An endangered flightless grasshopper with strong genetic structure maintains population genetic variation despite extensive habitat loss

Hoffmann, Ary A.[1] ORCID 0000-0001-9497-7645

White, Vanessa L.[2] ORCID 0000-0003-2544-9069

Jasper, Moshe [1] ORCID 0000-0003-4541-3223

Yagui, Hiromi[1] ORCID 0000-0001-5683-5873

Sinclair, Steve J.[3] ORCID 0000-0002-4282-1021

Kearney, Michael R. [2] ORCID 0000-0002-3349-8744

[1]Bio21 Institute, School of BioSciences, The University of Melbourne, Victoria 3010, Australia

[2]School of BioSciences, The University of Melbourne, Victoria 3010, Australia

[3]Arthur Rylah Institute, Department of Environment, Land, Water and Planning, Heidelberg, Victoria, 3084, Australia

Correspondence: Ary Hoffmann, Bio21 Institute, School of BioSciences, The University of Melbourne, Victoria 3010, Australia

ary@unimelb.edu.au

Running head: threatened grasshopper genetics

Keywords: *Keyacris*, morabine, small population area, fragmentation, isolation by distance, grassland

Article impact statement: Endangered *Keyacris* grasshopper populations are genetically unique and even those from small fragments deserve conservation.

Acknowledgments

This research was supported by the Australian Research Council Discovery Grant DP190100990. We thank Craig Moritz and Ted Deveson for support during the field work and Nick Bell for formatting assistance.

Abstract

Conservation research is dominated by vertebrate examples but the shorter generation times and high local population sizes of invertebrates may lead to very different management strategies. Here we investigate the genetic structure of an endangered flightless grasshopper, *Keyacris scurra*, which was used in classical evolutionary studies in the 60s. It had a wide distribution across New South Wales (NSW) and Victoria in pre-European times but has now become threatened because of land clearing for agriculture and other activities. We revisited remnant sites of *K. scurra*, with populations now restricted to only one area in Victoria and a few small patches in NSW and the Australian Capital Territory (ACT). Using DArTseq to generate SNP markers as well as mtDNA sequence data, we show that the remaining Victorian populations in an isolated valley are genetically distinct from the NSW populations and that all populations tend to be genetically unique, with large F_{ST} values up to 0.8 being detected for the SNP datasets. We also find that, with one notable exception, the NSW/ACT populations separate genetically into previously described chromosomal races (2n = 15 vs. 2n = 17). Isolation by distance was detected across both the SNP and mtDNA data sets, and there was substantial differentiation within chromosomal races. Genetic diversity as measured by heterozygosity was not correlated with the size of remaining habitat where the populations were found, with high variation present in some remnant cemetery sites. However, inbreeding correlated negatively with estimated habitat size at 25-500 m patch radius. These findings emphasize the importance of small habitat areas in conserving genetic variation across the species, and they highlight populations suitable for future translocation efforts.

Introduction

As with other animals, terrestrial invertebrates are increasingly being threatened by habitat destruction, climate change, invasive species, pesticides and other threats connected to human activities (Hafernik and Hafernik, 1992, Wagner and Van Driesche, 2010, Black and Vaughan, 2009), and population declines as well as extinction rates over the last few hundred years can match those of vertebrates and vascular plants (Thomas and Morris, 1994, Leidner and Neel, 2011). Despite this rate of decline and the role of threatened invertebrates in essential ecosystem services (Wagner and Van Driesche, 2010, Kim, 1993), there is still only a limited focus on their conservation around the world, including in Australia (Sands, 2018). Part of the problem resides in taxonomic issues, with many species undescribed and/or lacking basic taxonomic information (New and Sands, 2004, Kim, 1993), leading to the risk that some species may face extinction even before they are known. Yet in Australia, many threatened invertebrates represent unique evolutionary lineages that form an important component of biodiversity (Cranston, 2010).

