Population genomics of *Uperoleia daviesae* (Anura: Myobatrachidae) highlights the vulnerability of naturally fragmented short-range endemics to urban development

Shengyao Lin [1] (0000-0002-1236-6793), Peter J. McDonald [2] (0000-0001-6875-1466), Alistair Stewart [2], Carolyn J. Hogg [3,4] (0000-0002-6328-398X), Luke W. Silver [3,4] (0000-0002-1718-5756), Sam C. Banks[5] (0000-0003-2415-0057), Graeme R. Gillespie[2] (0000-0001-9727-8436), Raphael K. Didham [1] (0000-0001-6685-7005), and Renee A. Catullo [1] (0000-0002-1790-7085)

[1]School of Biological Sciences, The University of Western Australia, Perth, WA 6009, Australia

[2]Flora and Fauna Division, Department of Lands, Planning and Environment, Northern Territory Government, Darwin, NT, Australia

[3]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

[4]ARC Centre of Excellence in Innovations in Peptide and Protein Science, The University of Sydney, Sydney, NSW 2006, Australia

[5]Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Northern Territory, 0815, Australia

Corresponding author: Shengyao Lin, Shengyao.lin@research.uwa.edu.au

Abstract

Urbanisation and land use change threaten short-range endemic amphibians. *Uperoleia daviesae*, the Howard River toadlet, is a threatened frog species endemic to sandsheet heath, a unique, naturally patchy mosaic of habitats near Darwin in Australia's Northern Territory. We generated a chromosome-level genome assembly and performed genome-wide SNP analyses using data from 115 individuals across 15 sites to assess dispersal patterns, genetic diversity, anthropogenic impacts, and conservation targets. Our findings reveal a history of past connectivity, followed by recent genetic subdivision, partly due to urban expansion. Regions 2 (north of Girraween and Humpty Doo) and 3 (south of Bees Creek and Lloyd Creek) contain most of the genetic diversity within the species. The two isolated populations, Region 1 (north of Howard Springs) and Region 4 (near Wickham and Weddell), each harbour unique alleles. We recommend treating local populations as a metapopulation and developing targeted conservation actions in each region. In-situ conservation actions, such as establishing protected areas or translocation, should be considered to maintain gene flow and genetic diversity. These findings provide a foundation for evidence-based management of *U. daviesae* and contribute to broader discussions on the conservation of short-range endemic species in increasingly urbanised landscapes.

Keywords: Uperoleia daviesae, population genomics, urbanisation impacts, short-range endemic species, conservation management

Introduction

Human activities have profoundly impacted vertebrate populations worldwide (Allan et al., 2019; Hoffmann et al., 2010), and frogs are among the most severely affected taxa (Blaustein et al., 2011; Blaustein & Kiesecker, 2002; Wake & Vredenburg, 2008). Over the past century, urban expansion (Hamer & McDonnell, 2008), agriculture (Smalling et al., 2015), and infrastructure development have led to habitat loss and fragmentation (Cushman, 2006), resulting in a dramatic reduction of suitable environments for many frog species (Cushman, 2006; Gallant et al., 2007). Additionally, factors such as altered hydrological regimes have disrupted environmental processes critical to frog survival (Coleman et al., 2024; Hazell, 2003; Kupferberg et al., 2012). The cumulative effects of these pressures have resulted in severe population declines in many frog species (Carrasco et al., 2021) and increased extinction risk (Blaustein et al., 2011; Blaustein & Kiesecker, 2002), underscoring the urgent need for conservation efforts grounded in an understanding of ecology and genetics (Blaustein & Kiesecker, 2002; Carrasco et al., 2021).

The Howard River toadlet (*Uperoleia daviesae*) is a poorly known, tiny (22-26 mm) burrowing frog endemic to the Northern Territory (NT) of Australia (Young et al., 2005). It is the only threatened frog species in the NT, classified as Vulnerable under the Commonwealth of Australia and NT legislation and Endangered in the IUCN Red List. It is also one of Australia's 26 most at-risk frog species, with a predicted probability of extinction by 2040 of approximately 8% (Geyle et al., 2022). These frogs are restricted to a naturally patchy mosaic of habitat, the sandsheet heath, during the breeding season (Young et al., 2005), which occurs in seasonally inundated drainage depressions embedded in a matrix of upland eucalypt woodland. The entire distribution of the species is within the greater peri-urban area of the city of Darwin (NT capital city), with an estimated Area of Occupancy of 216 km^2 (Department of Environment, Parks and Water Security, 2021). Each local patch of sandsheet heath is typically less than 1 ha in size (Clancy, 2019), with typically sandy, acidic, and nutrient-poor soil. The ground vegetation stratum primarily consists of various sedges, herbs, and grasses, with an upper vegetation stratum of sparsely distributed medium- to tall-shrubs and low trees (Cuff & Brocklehurst, 2011; Liddle et al., 2017). During the non-breeding season, *U. daviesae* spend most of their time buried in the soil and are difficult to survey.

A key characteristic of sandsheet heath is the unique hydrology on which *U. daviesae* relies. The monsoonal Darwin climate has distinct wet and dry seasons, and wet season rainfall and

higher water tables inundate these heaths for several months (EcOz Environmental et al., 2013). The distinct topography and underlying geology mean it sustains a shallow sheet-like surface flow (depths of up to 6 cm; EcOz Environmental et al., 2013). Changes in hydrology due to landform disruption (including stream formation and water ponding) alter ecological processes, resulting in different botanical and hydrological attributes that make the habitat unsuitable for *U. daviesae* (EcOz Environmental et al., 2013). *Uperoleia daviesae* emerges from the soil during the wet season, which lasts from about December to March each year (Clancy, 2019; Reynolds & Grattidge, 2012). Males call from "debil-debil", small, raised mounds created by earthworms at the base of tussock grasses, which provide small islands in the sheeting water (Reynolds & Grattidge, 2012). During mating, the female lays each egg individually on the stem of plants in the inundated ground stratum. Although the number of clutches and eggs per clutch is unknown, estimates from other conspecific species suggest that females lay one clutch per year, with fewer than 50 eggs per clutch (Anstis, 2018).

Human activity is the primary threat to *U. daviesae*, including extractive sand mining, modifications to natural systems, and residential development and expansion (Department of Environment, Parks and Water Security, 2021). There is a significant conflict between the ecological and economic values of sandsheet heath (Northern Territory Environment Protection Authority, 2015). The sand from the sandsheet heaths where *U. daviesae* occurs is the primary target area for extractive sand mining operations (Department of Land Resource Management, 2013; Doyle, 2001). Approximately 21% of the potential habitat (460 ha) for *U. daviesae* has been undermined by sand mining operations over several decades (Department of Land Resource Management, 2013). Extraction sand mining activities have been subject to increased regulation over the past decade, with additional assessments of their potential impacts on habitat (EcOz Environmental et al., 2013; Northern Territory Environment Protection Authority, 2015). However, research is needed into the environmental requirements of threatened species inhabiting this habitat to guide environmental impact assessments, land-use planning and regulatory settings.

Over the last few decades, the size and population of the greater Darwin area have grown steadily (Kavaarpuo et al., 2022). Urban encroachment on the habitats of the *U. daviesae* is increasing (Department of Land Resource Management, 2013; Threatened Species Scientific Committee, 2021). Approximately 30% of the *U. daviesae*'s potential breeding habitat in the Darwin peri-urban area has already been highly modified (Threatened Species Scientific

Committee, 2021). The development and construction will affect the natural flow of surface water and groundwater, altering the hydrological system in the remaining natural areas (Hibbs & Sharp, 2012; Sharp, 2010), and affect the pattern and extent of inundation of the sandsheet heath, thereby influencing the breeding success of the *U. daviesae*.

Although we have some understanding of the species' ecology, such as dispersal range, habitat types, and breeding behaviour (Reynolds & Grattidge, 2012; Young et al., 2005), knowledge of the distribution of genetic diversity and connectivity between fragments is poor, limiting the capacity of policymakers and regulators to make well-informed decisions for the conservation and management of this significant ecosystem. Here, we generate the first genome assembly for *U. daviesae* and use genome-wide data to assess dispersal and gene flow patterns, quantify genetic diversity and inbreeding, and evaluate the effects of fragmentation on population structure. We also identify populations of high conservation value based on their contribution to overall genetic variation. These findings provide a foundation for evidence-based management of *U. daviesae* and contribute to broader discussions on the conservation of short-range endemic species in increasingly urbanised landscapes.

Materials and Methods

Sample collection

Fieldwork for tissue sample collection was conducted in January/February 2023. Toe clips from 129 individuals were collected from 15 locations (Figure 1, Table S1). We attempted to collect 10 samples at each location. The tissue samples were stored in 80 % ethanol at 4 °C before DNA extraction and SNP dataset generation (see below). A single male frog was collected for genome sequencing from site UD14 on 17 February 2023 (Figure 1). This individual was euthanised, and the heart, kidney, gonads and liver were collected for genome assembly and annotation. The heart was stored at -20 °C for genome sequencing. The gonads, kidney, and liver were stored in RNAlater® at -80 °C for transcriptomic analyses. All collections were undertaken under approval from Charles Darwin University Animal Ethics Committee (A21021) and a permit from NT Parks and Wildlife (Permit no. 71982).

Genome assembly and annotation

The DNA extraction for genome assembly and annotation was undertaken by the Australian Genome Research Facility (AGRF, Brisbane, Australia) using the Monarch® HMW extraction kit (New England BioLabs, Frankfurt) according to the manufacturer's instructions. Total RNA was extracted by the Garvan Medical Research Institute (Sydney, Australia) using the RNeasy® RNA extraction kit (Qiagen, Germany). RNA from the kidney and gonads was pooled into a single sample for sequencing.

Detailed information on our genome assembly and annotation methods is provided in the Supporting Information. Briefly, the reference sample was sequenced at the AGRF (Brisbane, Australia) using *PacBio HiFi* library preparation and *Revio* sequencing on two 25M SMRT Cells. Total RNA from the liver and the pooled kidney/gonads was sequenced at the Ramaciotti Centre for Comparative Genomics using an *Illumina* stranded mRNA prep. The RNA was sequenced on a NovaSeq X Plus 10B flow cell using 150 bp paired-end reads, resulting in 69 million reads per sample (kidney/gonads) and 81 million reads per sample (liver). To generate the reference genome for the *U. daviesae*, we used the Vertebrate Genome Project (*VGP)* assembly pipeline on the Galaxy Australia interface (Batut et al., 2018; Hiltemann et al., 2023; Lariviere et al., 2024; The Galaxy Community, 2022). The mitochondrial genome was assembled with *MitoHiFi v2* (Uliano-Silva et al., 2023) and visualised in *Proksee* (Grant et al., 2023). Final genome quality assessment was performed using *BUSCO v5.5.0* (Simao et al., 2015) and *Merqury v1.3* (Rhie et al., 2020). Repeat elements were identified and masked with *RepeatModeler v2.0.4* and *RepeatMasker v4.1.5* (Flynn et al., 2020). We also generated a reference-aligned transcriptome assembly to assist with genome annotation using workflows on the Galaxy Australia interface.

SNP data generation

Tissue samples were submitted to Diversity Arrays Technology Pty Ltd (Canberra) for commercial DNA extraction, library preparation, and *DArT-seq* 1.0 high-density sequencing, following the proprietary methods (Kilian et al., 2012; Sansaloni et al., 2011). In brief, the *PstI-SphI* restriction enzyme combination was used for DNA digestion, and all subsequent laboratory steps followed the protocol described in Gruber et al.'s paper (Gruber et al., 2018). Two rounds of high-level sequencing were conducted to increase read depth, given the large genome size of the species, and to enable read-depth-dependent analyses (see below). We

identified no differences in population structure between sequencing runs (data not shown) and therefore combined the individual fastq files using the Linux *cat* command, which was used in all subsequent bioinformatic pipelines.

We used "*process_radtags"* from Stacks v2.6.6 (Catchen et al., 2013; Catchen et al., 2011) to remove barcodes and quality-filter raw reads, then *Trimmomatic v0.39* (Bolger et al., 2014) to trim all reads to 68 bp (see Script S1 for specific parameters). Cleaned reads were aligned to the genome using *BWA-MEM2* (Vasimuddin et al., 2019) with default settings. *Samtools stats* and *Samtools view* (Danecek et al., 2021) were used to inspect and filter the BAM files generated by *BWA-MEM2*. We then used *MultiQC* to summarise the alignment results (Ewels et al., 2016). We used *the "gstacks"* and *"populations"* steps from *Stacks* v2.6.6 (Catchen et al., 2013; Catchen et al., 2011) to call SNPs and export VCF files for further analysis, using different settings and popmap files with varying numbers of individuals per region to meet the assumptions of the analysis methods (Figure S2, detailed in Supporting Information). We excluded site "UD31" because it contained only one individual after filtering.

Population genetic structure analysis

We used a customised parameter setting for the *"populations"* step (detailed in Script S3), followed by *VCFtools v0.1.17* (Danecek et al., 2011) to filter SNPs with a read depth between 10 and 50. Then, VCF files were filtered in R using the *dartR* package (Gruber et al., 2018). The key parameter settings for the pipeline used to filter the SNP data are provided in the supporting information (detailed in Script S3).

We first ran a Principal Coordinates Analysis (PCoA) using the *"gl.pcoa"* function in the *dartR* package to reduce the dimensionality of the multilocus genetic distance data and identify major axes of genetic differentiation among individuals (Figure S1). Then, we used the "*snmf"* (sparse non-negative matrix factorisation) function from the *LEA* package (Frichot & François, 2015) in R to estimate the number of genetic clusters (K; Figure S1). Given differences in sample sizes across sites, we first ran the analysis on all individuals. Then, to assess the effect of biased sampling (Schmidt et al., 2021; Sopniewski & Catullo, 2024), we re-ran the analysis, reducing each site to the same sample size (N = 5). For each analysis, we first used 5% of the SNP data to test K (the number of genetic clusters) from 1 to 6 and determine the optimal tolerance and alpha value settings based on the cross-entropy

criterion (Frichot et al., 2014). Then, using these tolerances and alpha values, we ran the full dataset for Ks ranging from 1 to 6, and repeated the analysis 100 times. The cross-entropy criterion was used to determine the optimal K-value (Frichot et al., 2014).

We then used the *"gl.dist.pop"* function in the *dartR* package to calculate the Euclidean and Nei's distance between regions with all individuals and used the "*mantel*" function in the *VEGAN* package (Dixon, 2003) to conduct the Mantel tests between genetic distances and geographic distances. To estimate the relative migration levels among sites with all individuals, we used the "*divMigrate()*" function from the *diveRsity* package (Keenan et al., 2013) in R. We used the number of migrants per generation (N_M) method (Alcala et al., 2014), with a filter threshold set to 0.4, and 1,000 bootstrap replicates (McLennan et al., 2025). The line width and opacity for showing the direction of gene flow in the figure were scaled based on the N_M value, ranging from 0.40 to 1.00.

Genetic diversity

We used the *poppr* (Kamvar et al., 2014) and *hierfstat* (Goudet, 2005) packages in R to calculate the number of private alleles (P_A) and allelic richness (A_R) for each region. We randomly selected the same number of individuals per site within each region, with 12 individuals in total per region ($N_{per\_site\_Region1} = 6$, $N_{per\_site\_Region2} = 2$, $N_{per\_site\_Region3} = 3$, $N_{per\_site\_Region4} = 6$, $N_{total} = 4 \times 12 = 48$; Figure S1), and ran the analysis 50 times. For heterozygosity, SNP-based estimates can be biased by study design and filtering parameters, such as sample size and missing-data thresholds (Schmidt et al., 2021; Sopniewski & Catullo, 2024). We first utilised the autosomal heterozygosity and nucleotide diversity generated during the "*populations"* step (-R = 1, Figure S1) at both the regional (12 individuals per region, $N_{total} = 48$, 50 iterations; Figure S1) and site levels ($N_{total} = 115$, Figure S1). Then, we used the *JeDi* pipeline (Pavlova et al., 2024) with a dataset containing all individuals after filtering ($N_{total} = 115$; Figure S1) to calculate unbiased genetic diversity metrics at both the regional and site levels (Figure S1), as the number of individuals does not affect the estimates.

We used the *"gl.report.diversity"* function in the *dartR* package to calculate genetic diversity across SNPs, with 12 individuals per region ($N_{total} = 48$) and 50 iterations (Figure S1). The diversity results include three "q-profile" results, each calculated using a different method.

The q = 0 profile was based on allelic richness, the q = 1 profile was based on Shannon information, and the q = 2 profile was based on heterozygosity (Sherwin et al., 2017).

The effective population size (N_e) for a species provides a broad-scale indication of genetic health, resilience to stochastic threats, and overall evolutionary potential (Hoban et al., 2020). We used *NeEstimator* v2 (Do et al., 2014) and selected the jackknife-on-samples method to estimate 95% confidence intervals for N_e at the whole-species and regional levels with all individuals (N_{total} = 115; Figure S1).

