DR. DANIEL J SCHMIDT (Orcid ID : 0000-0001-5638-497X)

Article type : Original Article

Monitoring age-related trends in genomic diversity of Australian lungfish.

Daniel J Schmidt1,7, Stewart Fallon2, David T Roberts3, Thomas Espinoza4, Andrew McDougall1, Steven G Brooks5, Peter K Kind4, Nick R Bond1,6, Mark J Kennard1, Jane M Hughes

1 Australian Rivers Institute, Griffith University, Nathan, Queensland, 4111, Australia
2 Radiocarbon Facility, Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia
3 Seqwater, 117 Brisbane Street, Ipswich, Queensland, 4305, Australia
4 Department of Natural Resources and Mines, PO Box 1167, Bundaberg, Queensland, 4760, Australia
5 Queensland Department of Agriculture, Fisheries and Forestry, Brisbane, Queensland, 4001, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.14791

This article is protected by copyright. All rights reserved
Abstract

An important challenge for conservation science is to detect declines in intraspecific diversity so that management action can be guided towards populations or species at risk. The lifespan of Australian lungfish (Neoceratodus forsteri) exceeds 80 years, and human impacts on breeding habitat over the last half century may have impeded recruitment, leaving populations dominated by old post-reproductive individuals, potentially resulting in a small and declining breeding population. Here we conduct a “single-sample” evaluation of genetic erosion within contemporary populations of the Australian lungfish. Genetic erosion is a temporal decline in intraspecific diversity due to factors such as reduced population size and inbreeding. We examined whether young individuals showed signs of reduced genetic diversity and/or inbreeding using a novel bomb-radiocarbon dating method to age lungfish non-lethally, based on $^{14}$C ratios of scales. A total of 15,201 single nucleotide polymorphic (SNP) loci were genotyped in 92 individuals ranging in age from 2 to 77 years old. Standardized individual heterozygosity and individual inbreeding coefficients varied widely within and between riverine populations, but neither was associated with age, so perceived problems with recruitment have not translated into genetic erosion that could be considered a proximate threat to lungfish populations. Conservation concern has surrounded Australian lungfish for over a century. However our results suggest that long-lived threatened species can maintain stable levels of intraspecific variability when sufficient reproductive opportunities exist over the course of a long lifespan.

Keywords: sequence-based genotyping; RADseq; allelic richness; gene diversity; standardized multilocus heterozygosity; identity disequilibrium, inbreeding coefficient.
1. Introduction

Intraspecific variation is both the product of, and the foundation of, evolutionary and biological processes from populations to ecosystems. Preserving intraspecific diversity is increasingly recognised as important for long-term sustainable management of populations and species (Mimura et al. 2017; Ottewell et al. 2016). This is because a critical determinant of species’ potential adaptation to environmental change is the availability of genetic variation underlying fitness (Allendorf et al. 2012). However, additive genetic variance is difficult to measure on the scale required to detect and manage global biodiversity decline. A more tractable metric - genotypic diversity (i.e. variation in genotypes among individuals) is amenable for monitoring programs that track ongoing changes in intraspecific variation that may have consequences for conservation of target species (Allendorf et al. 2012; Mimura et al. 2017). For example, using high-throughput sequencing, Hoffmann et al. (2014) demonstrated a link between the risk of inbreeding depression within a population of harbour seals and observed reductions in genotypic diversity of young individuals relative to older individuals. More generally, allelic diversity has been proposed as a candidate “essential biodiversity variable” that could be monitored for early detection of critical and potentially long lasting biodiversity change (Schmeller et al. 2017). Allelic diversity is simply the count of unique alleles per locus, often expressed as “allelic richness” which incorporates a transformation to facilitate comparison among samples of unequal size (see Methods, section 2.5). Allelic diversity is a measure of genetic diversity indicative of a population’s long-term potential for adaptability and persistence (Greenbaum et al. 2015). Allelic diversity is generally reduced in vertebrate species classed as threatened relative to those classed as non-threatened (Willoughby et al. 2015). Simulation studies indicate allelic diversity is a powerful metric for detection of genetic erosion of a population – particularly when measurements are based on thousands of loci as is now feasible using high-throughput sequencing methods (Hoban et al. 2014).

While natural resource and conservation managers are increasingly attentive to the need to monitor the genetic status and trends of wild populations’ genetic diversity, to date there is a limited number of empirical studies that have applied high-throughput genomic-scale data to temporal monitoring of diversity in target species (e.g. Chen et al. 2016). This deficiency is one example of a wider phenomenon termed “the...
The present study represents collaboration between conservation managers and population genetics researchers to monitor genome-wide diversity in an iconic target species where all extant populations are considered under threat.

The oldest fossil record of the Australian lungfish *Neoceratodus forsteri* is from the early Cretaceous, suggesting *N. forsteri* is possibly the oldest extant vertebrate species on the planet (Cavin & Kemp 2011). Lungfishes belong to the subclass Dipnoi – air-breathing fishes – that flourished in the Devonian period and are the closest living relatives of all tetrapods (Amemiya et al. 2013). Australian lungfish is a freshwater species that has motivated intense conservation interest for over 100 years. As early as the 1890’s there was sufficient concern over the scarcity of juveniles to prompt a translocation program to minimise the risk of extinction (Kind 2011). The Brisbane River is thought to have been populated via translocated individuals sourced from the Mary River in the late 1800’s (Hughes et al. 2015). There remains some conjecture over whether an endemic population existed in the Brisbane prior to this translocation attempt (Anonymous 2017). The natural distribution of Australian lungfish is otherwise limited to two rivers in southeast Queensland - the Burnett and Mary.