Although genetic data is critical in informing conservation strategies, helping to resolve taxonomic issues, defining patterns of connectedness across populations, and assessing the adaptive capacity of populations to future environmental changes, very little genetic data exists for threatened terrestrial invertebrate species. Older work using mtDNA, AFLPs, allozymes, microsatellites and other markers has been used to define management units for conservation (e. g. Roitman et al., 2017, Rotheray et al., 2012), examine gene flow and historical processes (e.g. Vogler et al., 1993, Crawford et al., 2011), and explore the consequences of management actions such as insect translocations (Witzenberger and Hochkirch, 2008). There are so far relatively few attempts to integrate modern genomic approaches based on genome wide SNPs or genome resequencing into invertebrate conservation efforts (e.g. Dupuis et al., 2020, Chen et al., 2017). These approaches can provide very detailed information on patterns of gene flow, hybridization and evolutionary potential in threatened species that can guide management actions (Allendorf et al., 2010).

Here we provide SNP and mtDNA based analysis of populations of an endangered morabine grasshopper, *Keyacris scurra* (formerly known as *Moraba scurra*). Morabines represent a unique group of Australian flightless grasshoppers, with a characteristic matchstick-like appearance. The morabines consist of ~250 species and 41 genera found across Australia on a range of plant types including grasses, trees and shrubs (Blackith and Blackith, 1969, Key, 1977).

Keyacris scurra is one of the better-known morabines. The genus *Keyacris* was named after the entomologist Ken Key (Day and Rentz, 2004) and studied by the Australian geneticist and evolutionary biologist Michael White (White, 1956, White et al., 1963). The species was used in pioneering work on adaptive genetic polymorphisms in collaboration with the American evolutionary biologist Richard Lewontin (e.g. Lewontin and White, 1960, White et al., 1963) which led to an ongoing debate about population processes affecting chromosomal polymorphisms and particularly the notion of adaptive landscapes (reviewed in Grodwohl, 2017).

The species was found in northeastern Victoria in the wheat/grazing belt and in the wheat/grazing belt of eastern NSW as far north as Goulburn. White (1956) noted that *Keyacris scurra* was already threatened when he indicated that they consist of “relatively minute "islands" in the general area within which the species occurs”. Most of these "ecological islands" studied by White were places which had escaped agricultural intensification and regular grazing, such as small rural cemeteries, small reserves and railway cuttings (Rowell and Crawford, 1995).

The species appears confined to habitats of a special type in which the tall perennial grass, *Themeda triandra*, usually predominates. This once dominant grass is removed by cropping and is grazing sensitive, and it now only dominates relict areas,

which are often also refuges for other similarly sensitive plant species (Dorrough and Scroggie 2008), including many daisies that *K. scurra* requires for food (White, 1956). Suitable habitats occur in grassland, savannah woodland and on the ecotones between the latter habitats and both "dry" and "wet" sclerophyll forest. *K. scurra* is an overwintering species, hatching in summer and with a univoltine life cycle. The species is unfortunately found within one of the most modified regions of Australia (Glanznig, 1995) where very little remnant habitat remains. The species has very limited dispersal ability due to its flightless habit. The main threat likely remains the management of vegetation (for instance cemeteries are now managed by repeated mowing close to ground level which destroys the habitat of *K. scurra*).

Here we provide information on the genetic structure of *K. scurra* by resampling areas where the species was previously found as well as new areas that appear to have suitable vegetation. We undertake genetic comparison across the species' range and find a high level of genetic differentiation across regions even when these are relatively close together, evidence for genetic isolation by distance in both nuclear and mtDNA markers, evidence of inbreeding in some populations and some genetic differentiation patterns unrelated to the chromosomal constitution of populations. We show that genetic variability varies among populations, and that inbreeding is related to the extent of habitat available in the proximity of the sampling sites. However, even very small habitat patches may support populations with valuable genetic resources. This information is critical in developing plans around the conservation of the species and emphasizes the importance of maintaining small habitat patches when conserving genetic variation across the species.

Methods

Sampling sites

Populations of *K. scurra* were collected from 17 locations from NSW and Victoria in 2019 for molecular work following an extensive survey to map the current distribution of this species (Fig. 1). These samples had been secured prior to the listing of the species as "endangered" in NSW (https://www.environment.nsw.gov.au/resources/threatenedspecies/determinations/CAMKeysMatchstickGrasshopperESPD.pdf) and with the approval of the Dept of the Environment, Land, Water and Planning in Victoria following the rediscovery of the species from Omeo. Only a few individuals were collected from the smaller populations (particularly at Bungonia, Gundagai South Cemetery, Windellama North) (Table 1). Grasshoppers were collected with aspirators across an area of >20 m^2 and preserved individually in 100% ethanol in Eppendorf tubes. They were then brought back to the laboratory for DNA processing.