Inbreeding and genetic drift

To estimate inbreeding levels at each region and site (N_{total} = 115), we used the F_h statistic (Kardos et al., 2015) with the "—het" function in *PLINK2* (Purcell et al., 2007). We used *TreeMix* version 1.13 (Pickrell & Pritchard, 2012) to investigate population relationships, assess genetic drift, and identify potential migration events (Figure S1). We further filtered the SNP dataset using a 100% call rate for loci and individuals to ensure no missing data. We ran *TreeMix* with all individuals (N_{total} = 115) 10 times, varying the number of migrations (m) from 1 to 5, with 500 bootstrap replicates. Then, we selected the optimum number of migrations using the *OptM* package (Fitak, 2021) in R. We then ran *TreeMix* 50 times (500 bootstrap replicates) with the chosen number of migrations (m = 1) and m = 0 as the null model. The best runs were selected based on the highest likelihood (Zecca et al., 2020). The consensus tree and bootstrap values were generated using the *BITE V2* package in R (Milanesi et al., 2017).

Prioritising sites in conservation

Firstly, we used the "*QDiver*" function (Smouse et al., 2017) in *GenAlEx* v6.51 b2 software (Peakall & Smouse, 2006, 2012) to investigate the partitioning of allelic diversity within and among sites and regions (Figure S1). Since the QDiver function is sensitive to missing data, we used the *"gl.filter.callrate"* function in the *dartR* package and set the callrate to one to remove all missing data.

Secondly, we used *Metapop2* v2.5.3 (Kilian et al., 2012) to investigate the contributions of each site to allelic and gene diversity (Figure S1). We first calculated the contributions of each site to the within-site (A_S, H_S) and among-site (D_A, D_G) allelic and gene diversity, as

well as the total allelic and gene diversity (A_T, H_T). Then, we calculated the optimal contributions of each site to allelic (K) and gene (H) diversity within a gene pool comprising 1000 randomly selected individuals. We randomly selected five individuals at each site, repeated the analysis 50 times, and averaged the results across all runs. Sites with positive contributions represent genetically distinctive sites or ones with unique or rare alleles, thereby boosting overall diversity measures when included in the hypothesised population. Negative contributions indicate sites where their inclusion in the hypothesised population reduces overall genetic diversity metrics, typically because these sites contain alleles commonly shared with others, rather than unique or rare genetic variants.

Thirdly, we used *Marxan* (Watts et al., 2009) to identify the minimum set of populations required to represent the maximum proportion of genome-wide neutral variation across the species (Figure S1). Each SNP variant (reference or alternate) was treated as a unit of genetic diversity to be represented, and each population was treated as a conservation unit. Because SNPs are biallelic, this metric reflects the representation of population-specific and shared variants rather than multiallelic diversity sensu microsatellites. We set the conservation cost of each site to equal and identified the optimal combination of sites for the maximum covered percentage of alleles within the species using the *prioritizr* (Hanson et al., 2024) and *lpsymphony* (Kim, 2019) packages in R. Each allele was considered as a feature to be conserved, and each site as a conserving unit. We randomly selected five individuals per site and ran the analysis with 100 iterations. We recorded the combinations of sites with the highest allele coverage at different numbers of sites being conserved, as well as the number of each site that had been selected under each different site size. This approach identifies populations that contribute unique or irreplaceable genomic variation, while reducing redundancy from populations that share the same allelic states.

Results

Genome assembly

Assembly with Hifiasm resulted in primary and alternate haplotype assemblies of 6.08 and 3.02 Gbp in size, with 7,390 and 19,207 contigs, respectively (Table S2). Analysis with BUSCO showed that the primary haplotype was highly complete (91.4% complete BUSCOs), but with 17.3% of BUSCOs duplicated. In contrast, the alternate assembly was much less complete (52.9% complete BUSCOs), indicating that purging of the primary

assembly was needed (Table S2). After purging, the primary and alternate assemblies were 4.77 and 3.68 Gbp with 3,327 and 14,048 contigs, respectively, with the mitochondrial genome determined to be on contig ptg003430 and 19,416 bases in length, containing 13 genes and 22 tRNAs (Table S2, Figure S2). Analysis of the purged assemblies showed a still highly complete primary assembly (90.7% complete BUSCOs) and increased completeness of the alternate assembly (68.6% complete BUSCOs), with a reduction in the duplicated BUSCOs in the primary assembly (2.0% duplicated BUSCOs). Additionally, Merqury analysis indicated the primary assembly to be 91.72% complete. Repeat masking of the genome masked 63% of bases, which is within the expected range for anuran species (Table S3).

Population structure

For population structure analyses, retaining equal samples at each site (N = 5), we obtained 28,878 SNPs from 70 individuals. The optimal number of genetic clusters (K) was one (Figure S3a), suggesting a history of gene flow across the region. However, when assessing additional values of K to identify subpopulations (Evanno et al., 2005; Janes et al., 2017), further clusters representing minor allele frequency shifts were identified (Figure 2c & 2d). These clusters comprised: 1) the two sites closest to Darwin city, 2) two sites to the far southwest of Darwin, and 3) all remaining sites in the distribution. Despite aggregating as a single cluster, this last broad cluster had minor but clear divergence, distinguishing Weddell from the Howard River Catchment (Figure 2b; Region 3 versus Region 2). The PCoA (Figure 2a) also identified the same four geographically delineated genetic clusters (hereafter referred to as 'regions') within the species. When analysed with differing sample sizes (the full dataset with 28,882 SNPs from 115 individuals), the optimal number of genetic clusters was three (Figure S3b), consistent with the regions outlined above.

Pairwise Euclidean and Nei's genetic distances (Table S4) were lowest between Regions 3 and 4 and highest between Regions 1 and 2. There was a modular genetic distance between Regions 2 and 3. This pattern is consistent with the spatial configuration of the regions and the SNMF clustering results. Moreover, a significant positive correlation between genetic and geographic distances across sites further supports the isolation-by-distance pattern (Mantel tests: Euclidean $r = 0.466$, $p = 0.0001$; Nei's $r = 0.451$, $p = 0.0002$).

Our migration estimates indicated a medium-to-high level of bidirectional gene flow among most sites in Region 3 (Figure 3). A moderate level of bidirectional gene flow was also detected from UD15 into Region 3. Other sites in Region 2 showed only low levels of bidirectional or directional gene flow with sites in Region 3. Within Region 2, most bidirectional gene flows among sites were low. In Region 1, there was no gene flow between the two sites, and only low levels of directional gene flow from UD09 to UD21B and from UD15 in Region 2. For Region 4, we detected no gene flow within or between regions.

Genetic diversity at the regional and site levels

We retained 28,741 SNPs across 48 individuals (12 per region) in all 50 replicate runs. Across the 50 replicate runs, the number of private alleles, allelic richness, and q-profile diversity per region were highly consistent, with only limited variation between runs, and the mean values were highly representative (Figures S4, S5, S6).

At the regional level, genetic diversity varied among the four regions (Table 1). Region 1 exhibited the highest number of private alleles (P_A = 284), followed by Region 4 (P_A = 120). Regions 3 and 2 had much lower counts (26 and 10, respectively). Allelic richness (A_R) was significantly higher in Regions 1 and 2 (A_R = 1.844 and 1.829, respectively, Kruskal–Wallis test $p << 0.001$). Unbiased Heterozygosity increased from Region 1 (Unb. H_O = 0.00206) to Region 4 (Unb. H_O = 0.00246, Table 1, Kruskal–Wallis test $p << 0.001$). The unbiased nucleotide diversity (Unb. Pi) was lowest in Region 1 (Unb. Pi = 0.00203), and highest in Region 3 (Unb. Pi = 0.00242, Table 1), but the differences were not significant (p = 0.22). There were differences in whether Regions 3 or 4 had the highest observed heterozygosity and nucleotide diversity between the autosomal results and the unbiased JeDi results. But in both methods, Region 1 had the lowest observed heterozygosity and nucleotide diversity, followed by Region 2.

The Q-profile analysis (Table 1), based on rare alleles, identified that Region 1 exhibited the highest diversity and Region 3 the lowest at either q=0 (rare alleles), q=1 (Shannon information), and q=2 (heterozygosity, Kruskal-Wallis test $p << 0.001$, Table 1). Within-region analyses indicated substantial variation among individual sites (Table S5). In Region 2, UD15 (m_0Da = 0.844, Table S5) had the highest allele richness diversity, whereas UD18 (m_0Da = 0.633, Table S5) had the lowest. In Region 3, UD21B (m_0Da = 0.875, Table S5)

and UD29 (m_0Da = 0.870, Table S5) showed high allelic diversity, reinforcing their importance in maintaining regional genetic diversity.

Effective population size (N_e) varied significantly across the four regions. Overall, the N_e estimate for the species was 116.1 (90.8–155.2). Region 3 had the highest N_e, estimated at 242 (126.0–1666.0), indicating a historically larger and more stable population. Region 2 also exhibited a relatively high N_e of 98.7 (66.5–176.5). Regions 1 (23.6; 18.2–32.2) and 4 (11.4; 8.3–16.4) showed much lower N_e values.

Genetic diversity significantly varied between sites within each region and among all sites (Table S6). Site UD10 had the highest unbiased observed heterozygosity, followed by sites UD01 and UD01B, while site UD14 had the lowest, followed by site UD15. For unbiased nucleotide diversity, site UD10 had the highest value, while UD14 had the lowest. In Region 1, UD09 had higher diversity than UD14. In Region 2, UD15 had the highest unbiased nucleotide diversity, while UD18 had the lowest. In Region 3, site UD05B had the highest diversity, while UD28 had the lowest. In Region 4, UD10 had higher genetic diversity than UD04.

Inbreeding and genetic drift

Across regions, F_h values showed clear spatial structure (Table 1), with the highest inbreeding occurring in Region 1 ($F_h = 0.1719 \pm 0.0078$, Table 1). Both sites in this region (UD09 and UD14) exhibited similarly elevated values, indicating consistently high inbreeding across the entire region. Region 2 showed moderate overall inbreeding (0.1095 ± 0.0079; Table 1), but substantial among-site heterogeneity. UD22, a geographic outlier, had markedly elevated F_h (0.1793 ± 0.0256; Table S6), exceeding those of all other Region 2 sites and approaching levels observed in Region 1. In contrast, Region 3 had the lowest overall inbreeding (0.0538 ± 0.0101, Table 1), with most sites displaying low to moderate values (e.g., UD21B = 0.0318; UD29 = 0.0276) consistent with its relatively intact habitat. Region 4 was characterised by strong site-level divergence, with one highly inbred site (UD04, $F_h = 0.1724 \pm 0.0118$, Table S6) and one low-inbreeding site (UD10, 0.0350 ± 0.0131, Table S6). Overall, all regions exhibited elevated F_h relative to Region 3, consistent with increases in inbreeding associated with anthropogenic habitat modification.

Branch lengths from *Treemix* were generally short and often poorly supported, and the topology of the *Treemix* result trees remained largely unchanged with additional migration events (Figure 4). Together, these data suggested that sites were generally well connected, with a topology concordant with our population-structure results. Large drift parameters were identified in the two isolated regions (Regions 1 and 4), while the two main regions (Regions 2 and 3) had shorter branch lengths. One migration event was identified, between UD10 and the relatively close UD29. Residuals from the m=0 and m=1 models highlighted populations that might not fit a strict tree model, with strongly positive residuals between some pairs (e.g., UD10 and UD28) suggesting they could be more closely related than the consensus tree indicates, again suggesting relatively recent connectivity.

High-priority sites for genetic conservation

The QDiver analysis revealed high total allelic diversity across the species (γ = 0.988; Table S7), suggesting that the species retain substantial genetic variation. Allelic diversity among regions was moderate (δ = 0.745, Table S7), indicating some genetic structuring at the region level, although this structure was not particularly strong, suggesting either contemporary gene flow or recent connectivity. Within-region allelic diversity was consistently high (σ range: 0.951-0.955, Table S7), with no significant difference among regions (Bartlett's test, p = 0.999, Table S7), implying that each region individually maintains considerable genetic variability. Allelic differentiation among sites within regions (β) varied from moderate to low (range: 0.500–0.819, Table S7), with Region 2 exhibiting higher differentiation among sites compared to other regions, although the difference was not significant (p = 0.142, Table S7), possibly due to limited recent connectivity at finer scales. Within-site allelic diversity relative to their region was consistently high (α range: 0.880–0.945, Table S7), and no significant variation was observed among sites ($p \geq 0.399$, Table S7).

Based on the Metapop2 analysis results, the rank of each site's contribution to total diversity was similar for allelic and genetic diversity (Figure 5a). Regions 3 and 4 contributed positively to the total allelic and genetic diversity (A_T, H_T), while Regions 1 and 2 contributed negatively. Only Region 3 had a positive contribution to the allelic and genetic diversity within sites (alpha diversity, A_S, H_S). Region 1 had the largest negative contributions. Regions 1 and 4 made significant positive contributions to allelic and genetic

diversity among sites (beta diversity, D_A, D_G). In an optimal sample of 1000 individuals (Figure 5b), all sites would contribute similarly to the number of alleles (K), but Regions 3 and 4 contribute the most to heterozygosity (H).

Our *MARXAN* results indicated that conserving either UD21B or UD29 sites (which are ~1 km apart in Region 3) was the most effective single site for conserving and capturing overall genomic variant richness, accounting for ~87% of all SNP variants detected (Figure S7, Table S8). Conserving at least two sites increased total variant representation to more than 90%. The number of sites required for conservation converged at four sites ($p < 0.01$, asymptotic regression). At least ten sites must be conserved to preserve all detected SNP variants within the species.

Discussion

Historical Connectivity and Contemporary Fragmentation

Our population genomic analysis of *U. daviesae* reveals an evolutionary history marked by historical connectivity followed by recent fragmentation at the edge of the species' range, as well as an older history of moderate isolation by catchment. Population structure and migration analyses (Figures 2, 3) found ongoing connectivity between habitat fragments in the relatively intact Region 3, and reduced gene flow in anthropogenically disturbed areas. This supports a scenario in which the species existed as a relatively continuous population from the Howard River catchment down to the Elizabeth River catchment, with gene flow occurring across the range. However, the detection of region-specific shifts in allele frequencies indicates that this continuity has been eroded by historical and modern processes (Brewer et al., 2020; Fountain et al., 2016; Fusco et al., 2021). *TreeMix* analyses further support this pattern, showing relatively short branch lengths for the central populations (Regions 2 and 3), consistent with recent gene flow preventing genetic drift. In contrast, the peripheral regions (Regions 1 and 4) exhibited longer branch lengths and increased signals of genetic drift, reflecting isolation (Peter, 2016; Pickrell & Pritchard, 2012). These findings are consistent with a metapopulation structure shaped by habitat discontinuities and anthropogenic fragmentation, particularly in the Darwin peri-urban region.

The barrier between Regions 2 and 3 was aligns with different catchments and the extensive peri-urban suburbs stretching from Howard Springs in the north to Humpty Doo and Bees

Creek in the south, and eastward to Girraween and Herbert (Figure 1). This peri-urban development area expanded rapidly after 1950 (Government, 2025; Taylor, 2011), removing significant areas of sandsheet heath habitat and increasing the fragmentation of the remaining habitat. Barriers between Regions 1 and 2, apart from different catchments, include peri-urban development, defence development, and sand mining at Howard Spring, north of Gunn Point Road, on a recent time scale, which may have further isolated the two regions. The separation of regions 3 and 4 is less clear. Specifically, while all sites in Region 3 lie within the Elizabeth Creek Catchment, the Region 4 grouping includes one Elizabeth Creek catchment site in Weddell (UD04) and a single site 13 km to the southeast in the Blackmore Creek catchment. Further genetic sampling in this region may be required to understand barriers to gene flow with Region 3.

While the species likely evolved with some degree of natural fragmentation due to habitat heterogeneity (Department of Land Resource Management, 2013; Liddle et al., 2017), our results suggest that contemporary landscape change, especially urban expansion and land clearing in the Darwin region, has exacerbated genetic subdivision and reduced connectivity among sites. This is particularly evident in Region 2, where several sites (e.g., UD15, UD18) show signs of increased genetic differentiation and low within-site diversity (see below) relative to Region 3, which is relatively undisturbed.

Regional Genetic Diversity and Signatures of Inbreeding

As expected, given landscape fragmentation, genetic diversity varied markedly among the four identified regions, particularly between core and peripheral populations. Region 3, situated in a relatively undisturbed area of the Elizabeth River catchment, exhibited the highest unbiased nucleotide diversity (Unb. Pi = 0.00242). This suggests that Region 3 has retained a substantial portion of the species' genetic variation and may represent a demographic and evolutionary stronghold for *U. daviesae*. In contrast, Region 2, located in a relatively disturbed area of the Howard River catchment, has the largest geographic extent but shows high among-site differentiation, most likely reflecting the effects of recent anthropogenic fragmentation. Genetic diversity in this region was moderate (Unb. Pi = 0.00216). Still, the increase in inbreeding and reduced within-site diversity in some sites relative to Region 3 (e.g., UD18) suggest that the negative genetic effects of urbanisation are already manifesting.