Mitogenome data indicate lungfish populations in the Burnett and Mary rivers have experienced isolation since the late Pleistocene (~150,000 years ago), although a more recent connection may have been facilitated by lowered sea levels ~12,500 years ago (Bishop et al. 2018).

Lungfish are listed as ‘vulnerable’ under the *Environment Protection and Biodiversity Conservation Act 1999*. Justification for their conservation status was based mainly on the threat posed by water impoundments, which alter natural flow regimes, restrict movement of adults, and adversely alter breeding habitats (Anonymous 2017). Impoundments can negatively impact the establishment of dense macrophyte beds required for successful spawning and recruitment. Recruitment appears to be naturally sporadic and adult mortality is low (Kind 2011). Therefore further restriction of breeding opportunities may lead to populations dominated by aging, post-reproductive individuals, as reproductive activity may be limited to 50 years of age (Kind 2011).

Under this scenario populations could undergo decline in the near future (Anonymous 2014, 2017). If the breeding population is declining we might expect to find erosion of genetic diversity in younger age fish relative to older fish due to reduced population
size and/or inbreeding. Testing this prediction is the goal of this study, along with
assessment of within catchment genetic variability which is a high priority under the
Australian Government’s Draft Recovery Plan for Australian Lungfish (Anonymous
2017).

Several challenges are involved in monitoring intraspecific genetic diversity of
Australian lungfish. Firstly, studies using traditional genetic markers (allozymes,
microsatellites, mitochondrial DNA) all indicate that population-level genetic
diversity is very limited in *N. forsteri* (Bishop *et al.* 2018; Frentiu *et al.* 2001; Hughes
*et al.* 2015). For example, mean observed heterozygosity based on eleven
polymorphic microsatellites was 0.35 (Hughes *et al.* 2015), which is low even in
comparison with threatened fish taxa (Willoughby *et al.* 2015). A large single
nucleotide polymorphism (SNP) dataset representative of genome-wide variation may
therefore provide better resolution of population genetic parameters in lungfish.

Restriction site associated DNA sequencing (RADseq) has recently emerged as a rapid
and economical means of genotyping thousands of SNPs in virtually any organism

Secondly, with a genome size of around 50 Gb, *Neoceratodus* ranks among the largest
known vertebrate genomes (15-fold larger than humans), likely due to duplication of
transposable elements (Metcalfe & Casane 2013). An extremely large genome
requires a SNP genotyping method targeting a low density of sequenced fragments.
We’ve employed a modified version of RADseq known as sequence-based-
genotyping (SBG) that has been successfully used in studies of plants with large
polyploid genomes (Gazave *et al.* 2016; Truong *et al.* 2012).

Monitoring genetic diversity of long-lived fish species has been successful where
annual spawning behaviour permits time-series analysis of larval genetic diversity
(e.g. endangered razorback sucker; Dowling *et al.* 2014). Australian lungfish spawn
irregularly in response to environmental conditions and larval stages are difficult to
collect, so monitoring diversity via larval sampling is not feasible (Kind *et al.* 2011;
Espinoza *et al.* 2014). An option for monitoring long-lived organisms recommended by
Mimura *et al.* (2017) was comparative evaluation of genetic variation in adults and
juveniles, an approach used previously for genetic monitoring of forest trees (Kettle *et al.*
2007; Vranckx *et al.* 2014). However, such comparisons depend on the ability to
accurately age individuals in the population, something that is difficult for species with indeterminate growth rates such as lungfish. Here we adopted the method of James et al. (2016), who developed a novel aging method using the signal of bomb $^{14}$C radiocarbon in lungfish scales to reliably estimate the age of individuals.

Our primary aim was to monitor genomic diversity and assess genetic erosion in the Australian lungfish by testing for a statistically significant decline in key diversity metrics (standardised multilocus heterozygosity, allelic richness and gene diversity) between age classes living contemporaneously within natural and translocated populations. We calibrated the age of individual fish using the bomb radiocarbon method of James et al. (2016) and estimated genetic diversity using the SBG method of Truong et al. (2012). Previous genetic datasets (allozymes, microsatellites, mtDNA) have yielded different insights into spatial subdivision and demography of extant lungfish populations. We therefore also assessed spatial subdivision for two natural lungfish populations (Burnett River, Mary River) and a translocated population (Brisbane River) using the SBG dataset, with the expectation that a large SNP dataset would provide improved resolution of population structure.

2. Materials and Methods

2.1. Fieldwork and sample selection

Australian lungfish sampling was conducted with an electrofishing boat in the Burnett, Mary and Brisbane Rivers from 2013 to 2015 (Australian Ethics Committee protocol number ENV/17/14/AEC; Queensland fisheries permit number 174232). A total of 1,492 individuals (Brisbane = 498, Mary = 488, Burnett = 506) were sampled from the three rivers as part of an associated study of population age structure. Tissue samples approximately 1cm² in size were removed from the dorsal or anal fin for genetic analysis, and two scales were removed from above the lateral line of the body for radiocarbon aging of the same individual. Fish were released alive at the point of capture. Radiocarbon aging and sequenced-based-genotyping could only be performed on a fraction of our sample due to cost, so for genetic analysis we subsampled ~15 individuals from the upper and lower tenth percentile of body length distribution in each river (~30 individuals per river). This approach was expected to capture a sample...
of the youngest and oldest individuals within each river. Sample data for the 92
selected individuals is available in Supplementary File S1.