CO1 PCR and sequencing

A total of 59 individuals was screened from NSW (13 populations, 43 individuals) and the ACT and Victoria (3 pops, 8 individuals each) with all populations in Table 1 represented. DNA was extracted using a Chelex® 100 Resin (Bio-Rad Laboratories, Hercules, CA) method on the upper half of a grasshopper hind limb. Tissue was crushed with 2 X 3 mm glass beads and 200 µL of 5% (w/v) Chelex® 100 suspension using a mixer mill. Extractions were incubated for 2 h at 60 °C with 5 µL proteinase K (20 mg/mL) (Roche Diagnostics Australia, Pty Limited, Castle Hill NSW, Australia) and heated at 90 °C for 10 min. Prior to polymerase chain reaction (PCR) amplification, extractions were spun at 14,000 rpm for 2 min, and DNA in solution was removed from just above the Chelex® resin.

PCR was performed to amplify approximately 700 base pairs (bp) of the mitochondrial cytochrome oxidase subunit 1(COI) gene using the primer combination LCO1490: 5'-ggtcaacaaatcataaagatattgg-3' and HCO2198: 5'-taaacttcagggtgaccaaaaaatca-3' (Folmer et al., 1994). A 50 µl reaction volume was used with 0.2 mM dNTPs, 0.1 mg/ml bovine serum albumin (New England Biolabs, Ipswich, MA), 1 X reaction buffer, 2.0 mM MgCl2, 2 units of Taq polymerase (New England Biolabs, Ipswich, MA), 0.20 µM forward and reverse primers and 4 µl of 1:10 diluted template DNA.
The PCR amplification profile for COI consisted of an initial denaturing step at 95 °C for 4 min (1 cycle), 40 cycles of denaturation at 95 °C for 45 sec, annealing at 53 °C for 45 sec and extension at 72 °C for 1 min then a final extension step at 72 °C for 5 min (1 cycle). All PCR reactions were conducted in Eppendorf Mastercycler S Gradient machines. PCR amplicons were sequenced from both directions using Sanger sequencing (Macrogen Inc., Geumcheongu, Seoul, South Korea), and the chromatograms were analysed using Geneious version 11.1.4 (http://www.geneious.com).

DArT-Seq™ processing

A high-throughput genotyping method using the DArT-Seq™ technology at Diversity Arrays Pty Ltd (Canberra, Australia) was employed. Here, complexity reduction is used to enrich nuclear genome representations with active genes and low copy sequences through combinations of restriction enzymes and reduction methods (https://www.diversityarrays.com/technology-and-resources/dartseq/). Implicit fragment size selection and next-generation sequencing of representations is subsequently performed with HiSeq2000 (Illumina, USA) (Georges et al., 2018, Kilian et al., 2012). This technology was considered appropriate for *K. scurra* to overcome sequencing problems associated with large genomes and high levels of repetitive DNA, gene duplications and pseudogenes which were expected in this orthopteran species (e.g. Palacios-Gimenez and Cabral-de-Mello, 2015, Wang et al., 2014).

Grasshopper hind limb tissue (upper half) was supplied to Diversity Arrays Pty Ltd (Canberra, Australia) for high-density (approx. 2.5 million sequences/sample used in marker calling) DArT-Seq™ assay. Eight samples were first tested with multiple restriction enzyme combinations, and an "optimal" set was determined based on the fraction of the genome represented controlling average read depth and a number of polymorphic loci (https://www.diversityarrays.com/technology-and-resources/dartseq/). DArT-Seq™ DNA extraction and sequencing and SNP genotyping methods are explained in detail elsewhere (Kilian et al., 2012).

Bioinformatics for nuclear data

Reads processing. Following adaptor and barcode sequence trimming, raw fastq files of DArT-Seq™ samples (HiSeq processing) were processed with the STACKs denovo_map.pl pipeline (version 2.0b, Catchen et al. 2013), as there is no reference genome for *K. scurra*. This pipeline assembles loci *de novo* within each individual, combines these loci into a catalogue, matches individuals to the catalogue, then performs SNP calling and haplotype phasing. Program settings were customised to allow four mismatches between sequence stacks within individuals (M = 4) and the same number between stacks between individuals (n = 4). Genotyped SNPs were output to the VCF file format for read filtering.