Regions 1 and 4 have been isolated from Regions 2 and 3 due to limited migration. Region 1 showed the lowest genetic diversity among the four regions (Unb. Pi = 0.00203) and the highest inbreeding (Fh = 0.1719). Site-level inbreeding for the two sites in this region was also high (Fh = 0.1712 and 0.1727), suggesting that isolation has already had genetic consequences. In contrast, Region 4 had relatively high region-level genetic diversity (Unb. Pi = 0.00232) and low inbreeding (Fh = 0.1037), indicating that it harbours unique diversity compared to the main regions. However, differences in site-level genetic diversity and inbreeding between UD04 and UD10 suggest that the genetic consequences of urbanisation vary among sites within regions.

Taken together, these patterns reveal that the species has been adversely affected by human activity, showing regional variation in genetic diversity, genetic drift, and overall inbreeding (Almeida-Rocha et al., 2020; DiBattista, 2008). These differences have important implications for developing conservation actions that account for the genetic resilience and vulnerability of each region.

Conservation Management of Uperoleia daviesae

The integration of genetic data into conservation planning for *U. daviesae* highlights both the challenges and opportunities of managing short-range endemic species threatened by urban development. Despite the species retaining high total allelic diversity (γ = 0.988; Table 6), the contributions of individual regions and populations to this diversity were unequal. Our QDiver and Metapop2 analyses identified that Regions 3 and 4 contribute disproportionately to total genetic and allelic diversity, especially in terms of heterozygosity and private alleles. But Region 4 also had a declining population (UD04), suggesting it may need more conservation interventions. The dispersal pattern we uncovered in *U. daviesae* closely mirrors that reported for other short-range endemic species inhabiting fragmented peri-urban heath and woodland systems worldwide (Hamer & McDonnell, 2008; Munshi-South et al., 2013). In these taxa, genetic diversity is typically concentrated within a limited number of "source" patches, whereas small, isolated sites accumulate unique alleles (Hantzschmann et al., 2021; Pröhl et al., 2021).

Because urbanisation in peri-urban landscapes is effectively irreversible, large-scale habitat restoration or re-establishment is unrealistic. A more pragmatic approach is to maintain intraspecific genetic diversity under current and foreseeable conditions while maximising opportunities for gene flow despite further urban encroachment. We recommend treating this species as a metapopulation and implementing tailored conservation actions in each region to sustain its gene flow and genetic diversity.

The significant within-region genetic variation detected by our conservation-priority analyses (σ = 0.951–0.955; Table 6), together with the relatively high within-site contributions (α = 0.880–0.945; Table 6), indicates that many *U. daviesae* sites individually capture most of the diversity present in their respective regions. This finding is encouraging for site-based conservation because it implies that protecting representative sites within each region can effectively preserve overall diversity.

Region 3 contains both the greatest genetic and most of the allelic diversity within the species. It is essential to recognise and maintain the landscape processes that support connectivity, which have likely enabled Region 3 to retain higher diversity than Region 2. For Region 2, actions should be taken to ensure these sites are not lost through inbreeding and drift. Translocation should be considered to enhance genetic connectivity, which appears to have been reduced by anthropogenic change (Ewen et al., 2012; Sheean et al., 2012). Furthermore, owing to its relatively high level of inbreeding, which distinguishes it from other Region 2 sites, UD22 soon requires conservation intervention.
In Region 4, the two sites differed substantially. Site UD04 has little unique allelic diversity and high inbreeding, thus translocations from Region 3 should be considered. In contrast to UD04, site UD10 had an exceptionally high allelic contribution to the species and low inbreeding, indicating it is a high-priority site for protection. Identifying and protecting other patches of sandsheet heath occupied by *U. daviesae* in nearby areas is also a high priority. Given its genetic significance, site UD10 is also a high priority for regular population monitoring to detect any future declines before they have genetic consequences.

Region 1 is isolated from the main region, with limited migration, and it shows the highest regional inbreeding among the four regions, as well as high inbreeding at both sites. Given its proximity to Palmerston City and Howard Springs, areas expected to face the earliest future urbanisation pressures in the Darwin peri-urban area, urgent management interventions are

warranted. Due to the unique allelic diversity, genetic supplementation may be warranted for all Region 1 sites and UD10 in Region 4 (Frankham, 2015). Any translocation into these areas should follow a low-intensity, multi-year introduction strategy that contributes a small proportion of migrants over several generations (Hedrick and Fredrickson, 2010). This approach reduces inbreeding while minimising the risk of overwhelming local unique allele frequencies.

While our findings provide several actionable insights for the conservation of this species, important knowledge gaps remain. We advocate that future landscape-genetic studies delineate the specific barriers across the four regions we identified and map the corridors that facilitate gene flow. Additionally, species persistence likely depends critically on the hydrological conditions within breeding habitats. The key hydrological features of *U. daviesae* habitats and their influence on reproduction remain insufficiently understood and warrant targeted study. This requires not only research from ecology and behavioural science, but also integration with interdisciplinary fields such as biogeology and hydrology.

Conclusion

Uperoleia daviesae exemplifies how short-range endemics with naturally fragmented distributions are particularly susceptible to genetic erosion from the demographic consequences of urban development. Conservation strategies must consider both central and peripheral populations to maintain genetic diversity and ensure long-term persistence, with special emphasis on protecting populations in Regions 3 and 4, which contribute disproportionately to the species' overall genetic health.

Acknowledgements

We would like to acknowledge the contributions of the Threatened Species Initiative and the Australian Amphibian and Reptile Genome Consortia to the generation of the data used in this publication. Both programs are supported by funding from Bioplatforms Australia, which is enabled by the Commonwealth Government's National Collaborative Research Infrastructure Strategy (NCRIS). TSI is a partnership between Bioplatforms Australia and the University of Sydney, with support from the Australian Government Department of Climate Change, Energy, the Environment and Water; WA Department of Biodiversity, Conservation & Attractions; and Amazon Web Services. LWS fellowship is supported by the Australian

Research Council (CE200100012). We thank Diarmid Brown, Anna Miller, Laura Ruykys, and Tony Griffiths for their assistance in the field. Brydie Hill helped develop the AusARG funding proposal.

References

Alcala, N., Goudet, J., & Vuilleumier, S. (2014). On the transition of genetic differentiation from isolation to panmixia: What we can learn from GST and D. *Theoretical Population Biology*, *93*, 75-84. https://doi.org/10.1016/j.tpb.2014.02.003

Allan, J. R., Watson, J. E. M., Di Marco, M., O'Bryan, C. J., Possingham, H. P., Atkinson, S. C., & Venter, O. (2019). Hotspots of human impact on threatened terrestrial vertebrates. *PLOS Biology*, *17*(3), e3000158. https://doi.org/10.1371/journal.pbio.3000158

Almeida-Rocha, J. M., Soares, L. A. S. S., Andrade, E. R., Gaiotto, F. A., & Cazetta, E. (2020). The impact of anthropogenic disturbances on the genetic diversity of terrestrial species: A global meta-analysis. *Molecular Ecology*, *29*(24), 4812-4822. https://doi.org/10.1111/mec.15688

Anstis, M. (2018). *Tadpoles and Frogs of Australia*. New Holland Publishers. https://books.google.com/books?id=NryBswEACAAJ

Batut, B., Hiltemann, S., Bagnacani, A., Baker, D., Bhardwaj, V., Blank, C., Bretaudeau, A., Brillet-Guéguen, L., Čech, M., Chilton, J., Clements, D., Doppelt-Azeroual, O., Erxleben, A., Freeberg, M. A., Gladman, S., Hoogstrate, Y., Hotz, H.-R., Houwaart, T., Jagtap, P.,…Grüning, B. (2018). Community-Driven Data Analysis Training for Biology. *Cell Systems*, *6*(6), 752-758.e751. https://doi.org/https://doi.org/10.1016/j.cels.2018.05.012

Blaustein, A. R., Han, B. A., Relyea, R. A., Johnson, P. T. J., Buck, J. C., Gervasi, S. S., & Kats, L. B. (2011). The complexity of amphibian population declines: understanding the role of cofactors in driving amphibian losses. *Annals of the New York Academy of Sciences*, *1223*(1), 108-119. https://doi.org/10.1111/j.1749-6632.2010.05909.x

Blaustein, A. R., & Kiesecker, J. M. (2002). Complexity in conservation: lessons from the global decline of amphibian populations. *Ecology Letters*, *5*(4), 597-608. https://doi.org/10.1046/j.1461-0248.2002.00352.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Brewer, V. N., Lane, S. J., Sewall, K. B., & Mabry, K. E. (2020). Effects of low-density urbanization on genetic structure in the Song Sparrow. *PLOS ONE*, *15*(6), e0234008. https://doi.org/10.1371/journal.pone.0234008

Carrasco, G. H., de Souza, M. B., & de Souza Santos, L. R. (2021). Effect of multiple stressors and population decline of frogs. *Environmental Science and Pollution Research*, *28*(42), 59519-59527. https://doi.org/10.1007/s11356-021-16247-6

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. https://doi.org/10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 Genes|Genomes|Genetics*, *1*(3), 171-182. https://doi.org/10.1534/g3.111.000240

Clancy, M. (2019). *Modelling the distribution of the Howard Springs Toadlet (Uperoleia daviesae) Unpublished Honours Thesis*. The School of BioSciences, University of Melbourne, Parkville, Victoria.

Coleman, D. W., Wood, R. J., & Healey, M. (2024). Frogs and flows: Using life-history traits and a systematic review to establish water-dependent functional groups for stream frogs in New South Wales, Australia. *Ecohydrology*, *17*(3), e2643. https://doi.org/10.1002/eco.2643

Cuff, N., & Brocklehurst, P. (2011). *Vegetation Mapping and Floristic Survey of the Proposed New Township of Weddell, Northern Territory*.

Cushman, S. A. (2006). Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation*, *128*(2), 231-240. https://doi.org/10.1016/j.biocon.2005.09.031

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Department of Environment, Parks and Water Security. (2021). *Evidence on Listing Eligibility and Conservation Actions 2021: Uperoleia daviesae Nomination*. https://www.dcceew.gov.au/sites/default/files/env/consultations/fabae2b7-6a50-4f47-8ec6-759f3d5f6b32/files/cam-assessment-uperoleia-daviesae.pdf

Department of Lands, Planning and Environment. (2014). Sandsheet heath vegetation in the Darwin region [Dataset]. Northern Territory Government. https://www.ntlis.nt.gov.au/metadata/export_data?type=html&metadata_id=0323B08857C95C65E050CD9B2144212B

Department of Land Resource Management. (2013). *Survey of the breeding populations of Uperoleia daviesae in the proposed township of Weddell*. Flora and Fauna Division, Department of Land Resource Management, Palmerston, Northern Territory.

Dibattista, J. D. (2008). Patterns of genetic variation in anthropogenically impacted populations. *Conservation Genetics*, *9*(1), 141-156. https://doi.org/10.1007/s10592-007-9317-z

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(6), 927-930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209-214. https://doi.org/10.1111/1755-0998.12157

Doyle, N. (2001). *Extractive minerals within the outer Darwin area Extractive minerals within the Outer Darwin area: Northern Territory Geological Survey Report 14*. Northern Territory Geological Survey. https://geoscience.nt.gov.au/gemis/ntgsjspui/bitstream/1/81550/3/NTGSRep14.pdf

EcOz Environmental, S., Grattidge, A., & Richardson, J. (2013). *Assessment Framework for Rehabilitation Sand Mining: Howard Sand Plains Site of Conservation Significance*. https://www.greeningaustralia.org.au/wp-content/uploads/2017/11/Howard-Sands-Assessment-Framework-for-Rehabilitation.pdf

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

Ewen, J., Armstrong, D., Parker, K., & Seddon, P. (2012). *Reintroduction Biology: Integrating Science and Management*. https://doi.org/10.1002/9781444355833

Fitak, R. R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1). https://doi.org/10.1093/biomethods/bpab017

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451-9457. https://doi.org/https://doi.org/10.1073/pnas.1921046117

Fountain, T., Nieminen, M., Sirén, J., Wong, S. C., Lehtonen, R., & Hanski, I. (2016). Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences*, *113*(10), 2678-2683. https://doi.org/10.1073/pnas.1600951113

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925-929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973-983. https://doi.org/10.1534/genetics.113.160572

Fusco, N. A., Pehek, E., & Munshi-South, J. (2021). Urbanization reduces gene flow but not genetic diversity of stream salamander populations in the New York City metropolitan area. *Evolutionary applications*, *14*(1), 99-116. https://doi.org/10.1111/eva.13025

Gallant, A. L., Klaver, R. W., Casper, G. S., & Lannoo, M. J. (2007). Global Rates of Habitat Loss and Implications for Amphibian Conservation. *Copeia*, *2007*(4), 967-979, 913. https://doi.org/10.1643/0045-8511(2007)7[967:GROHLA]2.0.CO;2

Geyle, H. M., Hoskin, C. J., Bower, D. S., Catullo, R., Clulow, S., Driessen, M., Daniels, K., Garnett, S. T., Gilbert, D., Heard, G. W., Hero, J.-M., Hines, H. B., Hoffmann, E. P., Hollis, G., Hunter, D. A., Lemckert, F., Mahony, M., Marantelli, G., McDonald, K. R.,…Gillespie, G. R. (2022). Red hot frogs: identifying the Australian frogs most at risk of extinction. *Pacific Conservation Biology*, *28*(3), 211-223. https://doi.org/10.1071/PC21019

Goudet, J. (2005). hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184-186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Government, N. T. (2025). *Northern Territory Government Gazette: No. S40-25*. Darwin, NT: Northern Territory Government Retrieved from https://hdl.handle.net/10070/1002784

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484-w492. https://doi.org/https://doi.org/10.1093/nar/gkad326

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*(3), 691-699. https://doi.org/10.1111/1755-0998.12745

Hamer, A. J., & McDonnell, M. J. (2008). Amphibian ecology and conservation in the urbanising world: A review. *Biological Conservation*, *141*(10), 2432-2449. https://doi.org/10.1016/j.biocon.2008.07.020

Hanson, J. O., Schuster, R., Strimas-Mackey, M., Morrell, N., Edwards, B. P. M., Arcese, P., Bennett, J. R., & Possingham, H. P. (2024). Systematic conservation prioritization with the prioritizr R package. *Conservation Biology*, e14376. https://doi.org/10.1111/cobi.14376

Hantzschmann, A. M., Sinsch, U., Göttlicher, C., & Pröhl, H. (2021). Conservation genetics of yellow-bellied toads (Bombina variegata): a matter of geographical scale and isolation. *Conservation Genetics*, *22*(1), 83-96. https://doi.org/10.1007/s10592-020-01320-3

Hazell, D. (2003). Frog ecology in modified Australian landscapes: a review. *Wildlife Research*, *30*(3), 193-205. https://doi.org/10.1071/WR02075

Hibbs, B. J., & Sharp, J. M. (2012). Hydrogeological Impacts of Urbanization. *Environmental & Engineering Geoscience*, *18*(1), 3-24. https://doi.org/10.2113/gseegeosci.18.1.3

Hiltemann, S., Rasche, H., Gladman, S., Hotz, H.-R., Larivière, D., Blankenberg, D., Jagtap, P. D., Wollmann, T., Bretaudeau, A., Goué, N., Griffin, T. J., Royaux, C., Le Bras, Y., Mehta, S., Syme, A., Coppens, F., Droesbeke, B., Soranzo, N., Bacon, W.,…Batut, B. (2023). Galaxy Training: A powerful framework for teaching! *PLOS Computational Biology*, *19*(1), e1010752. https://doi.org/10.1371/journal.pcbi.1010752

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R.,…Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, *248*, 108654. https://doi.org/10.1016/j.biocon.2020.108654

Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H. M., Carpenter, K. E., Chanson, J., Collen, B., Cox, N. A., Darwall, W. R. T., Dulvy, N. K., Harrison, L. R., Katariya, V., Pollock, C. M., Quader, S., Richman, N. I., Rodrigues, A. S. L., Tognelli, M. F.,…Stuart, S. N. (2010). The Impact of Conservation on the Status of the World's Vertebrates. *Science*, *330*(6010), 1503-1509. https://doi.org/doi:10.1126/science.1194442

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*(14), 3594-3602. https://doi.org/10.1111/mec.14187

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

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*(1), 63-72. https://doi.org/10.1038/hdy.2015.17

Kavaarpuo, G., Oppong-Yeboah, N. Y., & Vuin, A. (2022). Darwin: Towards the sustainability of the Larrikin of Australian capital cities. *Cities*, *120*, 103457. https://doi.org/10.1016/j.cities.2021.103457

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782-788. https://doi.org/10.1111/2041-210X.12067

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In F. Pompanon & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics: Methods and Protocols* (pp. 67-89). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_5

Kim, V. (2019). *lpsymphony: symphony integer linear programming solver in R.* In http://R-Forge.R-project.org/projects/rsymphony. https://projects.coin-or.org/SYMPHONY. http://www.coin-or.org/download/source/SYMPHONY/

Kupferberg, S. J., Palen, W. J., Lind, A. J., Bobzien, S., Catenazzi, A., Drennan, J., & Power, M. E. (2012). Effects of Flow Regimes Altered by Dams on Survival, Population Declines, and Range-Wide Losses of California River-Breeding Frogs. *Conservation Biology*, *26*(3), 513-524. https://doi.org/10.1111/j.1523-1739.2012.01837.x

Lariviere, D., Ostrovsky, A., Gallardo, C., Syme, A., Abueg, L., Pickett, B., Formenti, G., Sozzoni, M., & Nekrutenko, A. (2024). *VGP assembly pipeline: Step by Step (Galaxy Training Materials)*. Retrieved April 29 2024, from https://training.galaxyproject.org/training-material/topics/assembly/tutorials/vgp_genome_assembly/tutorial.html

Liddle, D. T., Cowie, I. D., Hirst, S. R., & Stuckey, B. M. (2017). A field guide to plants of Darwin Sandsheet Heath. *Report to Territory Natural Resource Management, Project NTRM00420. Top End Native Plant Society in collaboration with Northern Territory Herbarium, Northern Territory Government, Darwin*.