2.2. Radiocarbon aging of lungfish scales

Scale samples were prepared for radiocarbon aging using the methods of James et al.
(2016) and Fallon et al. (2015). A detailed description of the dating method and its
validation is available (Fallon et al. in review). In summary, each scale was cleaned to
remove the upper squamulae surface using a diamond burr. Scales were then sliced
into 1mm increments and approximately 1.5mg of material was combusted to CO$_2$
and then converted to graphite at the Radiocarbon Laboratory at The Australian
National University. Approximately 10 samples were measured on each scale. In order
to obtain calendar ages for the birth of the lungfish a $^{14}$C reference curve was
developed (Fallon et al. in review). These dates provided a basis for determining the
year of birth using the von Bertalanffy growth function following James et al. (2016).

2.3. Sequencing-based genotyping

Whole genomic DNA was isolated from fin tissue using the DNeasy Tissue Kit
(Qiagen, Valencia, CA) and treated with RNase A (Qiagen). Libraries were prepared
by Floragenex, Inc. (Eugene, OR, USA) following the SBG protocol of Truong et al.
(Truong et al. 2012), using ~500 ng of DNA, double-digested with restriction
enzymes PstI and MseI and size selected to target a low density of sequenced
fragments (typically 5,000 – 15,000). The library was sequenced on two lanes of the
Illumina HiSeq2000 platform with single-end 100 base pair (bp) chemistry. FASTQ
sequence data was demultiplexed by individual index and trimmed to 91bp using
Stacks v1.48 (Catchen et al. 2013). Two independent libraries were prepared for three
individuals serving as technical replicates for assessment of genotyping error.

Stacks v1.48 was used for de novo assembly and genotype calling of SBG loci
(Catchen et al. 2013). Low quality reads (phred score < 10 within 13 bp sliding
window) were discarded using process_radtags. Optimal de novo assembly parameters
were determined following the protocol of Rochette and Catchen (2017), involving
repeated runs of the denovo_map pipeline on a subsample of 12 individuals to assess
how variation in assembly parameters affected the number of loci assembled. The best
A combination of assembly parameters was minimum stack depth of three reads (-m, 3); up to two mismatches allowed between stacks (-M, 2) and maximum of two mismatches between catalog loci (-n, 2). These parameters were used to build an initial catalog of loci based on all individual samples. Corrections to genotype calls were then applied using \textit{rxstacks}, implementing the bounded SNP model (upper bound 0.05) and log likelihood filter threshold of -5. The threshold was chosen by inspecting the distribution of mean likelihood scores for all loci. A catalogue of corrected genotype calls was then rebuilt and matched to samples using \textit{cstacks} and \textit{sstacks}. The program \textit{populations} was used to produce a filtered set of genotype calls for each individual, where filtering constraints included:

1) all loci must be biallelic;
2) each locus must be scored in $\geq80\%$ of individuals in the entire sample of 92 individuals;
3) all loci are represented by a single SNP position (i.e. –write_single_snp flag).
4) any locus showing significant deviation from expected Hardy-Weinberg proportions in $>1$ river population sample was excluded from further analyses.

\section*{2.4. Analysis: relationship between age, individual inbreeding coefficient and multilocus heterozygosity.}

The inbreeding coefficient ($F$) is the probability that two alleles at a locus in an individual were inherited from a common ancestor, and can be estimated at the level of an individual or a population. A maximum likelihood-based (ML) estimate of $F$ for each individual was made using the function ‘inbreeding’ in the R package adegenet (Jombart et al. 2011). This function has numerical approximation issues when large numbers of loci are used simultaneously. Therefore an estimate of $F$ for each individual was taken as the median value from 1,000 repetitions, where at each repetition a ML estimate of $F$ was made on 500 loci sampled without replacement.

Standardized multilocus heterozygosity (sMLH), is an individual-based metric defined as the total number of heterozygous loci in an individual divided by the sum of average observed heterozygosities in the population, over the subset of loci successfully typed in the focal individual (Coltman \textit{et al.} 1999). sMLH was calculated.
using R package ‘inbreedR’ (Stoffel et al. 2016). The background sample used to
standardize sMLH for testing the relationship between $F$ and sMLH was the river of
capture for each individual lungfish. For testing the effect of individual age and/or
river on variance in sMLH, the pooled dataset was used to standardize sMLH.

Linear fixed effects models were used to test whether variance in a dependent variable
(sMLH or $F$ of individual fish), was explained by river of origin, individual age in
years, and/or interactions between river and age. Linear model results are presented
using the Burnett as the reference river, although the Mary was also tested since both
natural populations were predicted to have greater diversity relative to the translocated
Brisbane population. Evidence for some deviation of residuals from normality under a
linear model prompted additional fitting of the data to a quasipoisson generalized
linear model, and also calculation of p-values via a permutation method (R package
‘lmPerm’ Wheeler & Torchiano 2016).