The same dataset was also processed in parallel by Diversity Arrays Pty Ltd (Canberra, Australia) (Kilian et al., 2012): the resulting DArT-Seq™ matrix of nuclear SNP loci by individuals was used downstream for phylogenetic analyses.

Read filtering. Before further SNP filtering was carried out, one low-quality individual was removed from the dataset. To investigate the effects of sample selection and filtering assumptions on downstream measures of genetic diversity and differentiation, we created a range of differing datasets at the SNP filtering stage with VCFtools (Danecek et al. 2011). In all cases, only loci with less than 5% missing data across those individuals included within a dataset were retained.

Datasets were constructed by varying individuals per population, minor allele count (in the case of heterozygosity) and whether all SNPs or one SNP per sequence were included in the analysis. Minor allele count (MAC) rather than minor allele frequency was used as recently advocated for population structure analysis (Linck and Battey, 2019). A minimum minor allele count (MCAC) of 3 was used when assessing population structure, given that this filter appears to be optimal for this purpose particularly when using programs like STRUCTURE (Linck and Battey, 2019). However, a MAC of 1 was considered appropriate when characterizing variability within populations; if samples from populations are small, a MAC=3 filter would exclude a lot of SNPs and provide a potential bias towards alleles that are variable in populations with larger samples. Note that researchers normally use the same

filter when characterizing population structure and variation within populations, whereas in our case we used different options when characterizing heterozygosity, although the two estimates of heterozygosity based on these approaches were correlated in our dataset (see below). We also investigated whether the selection of even numbers of individuals from each population (i.e. equivalent to the lowest number of individuals sampled from any population) affected our estimates by comparing estimates when all populations were reduced to a randomly selected 3 individuals (2 in the case of three of the populations where 3 individuals were not available, see Table 1).

For phylogenetic analyses, the above-mentioned DArT-Seq™ matrix was read into a genlight object (from the adegenet package, Jombart, 2008) for processing by DARTR (version 1.1.6, Gruber et al., 2018) including a similar filtering scheme to the one above, in sequence: (a) Callrate > 95 % (b) Repavg > 95 % (c) Read depth between 2 and 48 (d) filter secondaries loci (e) filter monomorphic loci.

Heterozygosity. In characterizing individual heterozygosity, the SNP analysis was done in two ways; either one SNP per (80 bp) locus was randomly selected from the data sets, or all SNPs were analysed across sequences. In addition, as noted above, we chose one dataset with all individuals retained – 'all individuals' - and another dataset where at most three individuals were retained from each population – 'even populations' (158 vs 50 individuals in each set). After including both the MAC=1 and MAC=3 filters, the number of SNPs retained in each of the 8 combinations considered are shown in Table S1. In estimating heterozygosity, we also made one further estimate by filtering the SNPs under MAC=1 but then with this filtered set of SNPs calculating heterozygosity based on all individuals in a population, to minimise data loss from considering only a subset of individuals.

All datasets were passed to the R 'adegenet' package as genind objects for further calculations of heterozygosity and other statistics. Sites with only one individual were excluded from population measures but were included in individual heterozygosity assessments. The 'Hs' function from 'adegenet' was used to calculate expected heterozygosity for populations. We also ran the 'gl.report.heterozygosity' function from 'dartR' with default settings to calculate observed heterozygosity and expected heterozygosity per population but found inconsistent results with the adegenet analysis and only report on the 'adegenet' analysis where the above filtering options were more easily defined.

Population structure analyses. The 'all individuals' and 'MAC = 3' dataset was used as the basis of a run of the program STRUCTURE in its ADMIXTURE mode (Pritchard et al., 2000). In these analyses we did not want to bias towards variable regions

and therefore filtered by randomly selecting only one SNP per sequence. Following inference of lambda, MCMC runs were completed with a burn-in of 10,000 and a further 100,000 repeats for parameter inference. K values between K=1 and K=10 were investigated, with ten runs per value of K. Results were passed to the program CLUMPAK for collation and summarising, and evaluated according to various K-inference procedures (Kopelman et al., 2015). A further run was conducted with the 'even populations' and 'MAC=3' dataset, under the same conditions and a separate lambda inference.