McLennan, E. A., Kovacs, T. G. L., Silver, L. W., Chen, Z., Jaya, F. R., Ho, S. Y. W., Belov, K., & Hogg, C. J. (2025). Genomics identifies koala populations at risk across eastern Australia. *Ecological Applications*, *35*(1), e3062. https://doi.org/https://doi.org/10.1002/eap.3062

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J. F., Ajmone-Marsan, P., & Colli, L. (2017). BITE: an R package for biodiversity analyses. *bioRxiv*, 181610. https://doi.org/10.1101/181610

Munshi-South, J., Zak, Y., & Pehek, E. (2013). Conservation genetics of extremely isolated urban populations of the northern dusky salamander (Desmognathus fuscus) in New York City. *PeerJ*, *1*, e64. https://doi.org/10.7717/peerj.64

Northern Territory Environment Protection_Authority. (2015). *Environmental Quality Report: Biodiversity of the Howard Sand Plains Site of Conservation Significance*. https://ntepa.nt.gov.au/_media/about-ntepa/advice-policy-procedures/advice-recommendations/howard_sand_plains_env_quality_report.pdf

Pavlova, A., Tonkin, Z., Pearce, L., Robledo-Ruiz, D., Lintermans, M., Ingram, B., Lyon, J., Beitzel, M., Broadhurst, B., Rourke, M., Sturgiss, F., Lake, E., Castrejón-Figueroa, J., Stocks, J., & Sunnucks, P. (2024). A shift to metapopulation genetic management for persistence of a species threatened by fragmentation: the case of an endangered Australian freshwater fish. https://doi.org/10.22541/au.172801777.74141428/v1

Peakall, R., & Smouse, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288-295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*(19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Peter, B. M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, *202*(4), 1485-1501. https://doi.org/10.1534/genetics.115.183913

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pröhl, H., Auffarth, J., Bergmann, T., Buschmann, H., & Balkenhol, N. (2021). Conservation genetics of the yellow-bellied toad (Bombina variegata): population structure, genetic diversity and landscape effects in an endangered amphibian. *Conservation Genetics*, *22*(3), 513-529. https://doi.org/10.1007/s10592-021-01350-5

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559-575. https://doi.org/10.1086/519795

Reynolds, S., & Grattidge, A. (2012). Caring for Our Country Surveys for the Howard River Toadlet (*Uperoleia daviesae*). *EcOZ, Darwin, Northern Territory*.

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, *5*(7), P54. https://doi.org/10.1186/1753-6561-5-S7-P54

Sarkar, U. K., Chandran, R., Teena Jayakumar, T. K., & Ravi, C. (2024). In Situ Conservation: Tools, Strategies, and Challenges. In U. K. Sarkar, T. T. A. Kumar, N. Sood, R. K. Singh, R. Kumar, & L. K. Tyagi (Eds.), *Sustainable Management of Fish Genetic Resources* (pp. 71-89). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-5250-8_5

Schmidt, T. L., Jasper, M.-E., Weeks, A. R., & Hoffmann, A. A. (2021). Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and Evolution*, *12*(10), 1888-1898. https://doi.org/10.1111/2041-210X.13659

Sharp, J. (2010). The impacts of urbanization on groundwater systems and recharge. *Aqua Mundi*, *1*, 51-56. https://doi.org/10.4409/Am-004-10-0008

Sheean, V. A., Manning, A. D., & Lindenmayer, D. B. (2012). An assessment of scientific approaches towards species relocations in Australia. *Austral Ecology*, *37*(2), 204-215. https://doi.org/10.1111/j.1442-9993.2011.02264.x

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information Theory Broadens the Spectrum of Molecular Ecology and Evolution. *Trends in Ecology & Evolution*, *32*(12), 948-963. https://doi.org/10.1016/j.tree.2017.09.012

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212. https://doi.org/https://doi.org/10.1093/bioinformatics/btv351

Smalling, K. L., Reeves, R., Muths, E., Vandever, M., Battaglin, W. A., Hladik, M. L., & Pierce, C. L. (2015). Pesticide concentrations in frog tissue and wetland habitats in a landscape dominated by agriculture. *Science of The Total Environment*, *502*, 80-90. https://doi.org/10.1016/j.scitotenv.2014.08.114

Smouse, P. E., Banks, S. C., & Peakall, R. (2017). Converting quadratic entropy to diversity: Both animals and alleles are diverse, but some are more diverse than others. *PLOS ONE*, *12*(10), e0185499. https://doi.org/10.1371/journal.pone.0185499

Sopniewski, J., & Catullo, R. A. (2024). Estimates of heterozygosity from single nucleotide polymorphism markers are context-dependent and often wrong. *Mol Ecol Resour*, *24*(4), e13947. https://doi.org/10.1111/1755-0998.13947

Taylor, D. (2011). *The Highway One travel companion. Volume 1, Melbourne to Tweed Heads / David Taylor*. Boolarong Press.

The Galaxy Community. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, *50*(W1), W345-W351. https://doi.org/10.1093/nar/gkac247

Threatened Species Scientific Committee. (2021). Conservation Advice Uperoleia daviesae (Howard River Toadlet). *Canberra: Department of Agriculture, Water and the Environment*. Available from: http://www.environment.gov.au/biodiversity/threatened/species/pubs/85375-conservation-advice-13112021.pdf. In effect under the EPBC Act from 13-Nov-2021.

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Consortium, D. T. o. L., Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy, S. A. (2023). MitoHiFi: a Python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads. *bioRxiv*, 2022.2012.2023.521667. https://doi.org/https://doi.org/10.1101/2022.12.23.521667

Vasimuddin, M., Misra, S., Li, H., & Aluru, S. (2019, 20-24 May 2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS),

Wake, D. B., & Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences*, *105*(supplement_1), 11466-11473. https://doi.org/10.1073/pnas.0801921105

Watts, M. E., Ball, I. R., Stewart, R. S., Klein, C. J., Wilson, K., Steinback, C., Lourival, R., Kircher, L., & Possingham, H. P. (2009). Marxan with Zones: Software for optimal conservation based land- and sea-use zoning. *Environmental Modelling & Software*, *24*(12), 1513-1521. https://doi.org/10.1016/j.envsoft.2009.06.005

Young, J. E., Tyler, M. J., & Kent, S. A. (2005). Diminutive New Species of *Uperoleia* Grey (Anura: Myobatrachidae) from the Vicinity of Darwin, Northern Territory, Australia. *Journal of Herpetology*, *39*(4), 603-609, 607. https://doi.org/10.1670/77-05A.1

Zecca, G., Labra, M., & Grassi, F. (2020). Untangling the Evolution of American Wild Grapes: Admixed Species and How to Find Them [Original Research]. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.01814

Figure captions

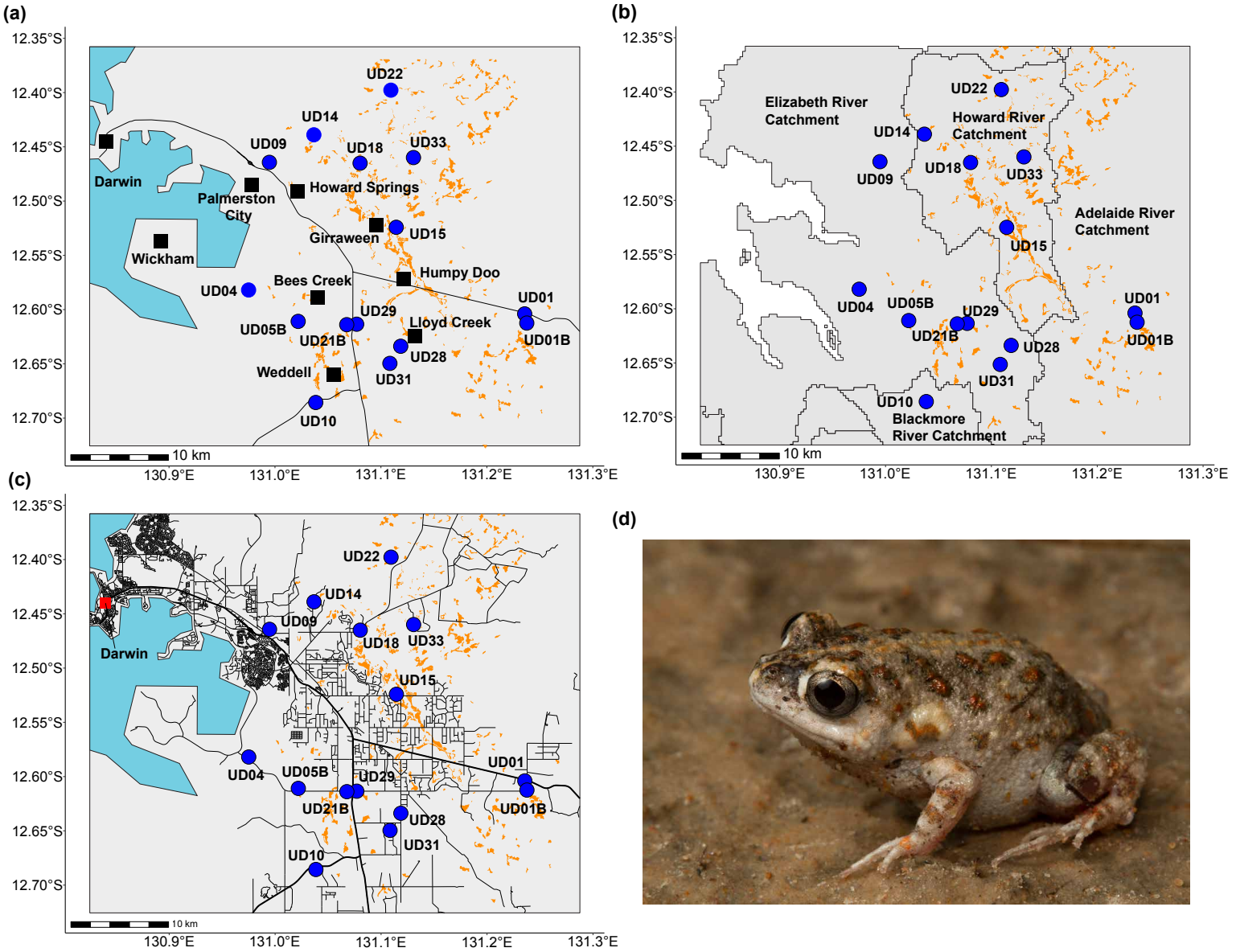
Figure 1. Sampling locations of *U. daviesae* (blue dots) across the broader Darwin peri-urban growth area, shown on: (a) major towns (black squares) and highways; (b) catchments; (c) road networks; and (d) with a photograph of *U. daviesae* by Matt Clancy. Areas shaded in brown represent the remaining sandsheet heath habitat (Department of Lands, Planning and Environment, 2014). Sampling locations cover the full known geographic extent of the species in all directions.

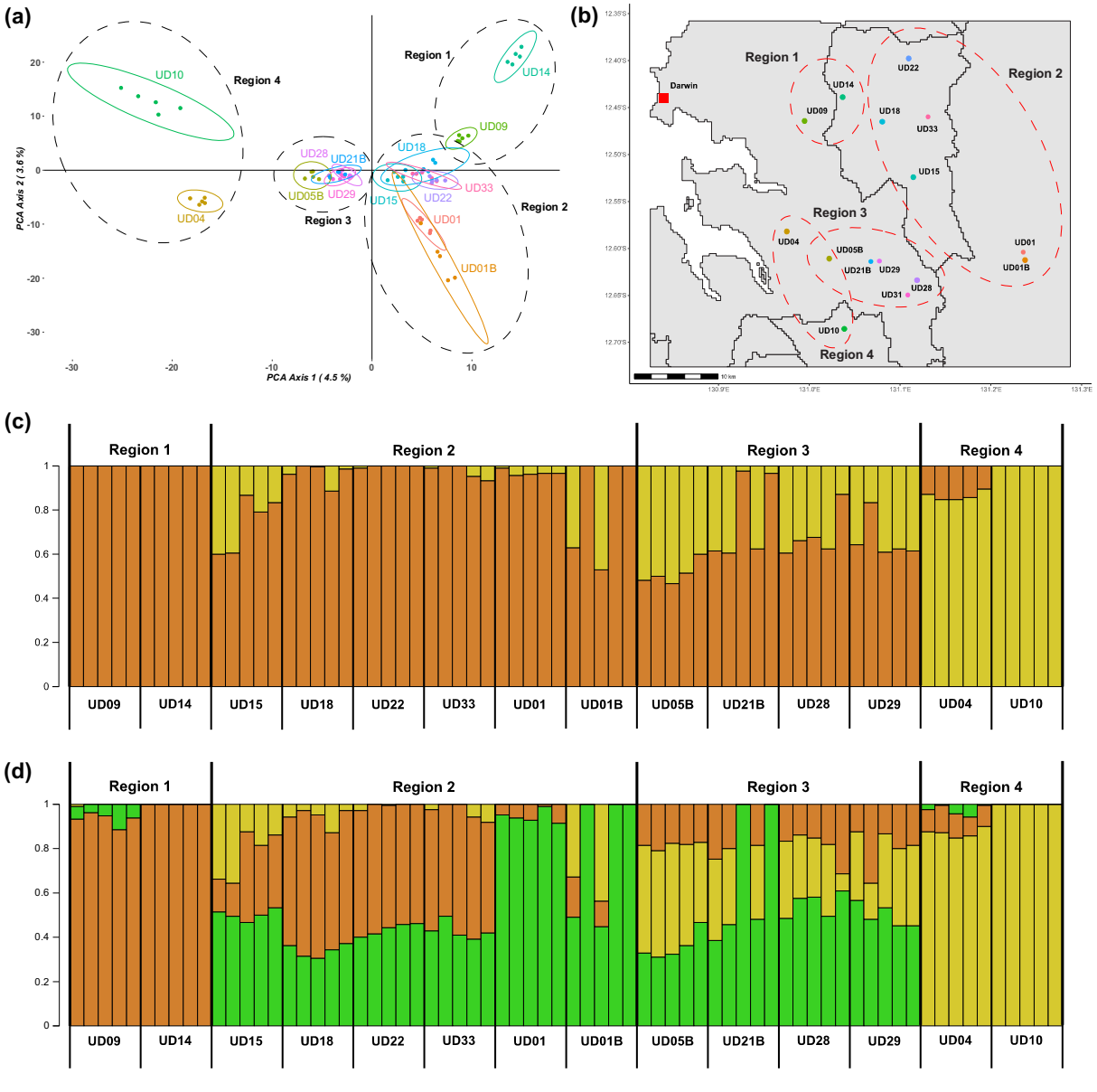
Figure 2. PCoA plot (a) with the sampling locations map (b) with region definitions on the catchments map. The ellipses in these two figures were just to encapsulate points for each site and no statistical meeting. In (c) and (d), the K=2 and K=3 analyses use the same number (N=5) of tissue samples at each site.

