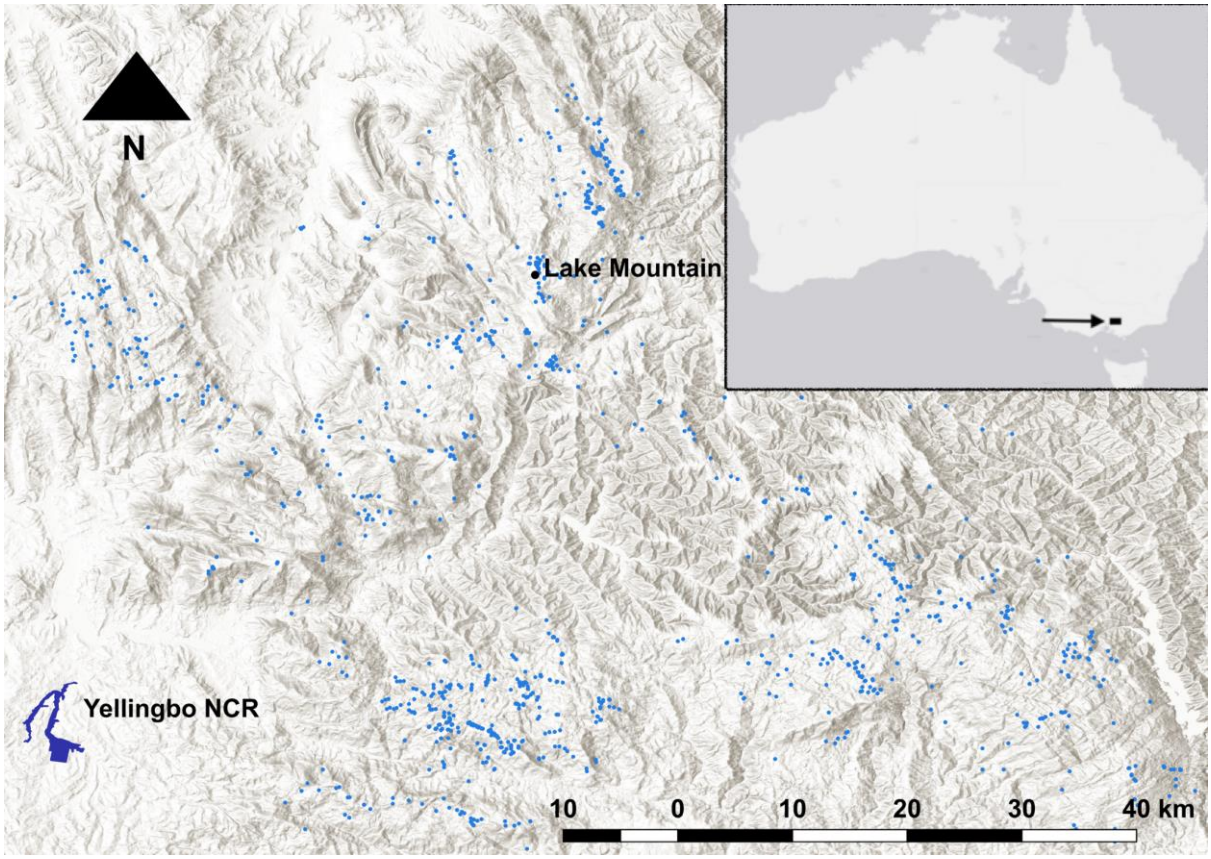

Figure 1: Contemporary distribution of Leadbeater's possum within Australia (black rectangle and arrow on the inset), including the range of the lowland population (Yellingbo NCR; dark blue area), and location presence records of the species in the Central Highlands (light blue points). The location of the Lake Mountain population used for comparison of genetic diversity with the Yellingbo samples is indicated.