2.5. Analysis: comparing diversity between old and young age classes

Two intraspecific diversity metrics, allelic richness ($A_R$) and gene diversity ($H_S$) were
used to test for a reduction in diversity between old and young age classes within each
of the three river samples. Allelic richness ($A_R$) is the count of unique alleles in a
sample, which can be either 1 or 2 for a single locus with biallelic SNPs. To account
for variation in sample size and amount of missing data per locus, estimates of $A_R$
were standardized by the rarefaction approach of El Mousadik and Petit (1996), where
allele counts for all loci are reduced to an expected value based on the minimum
sample size per locus, using function `allelic.richness` in the R package ‘hierfstat’
(Goudet 2005). Gene diversity ($H_S$) is Nei and Chesser’s (1983) estimator of within
population expected heterozygosity that is unbiased for small sample size. $H_S$
calculation was modified from the function `basic.stats` in R package ‘hierfstat’
(Goudet 2005) and the code is provided in Supplementary File S3.

Permutation procedures were used to evaluate whether $A_R$ and $H_S$ were significantly
greater in the old age class relative to the young age class within each riverine
population sample. Tests consisted of generating a null distribution of the test statistic,
by randomly shuffling individuals between the two groups (old, young) 1,000 times

This article is protected by copyright. All rights reserved
without replacement. The test statistic was the difference in mean A_R (or H_S) between old and young groups (e.g. mean H_S,old – mean H_S,young). Calculation of A_R on randomly permuted data results in some variation among replicates in the minimum sample size ("min.n") used for rarefaction of allele counts. Therefore permutation tests for A_R were based on a subset of permuted replicates where min.n was a fixed value (i.e. the highest frequency min.n among replicates). The one-sided p-value for the hypothesis that A_R (or H_S) is greater in the older age class relative to the younger age class was obtained from the frequency of simulated (null) values of the test statistic that were greater than or equal to the observed test statistic. Functions used for genetic erosion permutation tests are demonstrated with simulated data in an R tutorial provided as Supplementary File S3.

2.6. Analysis: population-level inbreeding and spatial genetic structure

Evidence for inbreeding within population samples was determined using the identity disequilibrium statistic (g2; Hoffman et al. 2014). Identity disequilibrium is the extent to which heterozygosity is correlated across loci and is a proxy for variance in inbreeding within a population. g2 was calculated using function g2_snps in the R package ‘inbreedR’ (Stoffel et al. 2016). F-statistics among the three river samples were calculated using functions in the R package ‘hierfstat’ (Goudet 2005). Spatial patterns of population structure were examined using sparse non-negative matrix factorization (sNMF; Frichot et al. 2014), implemented in the R package ‘LEA’ (Frichot & François 2015). Additional testing of spatial population structure was performed using discriminant analysis of principle components (DAPC), used to visualize relationships among groups of samples and assign individuals to these groups (Jombart et al. 2010). Firstly k-means clustering was used to identify differentiated genetic groupings within the sample of 92 individuals without predefined membership based on population of origin, and then DAPC analysis was performed on the genetically-defined clusters. The DAPC cross-validation procedure was used to select an appropriate number of principal components and discriminant functions, and was implemented in the R package ‘adegenet’ (Jombart & Ahmed 2011). The relationship among samples was also inferred using ‘fineRADstructure’, producing a coancestry matrix among individuals based on an expanded dataset including all SNPs (and all alleles) per locus (Malinsky et al. 2018). Data filtering
rules for fineRADstructure were as above except no constraint was placed on the
number of alleles per locus or on the number of SNPs per allele.

3. Results

3.1. Summary of radiocarbon aging.

A total of 32 fish were aged from the Burnett River population, 30 fish from the Mary
and 30 from the Brisbane River. The Burnett River fish aged from 4-72 years old. The
Mary River fish aged from 4-77 years old while the Brisbane River fish aged from 2-
62 years old (Fallon et al. in review). Analysis of genetic erosion based on comparison
of diversity between “old” and “young” age classes within each population required
dividing each river sample into two groups of approximately even size. Because age
structure of lungfish was different in the three river samples, this meant selecting an
arbitrary cut-off value for delineating old and young age classes in each river. In the
Burnett and Mary 40 years of age was the cut-off between age classes, while 23 years
of age was the cut-off in the Brisbane (Figure 1A). Note that age was also treated as a
continuous variable in linear models examining whether individual diversity metrics
were explained by an individual’s age and/or by river of origin.

3.2. Summary of genomic data.

Quality filtering of 206 million raw reads passed 192 million reads for denovo
assembly at an average of 2.00 million per individual (sd = 0.23 million). The initial
catalog generated by denovo_map contained 462,953 loci. Error correction using
rxstacks reduced the number of catalog loci to 436,849 of which 15,581 loci passed
filtering constraints no. 1-3. The fourth filtering constraint removed loci that deviated
significantly from HWP in > 1 river sample, leaving the final dataset with 15,201 loci.
Average number of loci scored across 92 individuals was 14,023 (sd = 1,057) and
average missing data per locus was 12.2%. Average read depth per locus across the 92
individuals was 9×, and genotyping error based comparison of 15,201 loci genotyped
independently across three duplicate individuals was 3.2%.
3.3. Tests for an effect of age on genetic diversity.