Datasets with all individuals were passed through PCO, PCA and DAPC multivariate analyses (via 'ade4' (Dray & Dufour, 2007) and 'adegenet' (Jombart, 2008). For principal components analysis, missing data were handled wherever possible by interpolation with the mean of the sampling location where the sample with missing data was found. The same principle was applied for the DAPC analyses.

AMOVA. A hierarchical analysis of molecular variance (AMOVA) was conducted using the "pegas" amova function implemented in "poppr" (Kamvar et al., 2014) for three levels: (a) individuals, (b) sites, and (c) regions. This last level was defined to include 6 levels: Cooma, Omeo, northern ACT (Mulligans, Hall), West (Gundagai, Wallendbeen, Boorowa), Northeast (Tarago, Windellama, Bungonia, Gundary, Goulburn) and Southeast (Bungendore, Kambah Pool, Burra). 1,000 permutations were conducted to test for significance across the differing levels.

F_{ST} and Isolation by distance based on SNPs. Pairwise F_{ST}s for each population were calculated via the "pairwise.WCfst" function in the R package "hierfstat" (Goudet, 2005). These were converted to a distance measure, $\frac{F_{ST}}{(1-F_{ST})}$, and compared to geographical distance between sites (in km) to check for isolation by distance. Mantel tests on distance matrices and genetic distance were also undertaken.

Phylogenetic analysis. As a further way of depicting variation among populations, we concatenated (a) sequence fragments (trimmed sequence tags with Single Nucleotide Polymorphisms or SNPs) and (b) SNPs only across loci for individuals where heterozygous positions were replaced by the standard ambiguity codes and exported these as fastA files using dartR (Gruber et al., 2018). Sequences were aligned and checked in MEGA (version 5.2, Tamura *et al*., 2011) and data from approach (b) was then converted to a nexus file format using the ape package (version 5.4, Paradis and Schliep, 2019) for R (version 1.3.959). The fastA file (a) and nexus file (b) were imported into CIPRES Science Gateway, version 3.3 (https://www.phylo.org) for Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic analyses respectively.

For ML, we used the program RAxML, version 8.2.12 (Stamatakis, 2014) on the "sequence fragments" = variant plus invariant data for improved branch length and topological accuracy in phylogenetic trees (Leaché et al., 2015). We assessed support for the best ML topology by performing 504 nonparametric bootstrap (BS) replicates using the autoMRE option with the GTR GAMMA site-rate substitution model. For BI, we used the program Mr. Bayes, version 3.2.7a (Ronquist and Huelsenbeck, 2003) on the "SNPs only" = variant data because of computational time constraints. We avoided uncertainty about what substitution model to use by sampling across the entire general time reversible (GTR) model space ("nst = mixed") and chose a "proportion of invariable sites + gamma" model of rate variation ("rates = invgamma") because this works well for many data sets (http://mrbayes.sourceforge.net/mb3.2_manual.pdf). Four independent Monte Carlo runs each with four Markov Chains (MCMC) were completed for 20,000,000 generations using random starting trees and a temperature parameter value of 0.1. Trees were sampled every 500 generations and the first 25% of generations were discarded as burn-in. The MCMC trace files were visualised and analysed in Tracer version 1.7.1 (Rambaut *et al*., 2018).

The best-scoring ML tree and consensus BI tree were imported into FigTree, version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree) for incorporation of branch length and support value (bootstrap for ML and probability for BI) information. Resultant files were then imported back into the R environment and final trees were read and visualised using the ggtree (version 2.0.4, Yu *et al*., 2017), ggplot2 (version 3.3.2, Wickham 2016) and treeio (version 1.10.0, Wang *et al*., 2020) packages.

mtDNA analysis

For all CO1 coding sequences, we first performed amino acid translations and searches for premature stop codons in Geneious version 11.1.4 (http://www.geneious.com) and confirmed sequence identity using BLASTn sequence homology searches against the National Center for Biotechnology Information (NCBI) nonredundant nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). A haplotype network was generated using PopART version 1.7 (http://popart.otago.ac.nz) and the Minimum Spanning network option. This program was considered appropriate because we had no sites with missing data.

Genetic and geographic distance matrices were created using the average number of base pair differences and latitude and longitude coordinates respectively for all pairwise population comparisons. A relationship among distance measures was investigated using Mantel tests performed with the "mantel.randtest" function in R package "ade4" (Dray & Dufour, 2007) with

1,000,000 permutations. Nuclear genetic distance was also compared to mtDNA distance with nuclear distance calculated as the Euclidean coancestry coefficient.