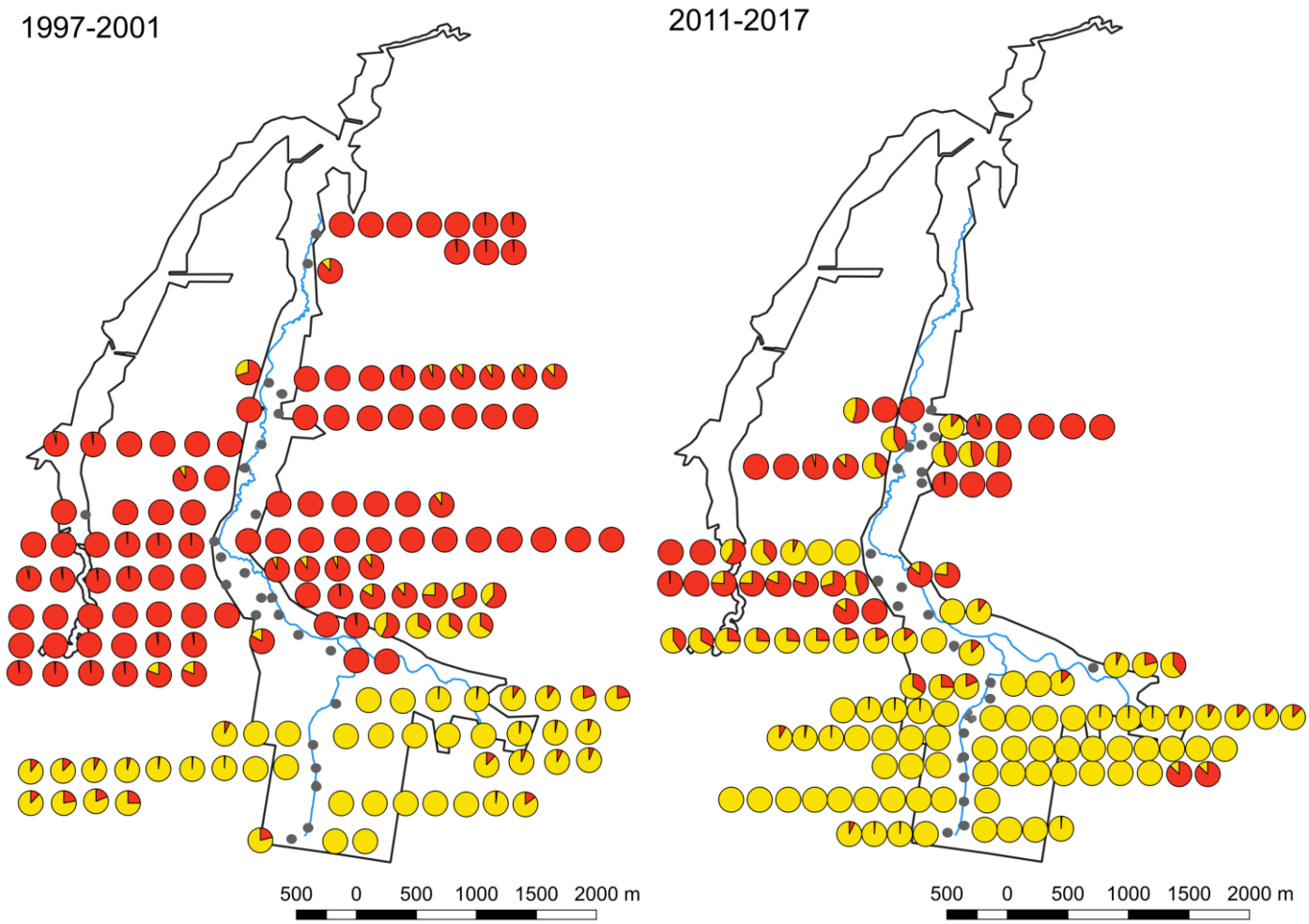

Figure 2: Geographic distribution of the two genotypic clusters (Cluster 1, prevalent in the north, red; Cluster 2, prevalent in the south, yellow) at Yellingbo for the two genetic sampling periods, 1997-2001 (n=157) and 2011-2017 (n=130). Sampling locations are denoted by grey dots, and the pie charts adjacent to each location indicate the membership coefficients (Q) of individuals in two genotypic clusters (STRUCTURE analysis, K=2, 1,207 loci). The main ephemeral creeks running through the reserve are indicated in blue. The same genetic structure was also detected using genetic PCA (Figure 3).