A linear fixed effects model tested whether variance in sMLH was explained by river of origin, individual age in years, and/or interactions between river and age. The full model explained a significant amount of variation in sMLH ($F_{5, 86} = 10.7, p < 0.0001$, $R^2 = 0.38, R^2_{\text{adjusted}} = 0.35$; Table 1). ANOVA showed age of individual fish did not predict variance in sMLH on its own ($p = 0.35$), nor interact with river ($p = 0.27$). River alone was a highly significant predictor of sMLH ($p < 0.0001$). Lungfish from Brisbane ($p < 0.0001$) and Mary ($p = 0.016$) showed significant reduction in sMLH relative to the Burnett river (Table 1). Likewise, when the Mary was used as reference population the Brisbane showed a significant reduction in sMLH ($p = 0.005$). Figure 1B illustrates the overlapping interquartile range of young and old age classes, and illustrates the reduction in sMLH observed in the Brisbane River population.

Estimates of individual inbreeding coefficient ($F$) ranged from ~0 to 0.32 with a mean of 0.13 (Figure 2). Modelling $F$ as a function of age and river reflected a similar pattern to sMLH reported above (Table 1), with individual age of fish not predicting variance in $F$ on its own ($p = 0.30$) nor interacting with river ($p = 0.12$). However river alone was a significant predictor of $F$ ($p = 0.0002$), with Mary exhibiting significantly higher $F$ relative to Burnett ($p = 0.004$; Table 1) and Brisbane ($p = 0.028$). Residual diagnostics on linear models suggested evidence for deviation from normality related to overdispersion. However use of a quasipoisson generalized linear model yielded essentially the same results as the linear models, except for a marginally significant negative interaction between age and $F$ in the Brisbane River (Table 1). Calculation of p-values via a permutation approach also confirmed the significant effect of river on genetic diversity, with no effect of age (Table 1).

Genetic variation at age class-level was quantified using allelic richness ($A_R$) and gene diversity ($H_S$) based on comparison of loci that were polymorphic within each river (Burnett = 10,282; Mary = 10,644; Brisbane = 7,837). Mean values of $A_R$ and $H_S$ were almost identical between age classes within each river (Table 2). Permutation tests for reduced variation in the young relative to the old age class using $A_R$ as the test statistic were not significant in any of the three rivers (Table 2; Supplementary File S2). Indeed, the mean observed $A_R$ was actually higher in young fish relative to
old fish within the Burnett and Brisbane River populations (Table 2). Permutation tests based on $H_S$ as test statistic yielded similar results to allelic richness – no evidence for significant reduction in $H_S$ of young fish (Table 2; Supplementary File S2).

### 3.4. Inbreeding and spatial genetic structure

Figure 2 summarizes overall relationships between inbreeding, individual heterozygosity, river of origin and individual age. Overall there was a strong negative correlation between individual heterozygosity (sMLH) and individual inbreeding coefficient ($F$) ($r = -0.899; p < 2.2e-16; $Figure 2$)$. The significant effect of river (see also Table 1) is clearly evident along with the non-significant effect of age (Figure 2).

On average, Mary River showed the highest inbreeding (mean $F = 0.17$), followed by Burnett and Brisbane (mean $F = 0.13$ and 0.11 respectively), although individual values varied widely (Figure 2). Inbreeding variance measured using identity disequilibrium ($g_2$) was estimated in the three riverine population samples and did not differ from zero in the Burnett River ($g_2 = 0.0053; p (g_2 > 0) = 1$), but was significantly greater than zero in both the Mary ($g_2 = 0.011; p(g_2 > 0) = 0.009$) and Brisbane ($g_2 = 0.0038; p(g_2 > 0) = 0.001$).

Global $F_{ST}$ among the three sample sites for 15,201 loci was 0.116 (95% CI: 0.113, 0.119). Admixture clustering using sNMF showed clear genetic subdivision ($k = 3$) between the three rivers (Figure 3A). Cross-validation suggested retention of the first 10 principal components in the DAPC, with proportion of conserved variance of 0.241. Optimal number of genetic clusters assessed using the Bayesian information criterion indicated three clusters, and these three groups corresponded to the three rivers of origin for 91 of the 92 individual fish. Relationships among the three rivers found by DAPC and fineRADstructure were consistent with admixture clustering (Figure 3B,C). DAPC scatterplot axis 1 separates the Burnett; axis 2 separates Mary and Brisbane; and the minimum spanning tree indicates closest affinity for the translocated Brisbane population was the Mary (Figure 3B). The fineRADstructure clustered coancestry matrix based on 31,822 loci also resolved three distinct populations, again showing the closest affinity for the translocated Brisbane population was the Mary (Figure 3C). One individual (sample code BNE_0153)
sampled from the Brisbane River assigned with posterior probability of 1.0 to the Mary River cluster in the DAPC analysis, and this individual clustered with the Mary group using both sNMF and fineRADstructure (Figure 3A,B,C).