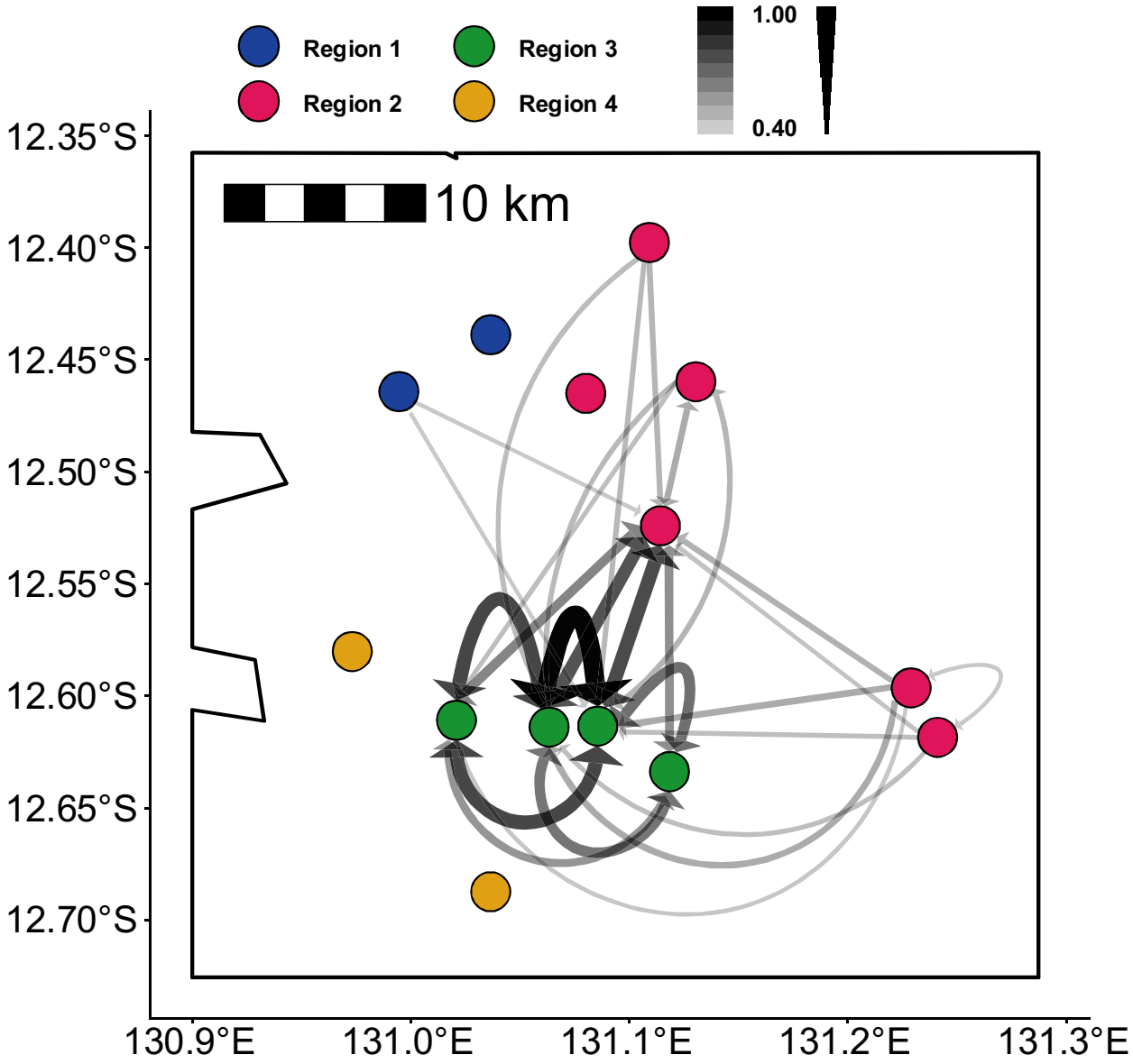
Figure 3. Direction of gene flow between sites based on the effective number of migrants (N_m). The width and opacity of gene flow lines were scaled using the N_M value, ranging from 0.40 to 1.00.

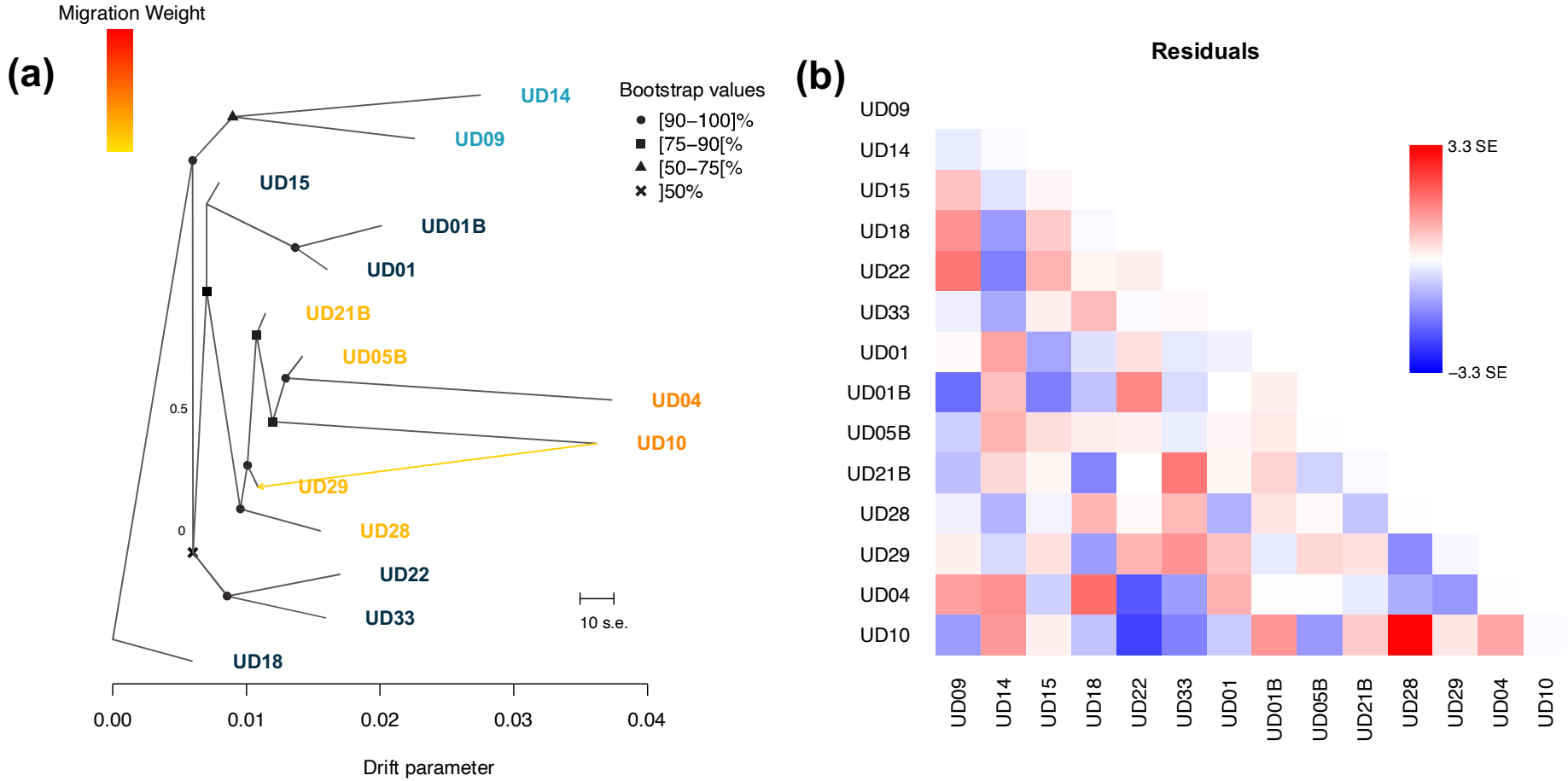
Figure 4. *TreeMix* consensus tree and bootstrap values displayed site relationships as a bifurcating maximum likelihood tree. (a) One migration edge (m=1), inferred as the best topology. (b) Associated residual fit of the observed versus the predicted squared allele frequency difference.

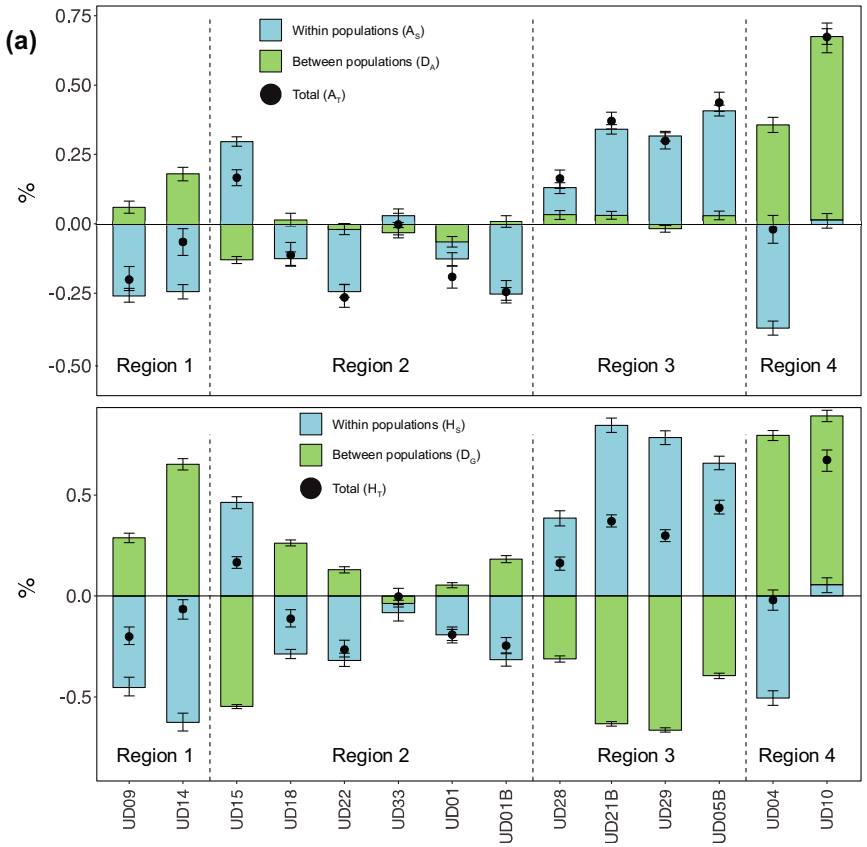
Figure 5. The contributions of each site to the allelic and gene diversity of the species. (a) contributions to within sites (A_S, H_S), among sites (D_A, D_G), and total (A_T, H_T) allelic and gene diversity. (b) Contributions to the allelic (K) and gene (H) diversity of a gene pool with randomly selected 1000 individuals.


(a)
12.35°S
12.40°S
12.45°S
12.50°S
12.55°S
12.60°S
12.65°S
12.70°S
130.9°E
131.0°E
131.1°E
131.2°E
131.3°E
10 km
Darwin
Palmerston City
Howard Springs
Wickham
Girraween
Bees Creek
Humpy Doo
Lloyd Creek
Weddell
UD22
UD14
UD09
UD18
UD33
UD15
UD04
UD01
UD29
UD05B
UD21B
UD01B
UD28
UD31
UD10
(b)
Elizabeth River Catchment
Howard River Catchment
Adelaide River Catchment
Blackmore River Catchment
(c)
(d)


(a)
Region 1
Region 2
Region 3
Region 4
UD10
UD14
UD09
UD18
UD28
UD21B
UD05B
UD29
UD15
UD22
UD33
UD01
UD01B
UD04
PCA Axis 2 (3.6 %)
PCA Axis 1 (4.5 %)
(b)
Region 1
Region 2
Region 3
Region 4
Darwin
UD14
UD22
UD09
UD18
UD33
UD15
UD04
UD05B
UD21B
UD29
UD01
UD01B
UD28
UD31
UD10
(c)
Region 1
Region 2
Region 3
Region 4
UD09
UD14
UD15
UD18
UD22
UD33
UD01
UD01B
UD05B
UD21B
UD28
UD29
UD04
UD10
(d)
Region 1
Region 2
Region 3
Region 4
UD09
UD14
UD15
UD18
UD22
UD33
UD01
UD01B
UD05B
UD21B
UD28
UD29
UD04
UD10


Region 1
Region 2
Region 3
Region 4
1.00
0.40
10 km
12.35°S
12.40°S
12.45°S
12.50°S
12.55°S
12.60°S
12.65°S
12.70°S
130.9°E
131.0°E
131.1°E
131.2°E
131.3°E


(a)
Migration Weight
Bootstrap values
[90–100]%
[75–90[%
[50–75[%
]50%
UD14
UD09
UD15
UD01B
UD01
UD21B
UD05B
UD04
UD10
UD29
UD28
UD22
UD33
UD18
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
Drift parameter
(b)
Residuals
3.3 SE
−3.3 SE
UD09
UD14
UD15
UD18
UD22
UD33
UD01
UD01B
UD05B
UD21B
UD28
UD29
UD04
UD10


(a)
Within populations (A_S)
Between populations (D_A)
Total (A_T)
%
0.75
0.50
0.25
0.00
-0.25
-0.50
Region 1
Region 2
Region 3
Region 4
Within populations (H_S)
Between populations (D_G)
Total (H_T)
%
0.5
0.0
-0.5
Region 1
Region 2
Region 3
Region 4
UD09
UD14
UD15
UD18
UD22
UD33
UD01
UD01B
UD28
UD21B
UD29
UD05B
UD04
UD10

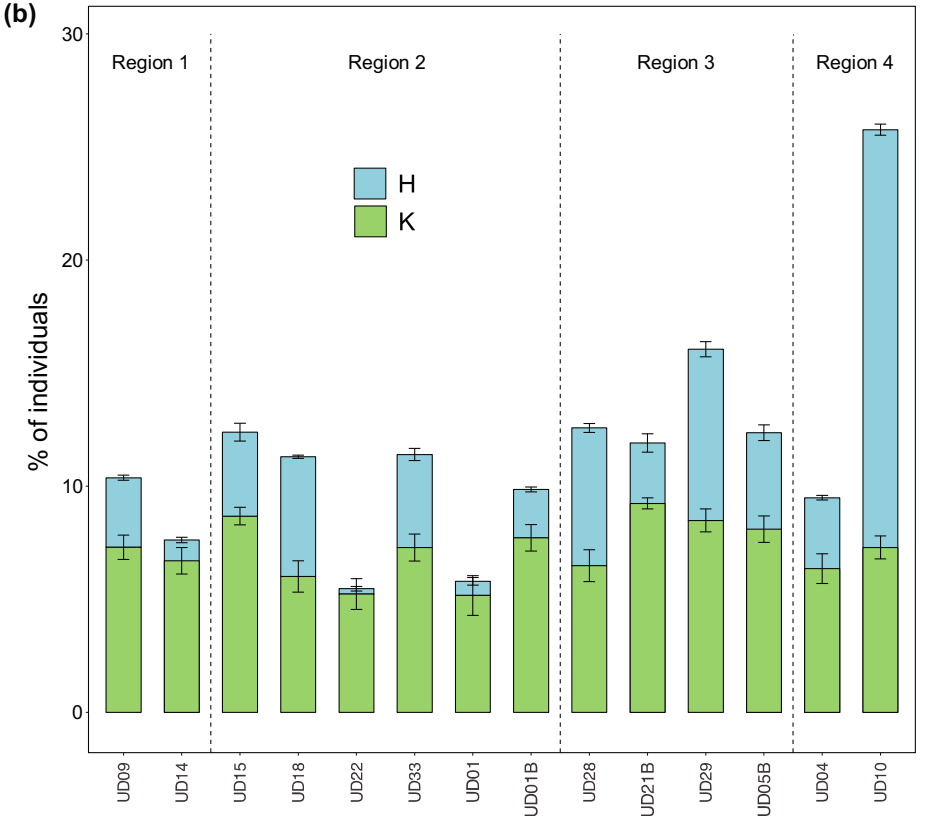

(b)
% of individuals
30
20
10
0
Region 1
Region 2
Region 3
Region 4
H
K
UD09
UD14
UD15
UD18
UD22
UD33
UD01
UD01B
UD28
UD21B
UD29
UD05B
UD04
UD10

Table 1. Genetic diversity among regions. N: sample size; P_A: number of private alleles; A_R: allelic richness; Auto. H_O: autosomal observed heterozygosity; Unb. H_O: JeDi unbiased heterozygosity; Auto. H_E: autosomal expected heterozygosity; Auto. Pi: autosomal nucleotide diversity; Unb. Pi: JeDi unbiased nucleotide diversity; F_h: genome-wide estimates of heterozygosity. The standard errors or deviations (for Q-profile diversity) are given in brackets.

Regions	N	PA	AR	Auto. HO	Unb. HO	Auto. HE	Auto. Pi	Unb. Pi	Q=0 diversity	Q=1 diversity	Q=2 diversity	Fh
Region 1	12	284 (19)	1.844 (0.002)	0.00127 (0.00001)	0.00206 (0.00003)	0.00132 (0.00001)	0.00138 (0.00001)	0.00203 (0.00005)	0.904 (0.294)	0.410 (0.209)	0.263 (0.156)	0.1719 (0.0078)
Region 2	12	10 (0)	1.829 (0.002)	0.00141 (0.00001)	0.00219 (0.00003)	0.00144 (0.00001)	0.00151 (0.00001)	0.00216 (0.00004)	0.893 (0.309)	0.387 (0.212)	0.246 (0.158)	0.1095 (0.0079)
Region 3	12	26 (1)	1.804 (0.002)	0.00154 (0.00001)	0.00221 (0.00003)	0.00155 (0.00001)	0.00162 (0.00001)	0.00242 (0.00002)	0.866 (0.341)	0.379 (0.220)	0.241 (0.162)	0.0538 (0.0101)
Region 4	12	120 (9)	1.803 (0.002)	0.00153 (0.00001)	0.00246 (0.00006)	0.00161 (0.00001)	0.00168 (0.00001)	0.00232 (0.00026)	0.875 (0.330)	0.390 (0.218)	0.250 (0.161)	0.1037 (0.0208)

Supplement Information

Population genomics of Uperoleia daviesae (Anura: Myobatrachidae) highlights the vulnerability of naturally fragmented short-range endemics to urban development

Shengyao Lin, Peter J. McDonald, Alistair Stewart, Carolyn J. Hogg, Luke W. Silver, Sam C. Banks, Graeme R. Gillespie, Raphael K. Didham, and Renee A. Catullo

Table of Contents

Methods

Genome assembly and annotation

The reference sample was sequenced at the AGRF (Brisbane, Australia) using *PacBio HiFi* library prep and *Revio* sequencing on two 25M SMRT cells. Total RNA from the liver and the pooled kidney/gonads was sequenced at the Ramaciotti Centre for Comparative Genomics using an *Illumina* stranded mRNA prep. The RNA was sequenced on a NovaSeq X Plus 10B flowecll, using 150bp paired-end reads, resulting in 69 (kidney/gonads) to 81 (liver) million reads per sample. To generate the reference genome for the *U. daviesae*, we used the Vertebrate Genome Project (*VGP)* assembly pipeline on the Galaxy Australia interface (Batut et al., 2018; Hiltemann et al., 2023; Lariviere et al., 2024; The Galaxy Community, 2022). Briefly, HiFi reads were quality-trimmed, and reads containing adapters were removed using *Cutadapt* v4.6 (Martin, 2011). Genome size and k-mers were estimated using *Meryl* v1.3 (Rhie et al., 2020) and *GenomeScope* v2.0 (Ranallo-Benavidez et al., 2020). Genome assembly was performed using *hifiasm* v0.19.8 in the solo mode with the purging level set to 'light' (Cheng et al., 2021) using HiFi reads to assemble contigs, with the assembly quality assessed using *gfastats* v1.3.6 (Formenti et al., 2022) and *BUSCO* v5.5.0 with both vertebrata_odb10 and tetrapoda_odb10 lineages (Simao et al., 2015). *BUSCO* revealed high numbers of duplicated genes in the primary assembly. Therefore, purging was performed using "purge_dups" to remove duplicate contigs from the primary assembly and relocate them to the alternate assembly. A second round of "purge_dups" was run on the alternate assembly.