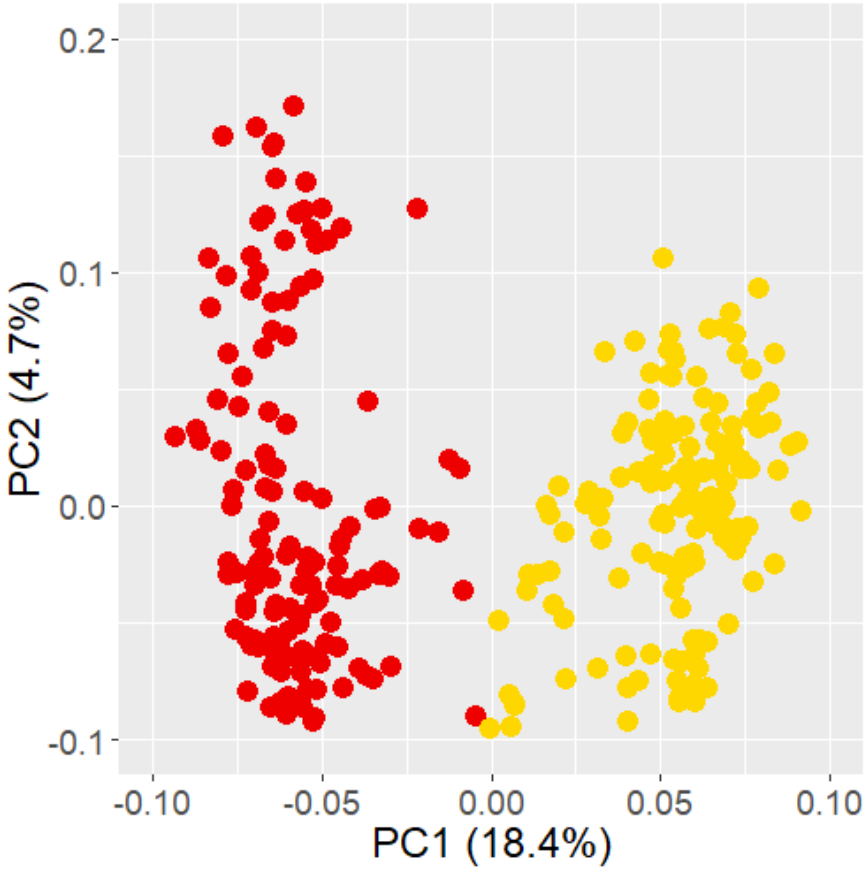

Figure 3: Plots of PC1 vs PC2 for principal components analyses of genetic variation within Yellingbo. Percentage of total variance explained by each principal component is indicated. Analyses were run for all genotyped Yellingbo individuals from 1997 to 2017 using 1,207 SNP loci. Individuals are coloured according to being assigned to northern (red) or southern (yellow) genetic groups. Assignments to these genetic groups were based on the membership in respective genotypic cluster detected by K=2 STRUCTURE (Figure 2: individuals with Q>0.5 in red Cluster 1 are assigned to northern group; those with Q>0.5 in yellow Cluster 2 to southern group).