4. Discussion

4.1. Genetic erosion and inbreeding

The Australian Government considers all Australian lungfish populations as under threat, due to uncertain population status, habitat degradation and flow regulation (Anonymous 2017). Existing adult population sizes based on mark recapture studies are on the order of thousands of individuals (Brooks, Roberts unpublished data), although juvenile lungfish are rarely encountered in surveys, and juvenile recruitment has long been a concern for population viability (Brooks & Kind 2002; Espinoza et al. 2013; Kind 2011). In addition, developmental abnormalities have been observed in some populations (Kemp 2014). Despite warnings that inbreeding and reduced diversity could be potential contributors to extinction risk (e.g. Frentiu et al. 2001), our analysis did not find evidence for genetic erosion in recent decades. We assessed wild-caught individuals ranging in age from 2 to 77 years and found no relationship of age with individual heterozygosity (sMLH) or inbreeding coefficient ($F$) when age was treated as a continuous variable. We also did not find a detectable reduction of allelic richness ($A_R$) or gene diversity ($H_S$) between old and young age classes when age was treated as a factor. Age classification (“young”, “old”) was necessarily different in the Brisbane River due to the different age structure of this population (see Fallon et al. in review), however permutation tests can be considered independent for the three river samples.

Previous studies have demonstrated that a relationship between age structure and genetic diversity can be detected in natural populations. For example, Labonne et al. (2016) found higher heterozygosity in older brown trout ($Salmo trutta$) relative to younger trout in a population where variance in inbreeding was high. The increased diversity with age was interpreted as a heterozygosity-fitness correlation, where old age is a proxy for higher fitness. The pattern of increased diversity with age in brown trout was suggested to be a product of viability selection against inbred individuals.
that are less likely to survive to old age (Labonne et al. 2016). Genetic diversity in
forest trees has also been shown to vary among age-classes, with older trees exhibiting
higher diversity than seedling cohorts, a pattern attributed to a negative impact of
recent forest fragmentation on recruitment success (Kettle et al. 2007; Yineger et al.
2014).

Variance in inbreeding ($g^2$) was significant in two population samples (Mary and
Brisbane), and wide variation in individual inbreeding coefficients was observed in all
three populations suggesting that heterogeneity in consanguineous and non-
consanguineous matings may be common in lungfish. Levels of inbreeding equivalent
to half sib mating ($F = 0.125$) were common, and several individuals had $F$ estimates
$>0.25$ suggesting their parents were related at approximately the level of full siblings
(see Figure 2). High variance in inbreeding coefficients is regularly observed in small
isolated populations and linked to selection against inbreeding (Coltman et al. 1999;
Hoffman et al. 2014; Kardos et al. 2018; Labonne et al. 2016). If selection against
inbreeding is active in Australian lungfish populations then it seems unlikely to be
related to viability or senescence as we found no relationship between inbreeding and
age.

Observed stability of genetic diversity with age in three lungfish populations indicates
that a catastrophic loss of genetic diversity over the last half century has not occurred.
It may be the case that relatively stable genetic diversity across age classes is a
product of intermittent and opportunistic reproduction events distributed over the
course of a long life span in Australian lungfish. Successful spawning and recruitment
is believed to be associated with a narrow range of environmental conditions including
stable low-flow rates, optimal temperature and dissolved oxygen range, suitable
macrophyte spawning substrate, and post hatching habitat and for developing
juveniles (Brooks & Kind 2002; Espinoza et al. 2013; Kemp 1984; Roberts et al.
2014). For an organism reaching reproductive maturity around 15-20 years old and
reproductive activity possibly continuing up to 50 years of age, an individual may
experience 30-35 seasonal windows-of-opportunity for reproduction and successful
recruitment (Kind 2011). From a genetic diversity point of view, it would seem that
sufficient seasonal opportunities have been successful over the past half century to
maintain stable levels of genetic variability in the lungfish populations studied (albeit
with highly variable levels of inbreeding). However, should the number of

This article is protected by copyright. All rights reserved
opportunities be reduced through various anthropogenic impacts, we could expect to see this pattern change and the gradual loss of genetic diversity to occur. To prevent this genetic erosion process from occurring, management efforts should be directed towards avoiding any reduction in the number of successful recruitment opportunities that occur over an adult lungfish lifespan.

4.2. *Spatial subdivision*

The SBG protocol used here generated thousands of polymorphic SNP loci, which resolved clear genetic subdivision among three riverine lungfish populations. Significant differentiation between Burnett and Mary River populations was consistent with findings of previous studies using microsatellites (Hughes *et al.* 2015) and mitogenomes (Bishop *et al.* 2018). Whether the Brisbane River population is natural or founded by translocation from the Mary is unresolved (Anonymous 2017). Mitogenome haplotypes were shared between Brisbane and Mary, and microsatellite differentiation between these rivers was also relatively low (Bishop *et al.* 2018; Hughes *et al.* 2015). New SNP data presented here adds further information to the debate over origins of Brisbane River lungfish, with reduced diversity in the Brisbane relative to Mary consistent with the translocation scenario. SNP-based resolution of the Mary and Brisbane populations yielded one potential case of a recent illegal translocation. An individual sampled from the Brisbane (code BNE_0153), was confidently assigned to the Mary River population using all three clustering approaches. This lungfish was only two years old (Supplementary File S1), so given that both of its parents appear to be of Mary river origin, this individual could represent a translocation, or a release from captivity (although we cannot fully rule out a technical error in laboratory preparation). Lungfish are protected from fishing under the *Queensland Fisheries Act 1994* but numerous illegal and unreported movements of lungfish are known to have occurred (Kind 2011).

If the Brisbane River population is purely derived from a translocation that took place ~100 years ago, equating to three or four lungfish generations, then divergence of Brisbane from its source population we observe in the SNP data is relatively rapid. But this can be reconciled with the biology of founder events known from other systems, where even small numbers of founders (as low as 2 – 4 individuals) may...
produce demographically viable populations that exhibit divergence and reduced genetic diversity relative to their source population as long as the source population is not itself highly inbred (Szűcs et al. 2017).