The mitochondrial genome was identified using *MitoHiFi* v2 (Uliano-Silva et al., 2023), mitofinder identified the Wokan cannibal frog (*Lechriodus melanopyga;* NC_019999) as the most taxonomically closely related mitogenome publicly available, which was used to search for the *U. daviesae* mitogenome. The mitogenome was then visualised using *Proksee* (Grant et al., 2023). Final genome analysis was conducted using *BUSCO* v5.5.0 with both vertebrata_odb10 (n = 3354) and tetrapoda_odb10 (n = 5310) lineages, *Merqury* v1.3 (Rhie et al., 2020) and the Fasta Statistics tools provided in Galaxy Australia. Repetitive elements of the genome were identified and masked using *RepeatModeler* v2.0.4 and *RepeatMasker* v4.1.5 (Flynn et al., 2020).

We generated a reference-aligned transcriptome assembly to assist with genome annotation using workflows on the Galaxy Australia interface. Briefly, we quality checked raw and trimmed reads using FastQC v0.12.1(Andrews, 2014). Reads were quality trimmed using trimmomatic v0.36 (Bolger et al., 2014) with parameters SLIDINGWINDOW:4:5, LEADING:5, TRAILING:5 and MINLEN:25 and ILLUMINACLIP:2:30:10 with the TruSeq3-PE adapters. Trimmed reads were then aligned to the masked genome using hisat2 v2.2.1 (Kim et al., 2019) and individual tissue transcript gtf files generated with stringtie v2.2.1 (Pertea et al., 2015). Stringtie merge was used to generate a global transcriptome, GTF files and coding probabilities were calculated using CPAT v3.0.5 (Wang et al., 2013) with coding and pseudogene sequences from the Corroboree frog (*Pseudophryne corroboree*) used as query sequences. Transcripts with coding probability higher than 50% were used as input to transdecoder v5.5.0 (Haas, 2022) to predict open reading frames and retain the longest isoform for each transcript with a minimum amino acid length of 30. Transcriptome completeness was assessed using BUSCO v5.5.0 (Simao et al., 2015) with the tetrapoda_odb10 (n = 5310) and vertebrata_odb10 (n = 3354) lineages. Genome annotation was also performed using FGenesH++ (Solovyev et al., 2006; Solovyev, 2002) using the longest predicted open reading frame, non-mammalian settings and optimised parameters for the generic Xenopus gene finding matrix. As the number of genes predicted was higher than expected, we conducted filtering by first removing gene predictions with no hits to the non-redundant database and those with hits to the non-redundant database with an e-value greater than $1.0 \times 10^{\wedge -130}$. We also removed genes that were predicted with mRNA evidence that were annotated on scaffolds different to the scaffold the mRNA had aligned to during transcriptome assembly. BUSCO v5.5.0 in protein mode was used to assess annotation completeness on the filtered protein dataset with the tetrapoda_odb10 (n = 5310) and vertebrata_odb10 (n = 3354) lineages on Galaxy Australia.

Figures

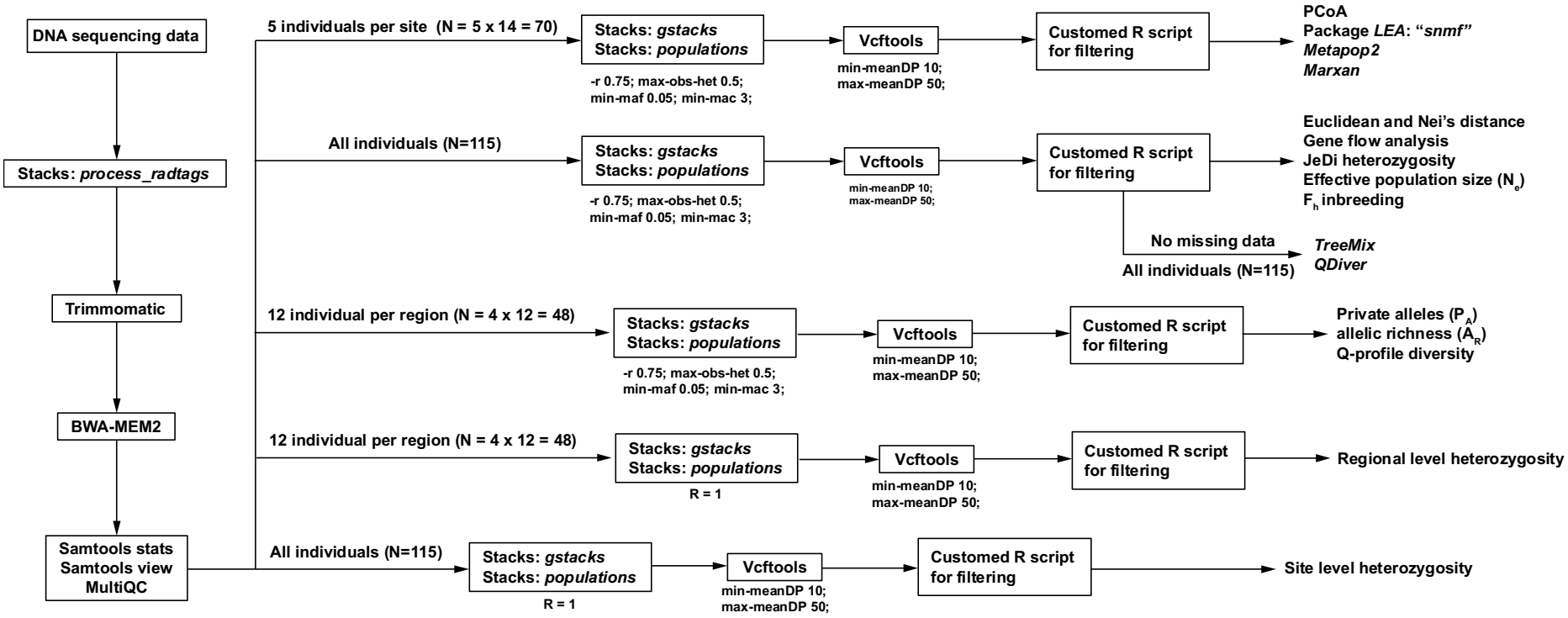


Figure S1. Bioinformatic workflow used to generate SNP datasets for population genomic analyses in *U. daviesae*. The pipeline is structured to accommodate different levels of missing data and site representation, depending on the requirements of downstream analyses. Six main workflows are shown: (1) a dataset with five individuals per site and moderate missingness for population genetic structure analysis, *Metapop2*, and *Marxan* analysis; (2) a dataset with all individuals and moderate missingness for genetic distances, gene flow analysis, JeDi heterozygosity, effective population size, and F_h inbreeding estimated; (3) a dataset with all individuals and no missingness for *TreeMix* and *QDiver* analysis; (4) a dataset with 12 individuals per region and moderate missingness for private alleles, allelic richness, and Q-profile diversity; (5) a dataset with

12 individuals per region and each locus present in all individuals for regional heterozygosity; (6) a dataset with all individuals and each locus present in all individuals for site-level heterozygosity. Filtering thresholds are indicated at each step to show how the data are tailored for specific analytical goals.

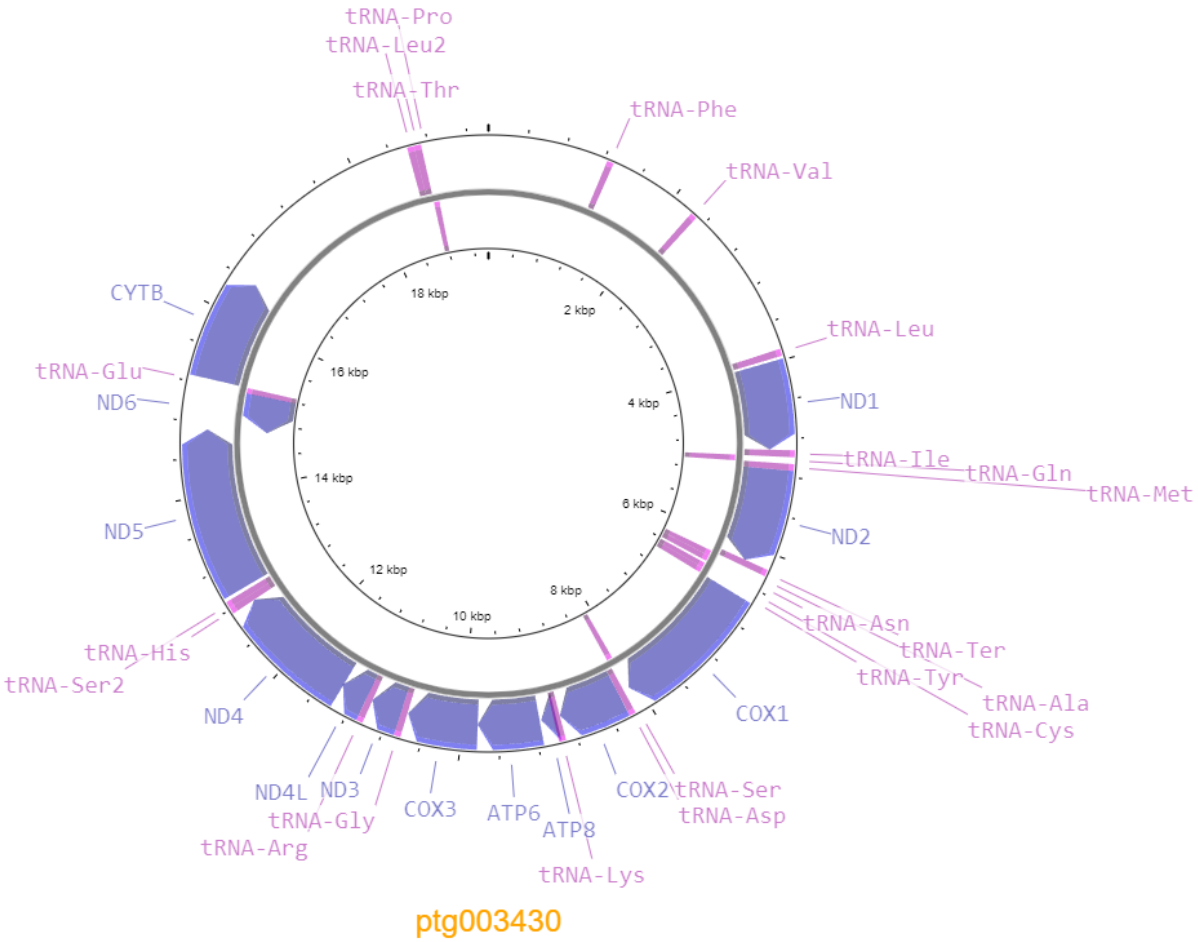


Figure S2. Depiction of the circularised mitogenome of the Howard River Toadlet generated using Proksee (Grant et al., 2023), gene elements on the outer ring are in the forward direction, and gene elements on the inner ring are in the reverse direction.

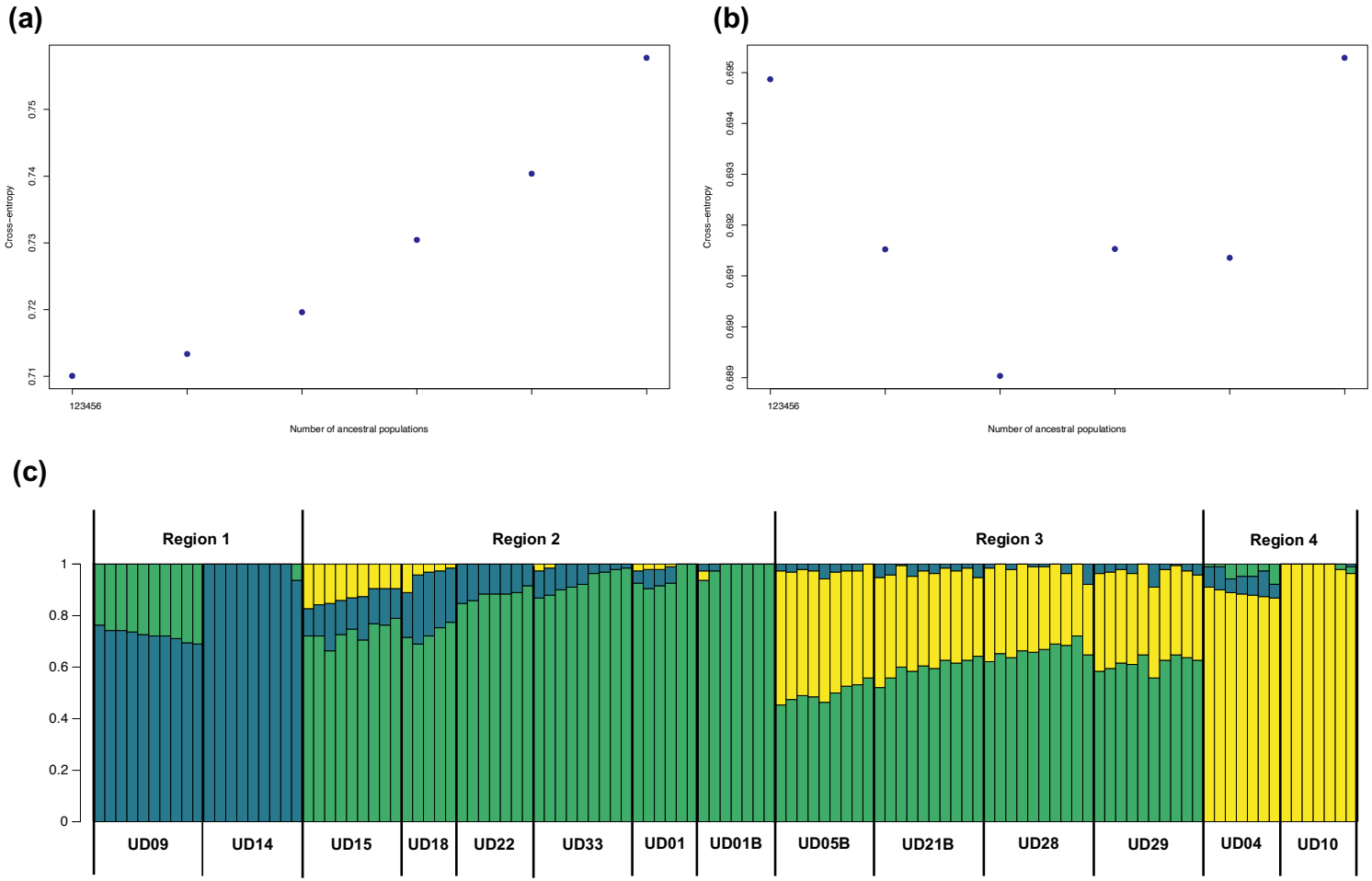


Figure S3. Cross-entropy plots generated by the LEA package for determining the genetic cluster number (K) using (a): the same individual numbers (N = 5) for each site; (b): different individual numbers for each site; (c): K = 3 structure plot with different individual numbers for each site.