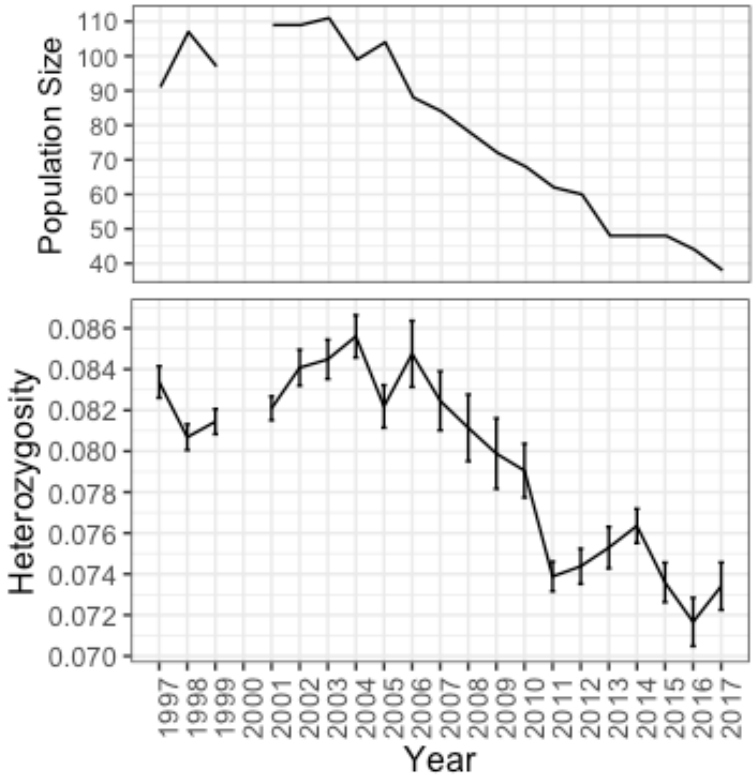

Figure 4: Decline in number of individuals (top graph) and genetic diversity (bottom graph) over time at Yellingbo. Mean individual heterozygosity (PHt) ± 1 SE is shown for each year of sampling. Heterozygosity values are based on individuals known to be alive in any given year for which genetic samples were available. Intense population monitoring was undertaken in all years except 2000 and genetic samples collected from 1997-2001 and 2011-2017.