5. Conclusion

We have combined a novel radiocarbon aging method with RADseq SNP genotyping to test whether genetic erosion can be detected within a contemporary population of a long-lived animal of high conservation significance, the Australian lungfish. Analyses of standardized multilocus heterozygosity, allelic richness, gene diversity and inbreeding showed no significant difference between old and young aged fish. Our results suggest that historical recruitment processes have been sufficient to maintain existing genetic diversity of the populations studied. However, the risk of future genetic diversity loss remains if recruitment processes are significantly altered from impacts such as dam expansions, hydrologic alternation or loss of critical habitats. We suggest the stability of age-related genetic diversity in lungfish could be due to their long reproductive lifespan, and further work is required to determine the crucial environmental conditions supporting successful spawning and recruitment opportunities, so that these conditions can be managed where possible to reduce the chance of genetic erosion. More work is also required to investigate possible association of fitness traits with variable inbreeding coefficients, and to further elucidate the history of extant populations (both natural and translocated) using SNP data, which offers the best resolving power for monitoring spatial and temporal trends in intraspecific diversity relevant for lungfish conservation.

Acknowledgements

Samples were collected under Fisheries Permit Numbers 140615 (Seqwater) and 174232 (Griffith University). We thank the following people for assistance: Kathryn Real (Griffith University); Sharon Marshall (DNRM); Jason Boone (Floragenex Inc.); Joel Huey (Western Australian Museum). Feedback and advice was provided by Giorgio Bertorelle, Robin Waples and anonymous reviewers. Funding for this study was provided by Australian Research Council Linkage grant LP13010018 with
support from South East Queensland Water (Seqwater), Queensland Department of Natural Resources and Mines (DNRM), Queensland Department of Agriculture, Fisheries and Forestry (DAFF).

**Data Accessibility**

Raw de-multiplexed sequences are available from NCBI Sequence Read Archive (SRA) (BioProject Accession# PRJNA477902). Lungfish sample information (specimen codes, age, length, location) and genetic summary data are provided in Supplementary File S1. R scripts used for analyses are supplied as Supplementary File S3.

**References**


conservation genetics and primary industry leads to mutually beneficial genomic

Evolutionary Histories and Patterns of Diversity across Subgenomes and
Subpopulations of Brassica napus L. Frontiers in Plant Science 7, 525.

Molecular Ecology Notes 5, 184-186.

Population Founding Events-A Stochastic Modeling Framework Incorporating Gene

Hoban S, Arntzen JA, Bruford MW, et al. (2014) Comparative evaluation of potential
indicators and temporal sampling protocols for monitoring genetic erosion.
Evolutionary Applications 7, 984-998.

Hoffman JI, Simpson F, David P, et al. (2014) High-throughput sequencing reveals inbreeding
depression in a natural population. Proceedings of the National Academy of Sciences
of the United States of America 111, 3775-3780.

Hughes JM, Schmidt DJ, Huey JA, et al. (2015) Extremely Low Microsatellite Diversity but
Distinct Population Structure in a Long-Lived Threatened Species, the Australian

Radiocarbon Dating the Scales of Australian Lungfish (Neoceratodus forsteri).
Radiocarbon 52, 1084-1089.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP
data. Bioinformatics 27, 3070-3071.

method for the analysis of genetically structured populations. Bmc Genetics 11.


Kemp A (1984) Spawning of the Australian lungfish, Neoceratodus forsteri (Krefft) in the
Brisbane River and in Enoggera Reservoir, Queensland. Memoirs of the Queensland
Museum 21, 391-399.

This article is protected by copyright. All rights reserved


**Figure Captions**

**Figure 1.** Evaluation of genetic erosion and population structure in Australian lungfish (*Neoceratodus forsteri*). **A.** Dotplot illustrating radiocarbon estimates of age (years) for lungfish from three riverine populations (Burnett, Mary and Brisbane rivers). Colours represent age classes (= young, old) used for permutation tests of genetic erosion; **B.** Boxplot showing variation in standardized multi-locus heterozygosity (sMLH) between age classes within riverine populations, based on 15,201 loci. Reference population used for calculation of sMLH was the pooled sample of 92 individuals.

**Figure 2.** Relationship between individual inbreeding coefficient (*F*), individual heterozygosity (sMLH), and age (years) for three riverine populations of
This article is protected by copyright. All rights reserved
Table 1. Summary of fixed effects models (linear and GLM) testing the effect of individual age and river of origin on individual heterozygosity (sMLH) and inbreeding coefficient (F). Reference river is Burnett. Model coefficients given with standard error in parentheses; p-values = *p<0.05; **p<0.01; ***p<0.001