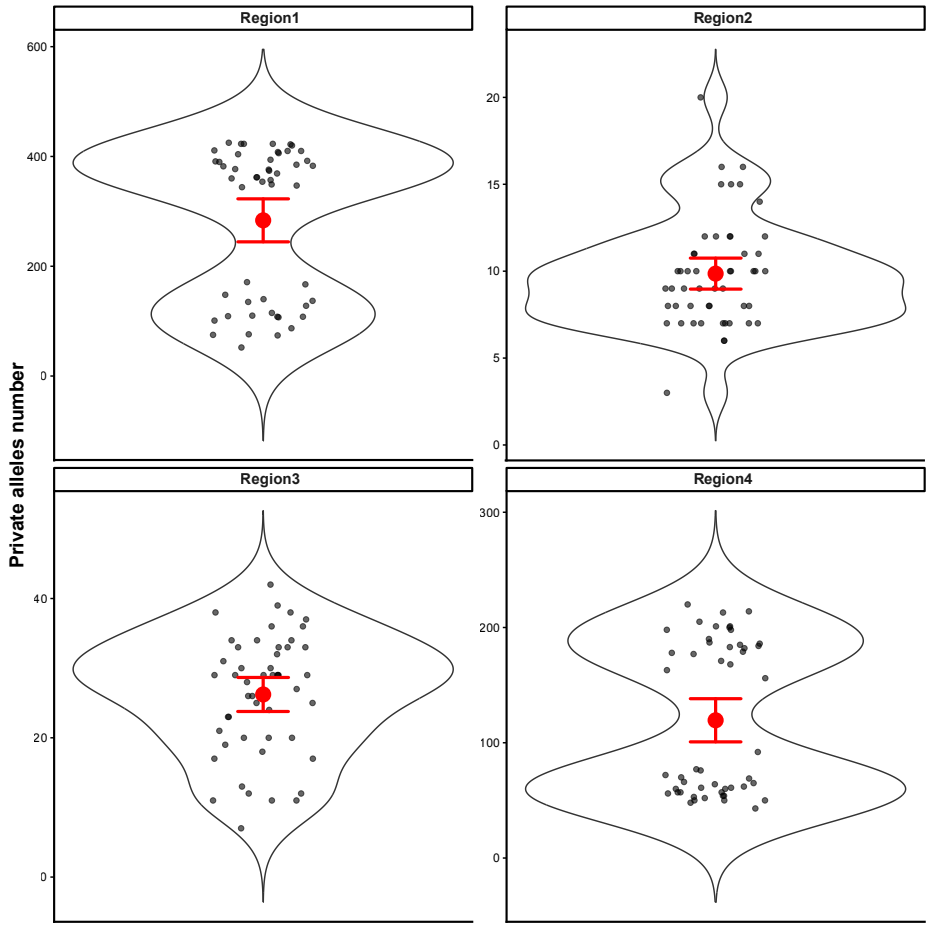


Figure S4. Variation in private allele counts across regions. Violin plots show the distribution of private allele numbers across 50 replicated runs for each region (Region 1–4). Black points represent individual replicate values. Red points indicate the mean private allele count for each region, with error bars showing the uncertainty around the mean (95% confidence interval).

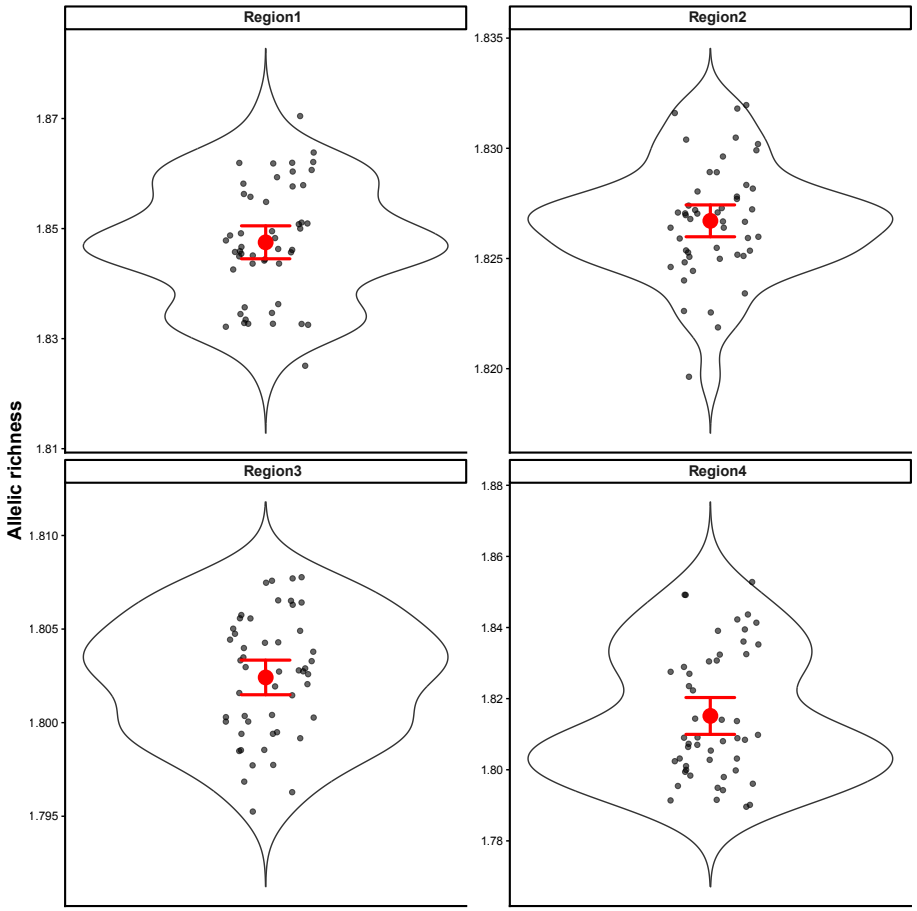


Figure S5. Variation in allelic richness across regions. Violin plots show the distribution of allelic richness across 50 replicated runs for each region (Region 1–4). Black points represent individual replicate values. Red points indicate the mean allelic richness for each region, with error bars showing the uncertainty around the mean (95% confidence interval).

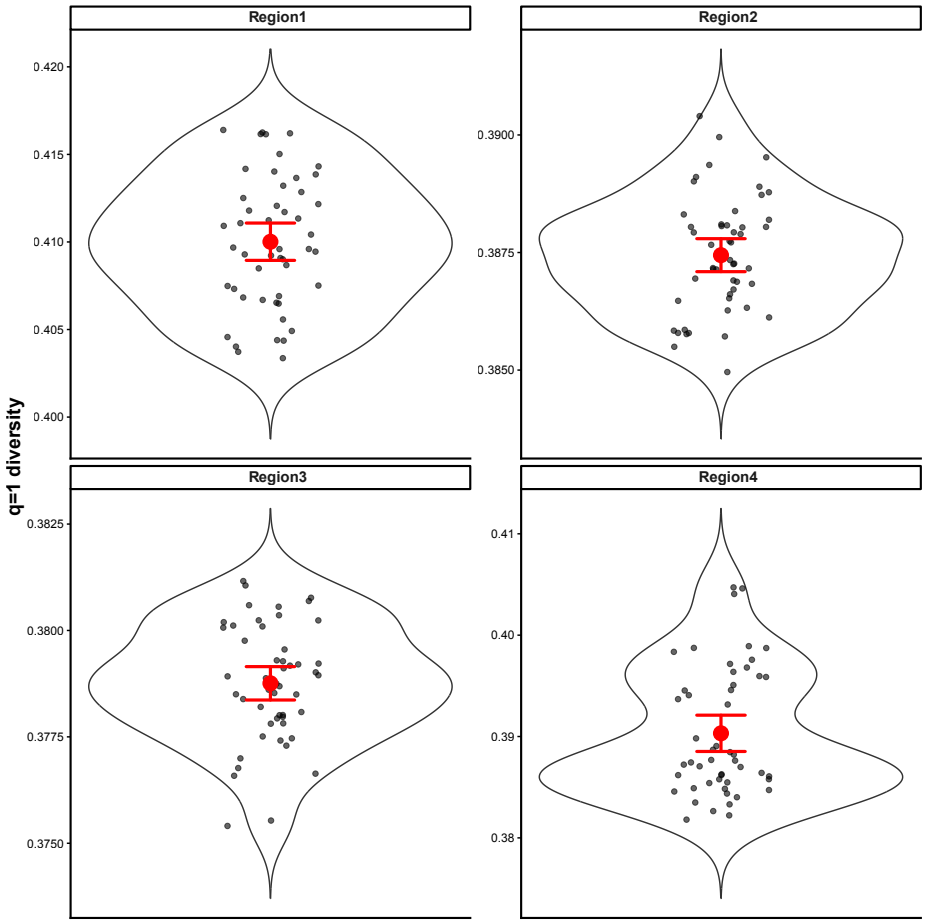


Figure S6. Variation in q=1 diversity (Shannon information) regions. Violin plots show the distribution of q=1 diversity across 50 replicated runs for each region (Region 1–4). Black points represent individual replicate values. Red points indicate the mean q=1 diversity for each region, with error bars showing the uncertainty around the mean (95% confidence interval).

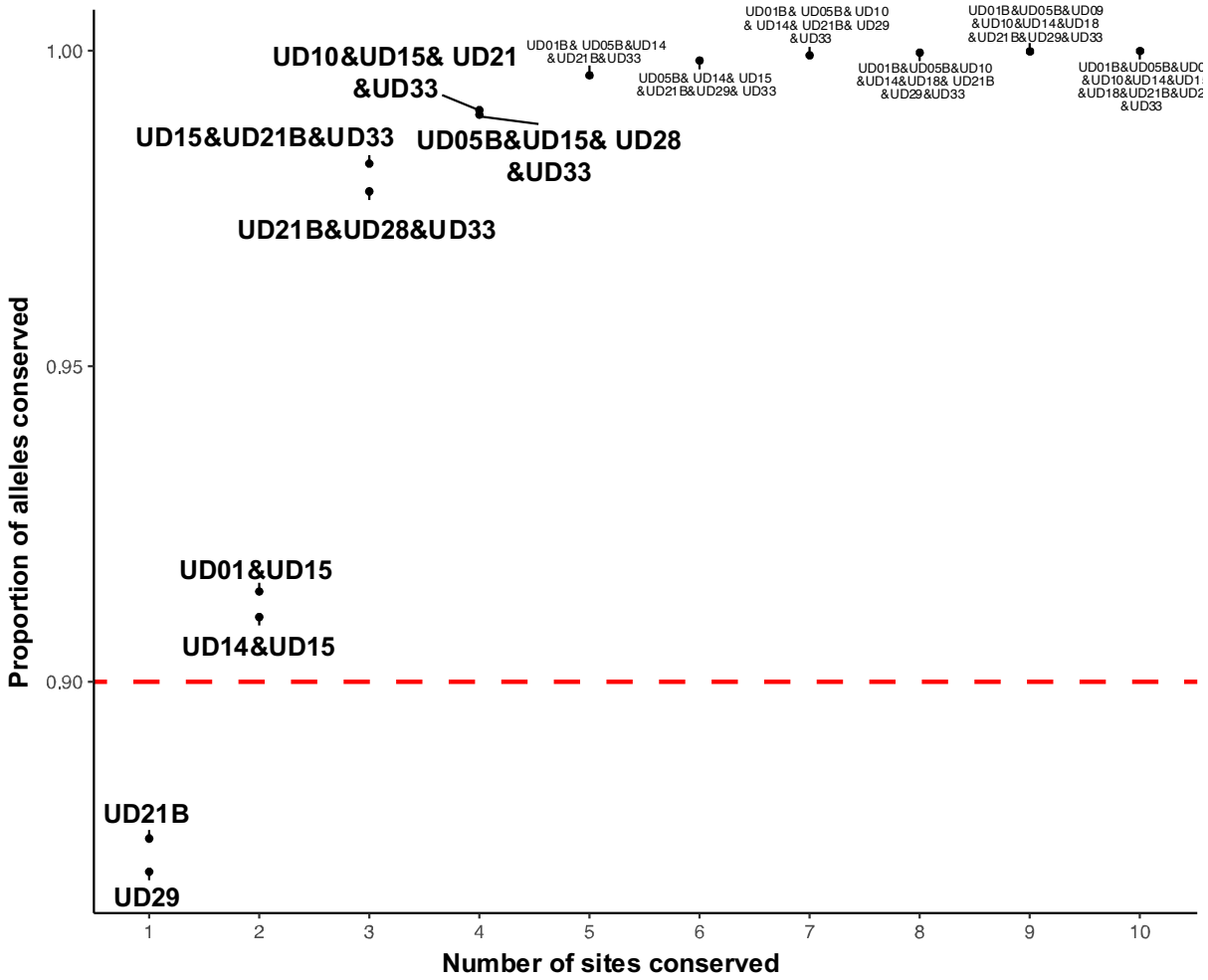


Figure S7. The effect of conserving differing site numbers on the proportion of the total number of alleles in the species, with the red line indicating that 90% of alleles are conserved.

Tables

Table S1. Spatial information for the sampled individuals during the fieldwork. The location coordinates have been rounded to hide sensitive information on threatened species.

ID	Site	Lat	Lon
455804	UD28	-12.63	131.12
455805	UD28	-12.63	131.12
455806	UD28	-12.63	131.12
455807	UD28	-12.63	131.12
455811	UD29	-12.61	131.08
455812	UD28	-12.63	131.12
455813	UD15	-12.52	131.11
455814	UD29	-12.61	131.08
455815	UD29	-12.61	131.08
455816	UD29	-12.61	131.08
455818	UD15	-12.52	131.11
455819	UD29	-12.61	131.08
455819	UD29	-12.61	131.08
455820	UD29	-12.61	131.08
455821	UD15	-12.52	131.11
455822	UD29	-12.61	131.08
455823	UD28	-12.63	131.12
455824	UD15	-12.52	131.11
455824	UD15	-12.52	131.11
455825	UD29	-12.61	131.08
455826	UD15	-12.52	131.11
455826	UD15	-12.52	131.11
455827	UD15	-12.52	131.11
455827	UD15	-12.52	131.11
455828	UD15	-12.52	131.11
455828	UD15	-12.52	131.11
455829	UD29	-12.61	131.08
455830	UD15	-12.52	131.11
455830	UD15	-12.52	131.11
455831	UD15	-12.52	131.11
455831	UD15	-12.52	131.11
455832	UD28	-12.63	131.12
455833	UD29	-12.61	131.08
455834	UD28	-12.63	131.12

455835	UD28	-12.63	131.12
455837	UD15	-12.52	131.11
455837	UD15	-12.52	131.11
455838	UD28	-12.63	131.12
455839	UD01	-12.60	131.24
455840	UD01	-12.60	131.24
455841	UD01	-12.60	131.24
455842	UD01	-12.60	131.24
455843	UD01	-12.60	131.24
455844	UD01	-12.60	131.24
455845	UD01	-12.60	131.24
455846	UD01	-12.60	131.24
455847	UD01	-12.60	131.24
455848	UD01	-12.60	131.24
455849	UD09	-12.46	130.99
455850	UD09	-12.46	130.99
455851	UD09	-12.46	130.99
455852	UD10	-12.69	131.04
455853	UD10	-12.69	131.04
455854	UD10	-12.69	131.04
455855	UD10	-12.69	131.04
455856	UD10	-12.69	131.04
455857	UD10	-12.69	131.04
455858	UD10	-12.69	131.04
455859	UD09	-12.46	130.99
455860	UD09	-12.46	130.99
455861	UD09	-12.46	130.99
455862	UD09	-12.46	130.99
455863	UD09	-12.46	130.99
455864	UD09	-12.46	130.99
455865	UD09	-12.46	130.99
455866	UD14	-12.44	131.04
455867	UD14	-12.44	131.04
455868	UD14	-12.44	131.04
455869	UD04	-12.58	130.98
455870	UD04	-12.58	130.98
455871	UD04	-12.58	130.98
455872	UD14	-12.44	131.04
455873	UD14	-12.44	131.04
455873	UD14	-12.44	131.04

455874	UD14	-12.44	131.04
455874	UD14	-12.44	131.04
455875	UD14	-12.44	131.04
455876	UD14	-12.44	131.04
455877	UD14	-12.44	131.04
455878	UD14	-12.44	131.04
455879	UD04	-12.58	130.98
455880	UD04	-12.58	130.98
455881	UD04	-12.58	130.98
455882	UD04	-12.58	130.98
455883	UD04	-12.58	130.98
455884	UD33	-12.46	131.13
455885	UD33	-12.46	131.13
455886	UD33	-12.46	131.13
455887	UD33	-12.46	131.13
455887	UD33	-12.46	131.13
455888	UD33	-12.46	131.13
455888	UD33	-12.46	131.13
455889	UD33	-12.46	131.13
455890	UD33	-12.46	131.13
455891	UD33	-12.46	131.13
455892	UD33	-12.46	131.13
455893	UD33	-12.46	131.13
455894	UD22	-12.40	131.11
455895	UD22	-12.40	131.11
455896	UD22	-12.40	131.11
455897	UD22	-12.40	131.11
455898	UD22	-12.40	131.11
455899	UD22	-12.40	131.11
455901	UD22	-12.40	131.11
455902	UD22	-12.40	131.11
455903	UD18	-12.47	131.08
455904	UD22	-12.40	131.11
455905	UD18	-12.47	131.08
455906	UD18	-12.47	131.08
455907	UD18	-12.47	131.08
455907	UD18	-12.47	131.08
455908	UD18	-12.47	131.08
455908	UD18	-12.47	131.08
455909	UD18	-12.47	131.08