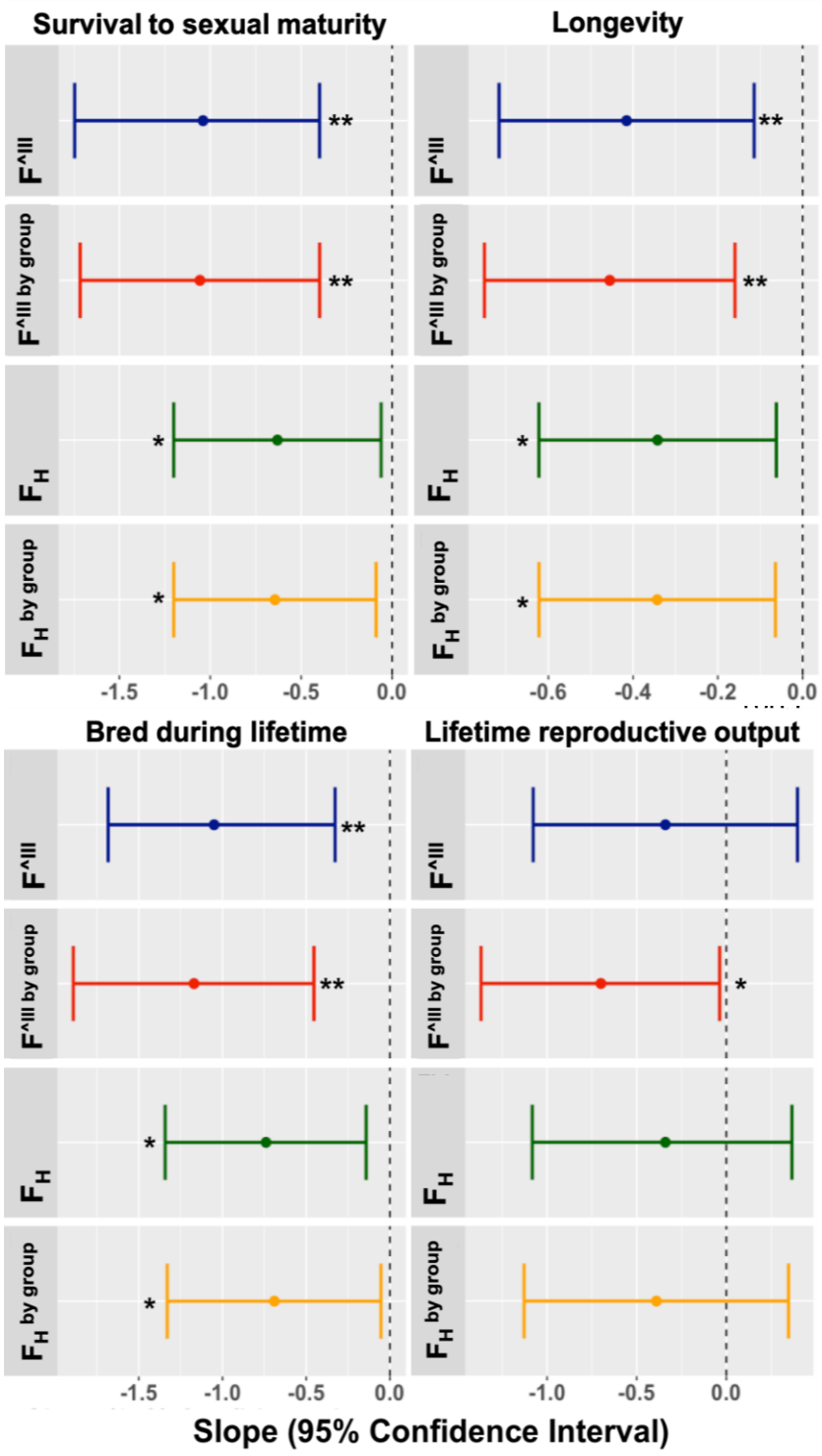

Figure 5: Slope estimates for genomic inbreeding coefficients $F^{\wedge III}$, F_H and their versions adjusted for population substructure $F^{\wedge IIIby\ group}$ and $F_H^{by\ group}$ as predictors of four fitness traits in GLMs. Estimated slopes with 95% confidence intervals of the average of the top models (i.e. ΔAICc <2) for each standardised predictor in models of survival to sexual maturity (n=237), longevity (n=183), whether individuals bred during their lifetime (n=211), and lifetime reproductive output (n=179). Full model estimates are given in Supporting Table S3 and predictor weights and importance are given in Supporting Tables S4-S11. Significance of predictors is indicated by asterisks (* p<0.05; ** p<0.01). $F^{\wedge IIIby\ group}$ was the only inbreeding coefficient that significantly predicted fitness for all traits.