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>sMLH OLS</th>
<th>sMLH glm: quasipoisson</th>
<th>Inbreeding coeff. (F) OLS</th>
<th>Inbreeding coeff. (F) glm: quasipoisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.0002 (0.0005)</td>
<td>-0.0002 (0.0004)</td>
<td>0.002 (0.003)</td>
<td>0.0003 (0.0004)</td>
</tr>
<tr>
<td>Mary River</td>
<td>-0.080 (0.032)</td>
<td>-0.078 (0.032)</td>
<td>0.534 (0.189)</td>
<td>0.082** (0.028)</td>
</tr>
<tr>
<td>Brisbane River</td>
<td>-0.142*** (0.031)</td>
<td>0.017 (0.026)</td>
<td>0.156 (0.208)</td>
<td>-0.144*** (0.031)</td>
</tr>
<tr>
<td>Age:Mary River</td>
<td>0.001 (0.001)</td>
<td>-0.001 (0.001)</td>
<td>-0.006 (0.004)</td>
<td>0.001 (0.001)</td>
</tr>
<tr>
<td>Age:Brisbane River</td>
<td>0.001 (0.001)</td>
<td>-0.001 (0.001)</td>
<td>0.001 (0.001)</td>
<td>0.001 (0.001)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.058*** (0.020)</td>
<td>0.117*** (0.017)</td>
<td>-2.140*** (0.135)</td>
<td>0.056** (0.020)</td>
</tr>
</tbody>
</table>

Note: In addition, p-values for linear models were calculated using a permutation approach giving same result.
Table 2. Comparison of age-class level genetic variability within three riverine populations of Australian lungfish.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Burnett Young</th>
<th>Burnett Old</th>
<th>Mary Young</th>
<th>Mary Old</th>
<th>Brisbane Young</th>
<th>Brisbane Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>n.poly.loci</td>
<td>10,282</td>
<td>10,282</td>
<td>10,644</td>
<td>10,644</td>
<td>7,837</td>
<td>7,837</td>
</tr>
<tr>
<td>n.ind</td>
<td>16</td>
<td>16</td>
<td>14</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>age mean (sd)</td>
<td>8.7 (4.7)</td>
<td>61.8 (7.0)</td>
<td>17.2 (11.4)</td>
<td>63.2 (7.0)</td>
<td>13.4 (6.9)</td>
<td>40.9 (13.0)</td>
</tr>
<tr>
<td>(A_R) mean</td>
<td>1.609</td>
<td>1.606</td>
<td>1.546</td>
<td>1.547</td>
<td>1.673</td>
<td>1.647</td>
</tr>
<tr>
<td>min.n</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>p-value ((A_R) young &lt; (A_R) old)</td>
<td>0.207</td>
<td>0.478</td>
<td>0.839</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H_S) (mean)</td>
<td>0.216</td>
<td>0.216</td>
<td>0.213</td>
<td>0.213</td>
<td>0.249</td>
<td>0.243</td>
</tr>
<tr>
<td>p-value ((H_S) young &lt; (H_S) old)</td>
<td>0.296</td>
<td>0.449</td>
<td>0.771</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.loci = number of polymorphic loci within a riverine population sample; n.ind = number of individual lungfish per sample; age = mean age of individual lungfish in years derived from bomb radiocarbon dating; \(A_R\) = allelic richness; min.n = number of alleles/locus/sample used for rarefaction of allelic counts in permutation test; \(H_S\) = gene diversity.
List of Supplementary Files

S1. Table of summary statistics for 92 individual lungfish. Data includes: sample code; river of origin; body length (mm); age class used in genetic erosion permutation tests; radiocarbon estimated age of birth; estimated age (years); number of loci scored after data filtering (see materials and methods); mean coverage per locus; standardized multi-locus heterozygosity where reference population is total pooled dataset (sMLH total); standardized multi-locus heterozygosity where reference population is population of origin (sMLH pop); inbreeding coefficient (F).

S2. Permutation tests: genetic erosion. Permutation tests for genetic erosion within three riverine populations of Australian lungfish to determine whether genetic diversity was lower in the young age class relative to the old age class. A-C.

Histograms showing null distribution of mean difference in allelic richness (A<sub>R</sub>) between age classes for three riverine populations. Dashed red line denotes observed difference in A<sub>R</sub> between age classes for each population. Red bars denote upper and lower 95% quartiles of distribution. Number of permutations retained for each test (n) and minimum sample size used for rarefied allele counts (min.n): A. n = 954, min.n = 12; B. n = 933, min.n = 10; C. n = 910, min.n = 12. Associated p-values are given in Table 1 of main paper. D-F. Null distribution of mean difference in gene diversity (H<sub>S</sub>) between age classes based on 1,000 permutations. Dashed red line denotes observed difference in H<sub>S</sub> between age classes for each population. Red bars denote upper and lower 95% quartiles of distribution. Associated p-values are given in Table 1 of main paper.

S3. R tutorial: “Genetic erosion analysis: example with simulated data”.

RMarkdown document illustrating functions and associated code used to evaluate genetic erosion using permutation tests of gene diversity (H<sub>S</sub>) and allelic richness (A<sub>R</sub>) between age classes (or between any two sample groups). Document is a self-contained tutorial, where an example dataset is simulated within the tutorial script. Requires following R libraries installed: hierfstat, parallel, ggplot2.
Age class: Young, Old

A

Burnett Mary Brisbane

Age (Years)

B

Burnett Mary Brisbane

Ind. Heterozygosity (SMLH)
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Schmidt, DJ; Fallon, S; Roberts, DT; Espinoza, T; McDougall, A; Brooks, SG; Kind, PK; Bond, NR; Kennard, MJ; Hughes, JM

Title:
Monitoring age-related trends in genomic diversity of Australian lungfish

Date:
2018-08-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/285177