455910	UD21B	-12.61	131.07
455911	UD21B	-12.61	131.07
455912	UD21B	-12.61	131.07
455912	UD21B	-12.61	131.07
455913	UD01B	-12.61	131.24
455914	UD21B	-12.61	131.07
455914	UD21B	-12.61	131.07
455915	UD01B	-12.61	131.24
455917	UD01B	-12.61	131.24
455919	UD01B	-12.61	131.24
455919	UD01B	-12.61	131.24
455920	UD05B	-12.61	131.02
455921	UD05B	-12.61	131.02
455922	UD05B	-12.61	131.02
455923	UD05B	-12.61	131.02
455924	UD05B	-12.61	131.02
455924	UD05B	-12.61	131.02
455925	UD05B	-12.61	131.02
455925	UD05B	-12.61	131.02
455926	UD05B	-12.61	131.02
455927	UD05B	-12.61	131.02
455928	UD05B	-12.61	131.02
455929	UD21B	-12.61	131.07
455929	UD21B	-12.61	131.07
455930	UD21B	-12.61	131.07
455930	UD21B	-12.61	131.07
455931	UD21B	-12.61	131.07
455931	UD21B	-12.61	131.07
455932	UD01B	-12.61	131.24
455933	UD01B	-12.61	131.24
455934	UD31	-12.65	131.11
455935	UD31	-12.65	131.11
455935	UD31	-12.65	131.11
455936	UD05B	-12.61	131.02
455937	UD21B	-12.61	131.07
455937	UD21B	-12.61	131.07
455938	UD21B	-12.61	131.07
455938	UD21B	-12.61	131.07
455939	UD21B	-12.61	131.07
455939	UD21B	-12.61	131.07

455940	UD01B	-12.61	131.24
455941	UD01B	-12.61	131.24
455942	UD01B	-12.61	131.24

Table S2. Methods used for each genome assembly version and genome statistics for each genome version

	Primary_contigs	Alternate_contigs	aUpedav1.alt.20243004	aUpedav1.pri.20243004
Methods	Primary haplotype assembled with hifiasm	Alternate haplotype assembled with hifiasm	Purged alternate assembly	Purged primary assembly
Number bases (Mb)	6083974420	3021533494	3677698490	4771429782
GC%	44.49	44.42	44.48	44.55
Gaps (%)	0	0	0	0
Contig L50	558	1587	1741	358
Contig L90	2740	8227	6730	1444
Contig N50 (Mb)	2.75	0.52	0.61	3.59
Contig N90 (Mb)	0.41	0.06	0.13	0.78
Number contigs	7390	19207	14048	3327
Longest contig (Mb)	37.40	4.08	5.99	37.40
BUSCO Vertebrata_odb10 (n:3354)	C:95.9%[S:76.6%,D:19.3%],F:1.6%,M:2.5%	C:55.7%[S:54.4%,D:1.3%],F:6.0%,M:38.3%	C:72.7%[S:71.4%,D:1.3%],F:6.6%,M:20.7%	C:94.8%[S:92.6%,D:2.2%],F:2.4%,M:2.8%
BUSCO Tetrapoda_odb10 (n:5310)	C:91.4%[S:74.1%,D:17.3%],F:2.6%,M:6.0%	C:52.9%[S:51.6%,D:1.3%],F:4.6%,M:42.5%	C:68.6%[S:67.3%,D:1.3%],F:5.5%,M:25.9%	C:90.7%[S:88.7%,D:2.0%],F:2.8%,M:6.5%

Table S3. Classification and number of repeat elements masked in the aLitaur1.pri.20240424 genome assembly

	Number of Elements	Length (bp)	% of Sequence
Retroelements	1471787	895323119	18.76
SINEs	99574	17865329	0.37
Penelope	6445	756297	0.02
LINEs	604508	305980414	6.41
L2/CR1/Rex	321608	155360738	3.26
R1/LOA/Jockey	7496	524829	0.01
RTE/Bov-B	489	146499	0.00
L1/CIN4	101623	85326707	1.79
LTR elements	767705	571477376	11.98
BEL/Pao	820	1332796	0.03
Ty1/Copia	29659	10728688	0.22
Gypsy/DIRS1	474415	48646646	10.20
Retroviral	61135	26860128	0.56
DNA transposons	1521873	429272806	9.00
hobo-Activator	870063	305552506	6.40
Tc1-IS630-Pogo	370167	71958929	1.51
Piggybac	3787	4778127	0.10
Tourist/Harbinger	119369	20439618	0.43
Other (Mirage, P-element, Transib)	41886	3028514	0.06
Rolling-circles	17362	3319600	0.07
Unclassified	8289187	1724910956	36.15
Total interspersed repeats		3050263178	63.93
Small RNA	27293	6185734	0.13
Satellites	46828	9198436	0.19
Simple Repeats	105786	15407280	0.32

Table S4. Pairwise Euclidean (lower) and Nei's (D_{XY}) distances (upper) between each region.

	Region 1	Region 2	Region 3	Region 4
Region 1	-	0.0524	0.0282	0.0365
Region 2	33.2774	-	0.0285	0.0250
Region 3	24.3385	24.6740	-	0.0122
Region 4	27.6859	23.1506	16.0679	-

Table S5. The Q-profile genetic diversity results for each region and population.

Regions	Sites	N	m_0Da (sd_0Da)	m_1Da (sd_1Da)	m_2Da (sd_2Da)
Region 1	-	19	0.821 (0.384)	0.373 (0.240)	0.241 (0.174)
	UD09	10	0.708 (0.455)	0.339 (0.261)	0.221 (0.185)
	UD14	9	0.645 (0.479)	0.324 (0.273)	0.214 (0.192)
Region 2	-	43	0.980 (0.141)	0.409 (0.192)	0.259 (0.149)
	UD15	9	0.844 (0.363)	0.387 (0.226)	0.249 (0.165)
	UD18	5	0.633 (0.482)	0.328 (0.273)	0.217 (0.191)
	UD22	7	0.686 (0.464)	0.337 (0.263)	0.221 (0.185)
	UD33	9	0.747 (0.435)	0.356 (0.252)	0.231 (0.180)
	UD01	6	0.685 (0.464)	0.342 (0.263)	0.224 (0.185)
	UD01B	7	0.683 (0.465)	0.337 (0.264)	0.221 (0.186)
Region 3	-	39	0.987 (0.115)	0.424 (0.183)	0.269 (0.145)
	UD05B	9	0.839 (0.367)	0.394 (0.229)	0.255 (0.167)
	UD21B	10	0.875 (0.331)	0.401 (0.218)	0.258 (0.161)
	UD28	10	0.810 (0.392)	0.378 (0.239)	0.245 (0.173)
	UD29	10	0.870 (0.336)	0.399 (0.220)	0.257 (0.162)
Region 4	-	14	0.843 (0.364)	0.397 (0.236)	0.259 (0.172)
	UD04	7	0.645 (0.479)	0.323 (0.272)	0.213 (0.191)
	UD10	7	0.691 (0.462)	0.356 (0.270)	0.236 (0.191)

Table S6. Genetic diversity by site. Auto. H_O: autosomal observed heterozygosity; Unb. H_O: JeDi unbiased heterozygosity; Auto. H_E: autosomal expected heterozygosity; Auto. Pi: autosomal nucleotide diversity; Unb. Pi: JeDi unbiased nucleotide diversity; F_h: genome-wide estimates of heterozygosity. The standard errors are given in brackets.

Regions	Sites	N	Auto. H_O	Unb. H_O	Auto. H_E	Auto. Pi	Unb. Pi	F_h
Region 1	UD09	10	0.00136 (0.00001)	0.00212 (0.00005)	0.00138 (0.00001)	0.00145 (0.00001)	0.00208	0.1712 (0.0076)
	UD14	9	0.00127 (0.00001)	0.00199 (0.00003)	0.00126 (0.00001)	0.00133 (0.00001)	0.00198	0.1727 (0.0148)
Region 2	UD15	9	0.00154 (0.00001)	0.00203 (0.00004)	0.00158 (0.00001)	0.00167 (0.00001)	0.00232	0.0611 (0.0053)
	UD18	5	0.00125 (0.00001)	0.00207 (0.00006)	0.00119 (0.00001)	0.00132 (0.00001)	0.00205	0.1054 (0.0330)
	UD22	7	0.00129 (0.00001)	0.00208 (0.00003)	0.00135 (0.00001)	0.00146 (0.00001)	0.00212	0.1793 (0.0256)
	UD33	9	0.00143 (0.00001)	0.00229 (0.00004)	0.00142 (0.00001)	0.00150 (0.00001)	0.00218	0.1028 (0.0072)
	UD01	6	0.00133 (0.00001)	0.00233 (0.00003)	0.00132 (0.00001)	0.00145 (0.00001)	0.00218	0.1166 (0.0066)
	UD01B	7	0.00140 (0.00001)	0.00235 (0.00003)	0.00134 (0.00001)	0.00145 (0.00001)	0.00211	0.1075 (0.0041)
Region 3	UD05B	9	0.00156 (0.00001)	0.00220 (0.00009)	0.00164 (0.00001)	0.00174 (0.00001)	0.00245	0.0615 (0.0144)
	UD21B	10	0.00165 (0.00001)	0.00208 (0.00005)	0.00168 (0.00001)	0.00177 (0.00001)	0.00245	0.0318 (0.0048)
	UD28	10	0.00153 (0.00001)	0.00222 (0.00006)	0.00159 (0.00001)	0.00168 (0.00001)	0.00236	0.0953 (0.0336)
	UD29	10	0.00160 (0.00001)	0.00232 (0.00003)	0.00163 (0.00001)	0.00172 (0.00001)	0.00244	0.0276 (0.0054)
Region 4	UD04	7	0.00128 (0.00001)	0.00232 (0.00003)	0.00127 (0.00001)	0.00137 (0.00001)	0.00206	0.1724 (0.0118)
	UD10	7	0.00169 (0.00001)	0.00259 (0.00009)	0.00160 (0.00001)	0.00172 (0.00001)	0.00258	0.0350 (0.0131)

Table S7. QDiver diversity results of *U. daviesae*

Parameters	Regions	Sites	Diversity value	p value (Bartless's test)
γ = Q(GT)			0.988	
δ = Q(AR)			0.745	
σ = Q(WR)	Region 1		0.951	0.999
	Region 2		0.955	
	Region 3		0.955	
	Region 4		0.955	
β = Q(AP/WR)	Region 1		0.500	0.142
	Region 2		0.819	
	Region 3		0.729	
	Region 4		0.500	
α = Q(WP/WR)	Region 1	UD09	0.935	0.639
		UD14	0.932	
	Region 2	UD15	0.944	0.943
		UD18	0.880	
		UD22	0.929	
		UD33	0.944	
		UD01	0.917	
		UD01B	0.918	
	Region 3	UD05B	0.944	0.399
		UD21B	0.930	
		UD28	0.935	
		UD29	0.945	
	Region 4	UD04	0.929	1.000
		UD10	0.929	

GT: total diversity, AR: among regions, WR: within regions, AP/WR: among sites within regions, WP/WR: with sites within regions.

Table S8. The number of each population has been selected under each N scenario.

Region	Site	N=1	N=2	N=3	N=4	N=5	N=6	N=7	N=8	N=9	N=10
Region 1	UD09	0	5	17	20	25	20	2	9	33	33
	UD14	0	4	7	21	44	56	26	34	39	39
Region 2	UD15	1	2	17	27	28	43	13	17	27	27
	UD18	0	5	13	6	12	24	8	17	22	22
	UD22	0	4	12	17	20	22	5	14	13	13
	UD33	0	7	24	28	31	40	25	34	34	34
	UD01	0	7	17	27	24	31	2	2	4	10
	UD01B	0	6	16	22	25	26	6	18	38	28
Region 3	UD21B	1	0	4	21	31	51	22	36	35	35
	UD28	0	0	23	33	27	24	17	28	25	25
	UD29	1	0	6	20	22	37	20	31	33	33
	UD05B	1	0	4	25	43	44	15	19	33	23
Region 4	UD04	0	6	24	26	31	20	3	10	29	29
	UD10	0	0	11	35	62	72	25	35	39	39

Scripts

Script S1. Scripts for the *Trimmomatic* step.

Trimmed to 68bp

java -jar trimmomatic-0.39.jar SE -phred64 ILLUMINACLIP: TruSeq3-SE.fa:2:30:10 LEADING:5 SLIDINGWINDOW:4:5 CROP:68 MINLEN:68

Script S2. Scripts for each run of the *populations* program for different analyses.

All other analyses except heterozygosity

-r 0.75

Minimum percentage of individuals in a population required to process a locus for that population

--max-obs-het 0.75

Specify a maximum observed heterozygosity required to process a nucleotide site at a locus (applied to the metapopulation).

--min-mac 3

##Specify a minimum minor allele count required to process a SNP (applied to the metapopulation)

--min-maf 0.05

Specify a minimum minor allele frequency required to process a nucleotide site at a locus (0 < min_maf < 0.5; applied to the metapopulation)

For autosomal heterozygosity

-R 1

Minimum percentage of individuals across populations required to process a locus.

Script S3. SNP data filtering process in R.

Steps	Settings
Genotyping Quality	GQ <- 25
Minimum average locus read depth and maximum	minCount <- 15 maxCount <- 80
Minimum average SNP read count	minSNP <- 4
Maximum frequency of hets to remove potential paralogues	maxHet <- 0.6
Allele balance - max difference in read depth between REF and SNP alleles	maxDiff <- 0.6
Overall callrate light filtering	minCall <- 0.6
Proportion of populations not in HWE	minHWE <- 0.1
Reproducibility	rp <- 98
MAF light filtering	maf <- 0.005
Individual callrate	threshold = 0.9
Locus callrate	threshold = 0.9
MAF	MAC3 <- 3 / (2 * (sample numbers)

References

Andrews, S. (2014). *FastQC A quality control tool for high throughput sequence data*. Retrieved 21 September 2020 from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Batut, B., Hiltemann, S., Bagnacani, A., Baker, D., Bhardwaj, V., Blank, C., Bretaudeau, A., Brillet-Guéguen, L., Čech, M., Chilton, J., Clements, D., Doppelt-Azeroual, O., Erxleben, A., Freeberg, M. A., Gladman, S., Hoogstrate, Y., Hotz, H.-R., Houwaart, T., Jagtap, P.,…Grüning, B. (2018). Community-Driven Data Analysis Training for Biology. *Cell Systems*, *6*(6), 752-758.e751. https://doi.org/https://doi.org/10.1016/j.cels.2018.05.012

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/https://doi.org/10.1093/bioinformatics/btu170

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170-175. https://doi.org/10.1038/s41592-020-01056-5

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451-9457. https://doi.org/https://doi.org/10.1073/pnas.1921046117

Formenti, G., Abueg, L., Brajuka, A., Brajuka, N., Gallardo-Alba, C., Giani, A., Fedrigo, O., & Jarvis, E. D. (2022). Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, *38*(17), 4214-4216. https://doi.org/10.1093/bioinformatics/btac460

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484-w492. https://doi.org/https://doi.org/10.1093/nar/gkad326

Haas, B. J. (2022). *TransDecoder (find coding regions within transcripts)*. In https://github.com/TransDecoder/TransDecoder

Hiltemann, S., Rasche, H., Gladman, S., Hotz, H.-R., Larivière, D., Blankenberg, D., Jagtap, P. D., Wollmann, T., Bretaudeau, A., Goué, N., Griffin, T. J., Royaux, C., Le Bras, Y., Mehta, S., Syme, A., Coppens, F., Droesbeke, B., Soranzo, N., Bacon, W.,…Batut, B. (2023). Galaxy Training: A powerful framework for teaching! *PLOS Computational Biology*, *19*(1), e1010752. https://doi.org/10.1371/journal.pcbi.1010752

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907-915. https://doi.org/https://doi.org/10.1038/s41587-019-0201-4

Lariviere, D., Ostrovsky, A., Gallardo, C., Syme, A., Abueg, L., Pickett, B., Formenti, G., Sozzoni, M., & Nekrutenko, A. (2024). *VGP assembly pipeline: Step by Step (Galaxy Training Materials)*. Retrieved April 29 2024 from https://training.galaxyproject.org/training-material/topics/assembly/tutorials/vgp_genome_assembly/tutorial.html

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads [next generation sequencing; small RNA; microRNA; adapter removal]. *2011*, *17*(1), 3. https://doi.org/10.14806/ej.17.1.200

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq

reads. *Nature Biotechnology*, *33*(3), 290-295. https://www.nature.com/articles/nbt.3122.pdf

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212. https://doi.org/https://doi.org/10.1093/bioinformatics/btv351

Solovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology*, *7*(S1), S10.

Solovyev, V. V. (2002). Finding genes by computer: Probabilistic and discriminative approaches. In Y. X. Tao Jiang, Michael Q. Zhang (Ed.), *Current Topics in Computational Molecular Biology* (pp. 201-248). MIT Press.

The Galaxy Community. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, *50*(W1), W345-W351. https://doi.org/10.1093/nar/gkac247

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Consortium, D. T. o. L., Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy, S. A. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads. *bioRxiv*, 2022.2012.2023.521667. https://doi.org/https://doi.org/10.1101/2022.12.23.521667

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., & Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Research*, *41*(6), e74-e74. https://doi.org/10.1093/nar/gkt006