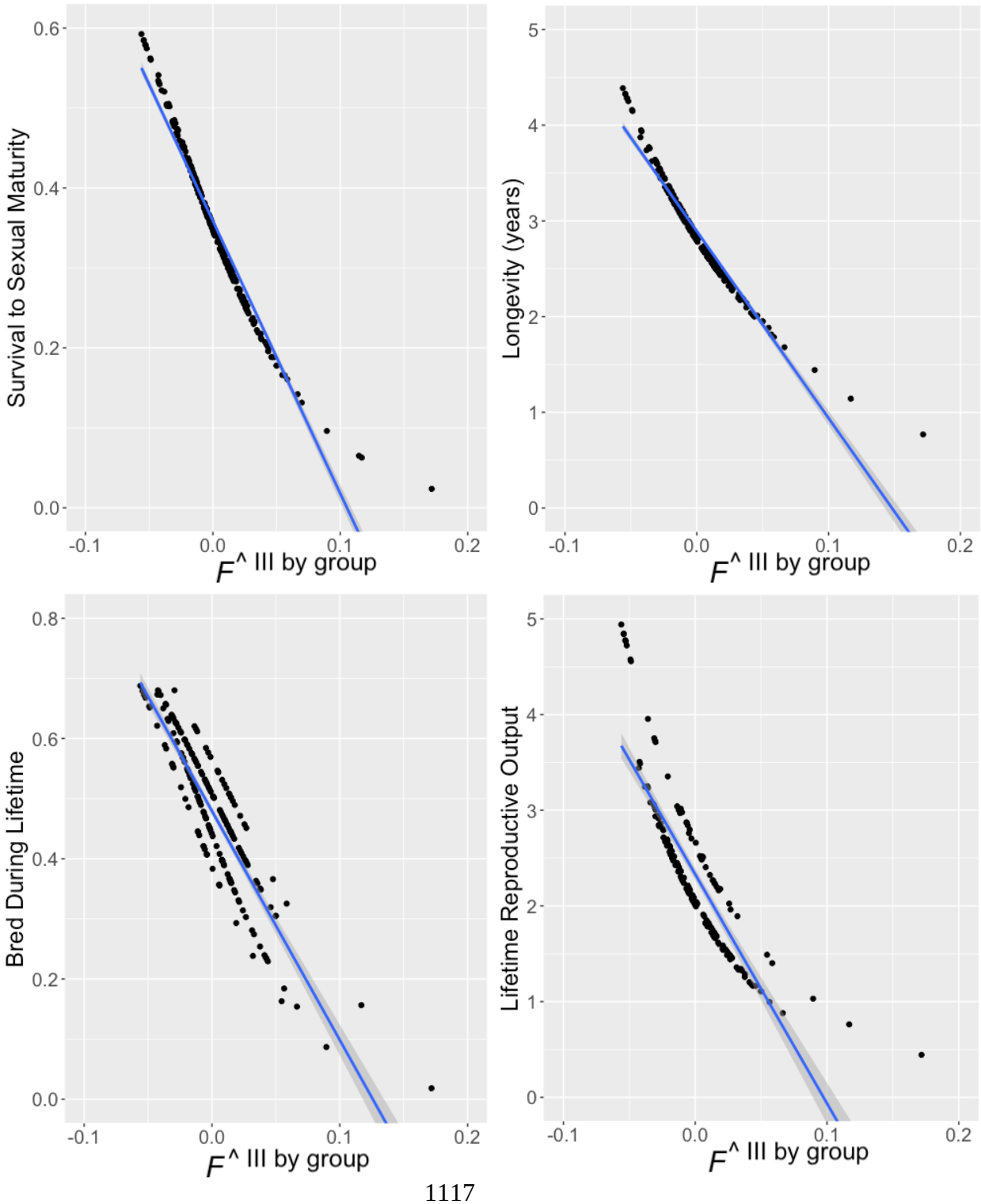

Figure 6: Relationships between inbreeding coefficient $F\hat{}^{\text{IIIby group}}$ and predicted values of fitness: survival to sexual maturity, longevity, whether individuals bred during their lifetime, and lifetime reproductive output. Values were calculated by averaging predictions of the top models (i.e. ΔAICc <2). The 95% confidence intervals for linear regression are also shown (dark grey). The full range of observed values for longevity and lifetime reproductive output are shown in Supporting Figure S14.

1125

Table 1: The number of individuals first captured as juveniles (juvs) and adults (ads) and the percentages of these that had parents assigned to them using genetic assignment (Juvs CERVUS), observation of pair bonding in colonies (Juvs Observed), or the sum of these (CERVUS + Obs.).

Period	**Juvs**	**Ads**	**Parent**	**Juvs CERVUS**	**Juvs Observed**	**Juvs CERVUS + Obs.**	**Ads CERVUS**
1995-2001	150	89	Mother	64 (42.7%)	145 (96.7%)	150 (100%)	36 (40.4%)
			Father	67 (45%)	140 (95.0%)	145 (97.0%)	30 (33.7%)
2002-2010	149	113	Mother	2 (1.3%)	143 (96.0%)	144 (96.6%)	5 (4.4%)
			Father	4 (3.0%)	136 (91.0%)	136 (91.0%)	5 (4.4%)
2011-2017	60	45	Mother	36 (60.0%)	47 (78.3%)	47 (78.3%)	30 (66.7%)
			Father	25 (42.0%)	44 (73.0%)	45 (75.0%)	29 (64.4%)

1126

Table 2: Haploid lethal equivalents for four fitness traits. Estimates are given as the regression slope and 95% confidence interval of the relationship between inbreeding coefficient and fitness trait values. NA indicates where estimates were not generated because there was no significant relationship between variables in GLMs.

	Survival to sexual maturity	**Longevity**	**Bred during lifetime**	**Lifetime reproductive output**
$F^{\wedge III}$	8.60 (3.66-14.25)	5.08 (1.50-8.74)	6.36 (2.42-10.83)	NA
$F^{\wedge III by\ group}$	11.46 (5.13-18.49)	8.06 (2.84-13.44)	9.13 (3.94-14.63)	13.36 (2.37-25.99)
F_H	1.35 (0.18-2.63)	1.17 (0.20-2.15)	1.28 (0.27-2.31)	NA
$F_H^{by\ group}$	1.28 (0.21-2.37)	1.09 (0.21-1.98)	1.02 (0.15-1.88)	NA

1127