Vegetation analysis

To analyse the extent of available suitable habitat at each collection point, we compared between the ArcGIS raster of the National Vegetation Information System (NVIS) Version 5.1, a High-resolution Satellite Imagery, and a pre-existing likelihood model of intact native grassland (hereafter 'grassland model', Sinclair and White, unpublished). This last one was selected for the final analysis because its projection of 25 m x 25 m cell size, compared to the 100 m x 100 m of the NVIS raster, made it more concordant with the scale of our study. The grassland model was originally built for the Victorian State Government, and it was extrapolated to cover sites in NSW (Grassland model details provided in Supplementary material).

Given the low vagility of the *K. scurra*, we considered available habitat within radii 25 m, 50 m, 100 m, 250 m, and 500 m around each collection point using the ArcGIS 10.6 buffer tool. Available habitat was quantified by the sum of the pixel values of the grassland model within the relevant radius (Table S4). The computed available habitat measures were then correlated with the observed heterozygosity and inbreeding (F_{IS}).

Results

Population variation

A total of 195,172 SNPs markers were obtained from the DArT-Seq™ matrix of which 8,357 were retained following filtering at MAC=3, 13,518 were retained at MAC=1, while 5,608 were kept after the Diversity Arrays filtering used prior to the phylogenetic analysis. The numbers remaining after further filtering with 1 SNP per locus being retained and a similar number of individuals per population are given in Table S1. The four measures of heterozygosity variation we obtained across individuals with the MAC of 1 or 3 and including all individuals or equal numbers of individuals (N = 3 per population except where only 2 were available) were mostly highly correlated, regardless of whether we randomly selected one SNP per sequence or included all sequences (Table S2). We ran all further analyses with one SNP retained per sequence with all individuals considered and MAC=1, which maximises the retention of variation in populations where sample sizes are small, but nevertheless allows all individuals from a population to be used in computing heterozygosity (Figure 2 and Table 1). This measure also tended to be quite well correlated with the other measures of heterozygosity at MAC=1 (Table S2).

There was a significant difference in heterozygosity among populations when individual heterozygosities were considered (Figure 2 and P<0.001 for all ANOVAs regardless of how data was filtered prior to computing heterozygosity). Observed heterozygosity varied from expected heterozygosity in some cases, resulting in F_{IS} values that were substantial and positive in some several cases (e. g. Bungonia, Boorowa) (Table 1). These results suggest inbreeding in some populations. At the population level there was nevertheless a strong correlation between the observed and expected heterozygosities (r = 0.89).

Cooma and the Bungonia cemetery had particularly low levels of heterozygosity (Figure 2). However, the low level of heterozygosity at Bungonia may partly reflect inbreeding given the large F_{IS} and the fact that the expected heterozygosity is similar to values in several other populations (Table 1). Another cemetery population (Gundagai South) had a high level of heterozygosity (Fig. 2) while populations not involving cemeteries such as Omeo were relatively low in heterozygosity despite not showing inbreeding. Tarago was noteworthy in showing high variability in heterozygosity estimates which may reflect the inclusion of some inbred individuals (Fig. 2). These patterns suggest a range of genetic variability levels in *K. scurra* populations and high levels of variation even in some populations where suitable habitat appears limited.

Vegetation associations

We assessed whether genetic variation is related to the area of available habitat, measured at different scales. Such an association might be expected given that *K. scurra* is closely associated with relict patches of *Themeda* grasslands through its basic requirements for food and shelter (White, 1956). We found a significant negative relationship with F_{IS}, suggesting that populations in smaller habitat patches are more inbred (Figure 3, right column). This relationship was slightly stronger when smaller radii were used to define available habitat. There was a weak non-significant negative association with observed heterozygosity (Figure 3, left column).

Population structure

mtDNA

The network diagram of mtDNA variation (Fig. 4) indicates clear separation of the three Omeo sites from the other populations, with Cooma falling in between them and the other populations. For the remaining populations, two of the 2n=17 populations (Wallendbeen, Boorowa), as determined from the earlier cytological work, fall apart from most of the populations but the other

2n=17 sites (Gundagai South, referred to here as Gundagai, and Gundagai Cemetery) are not separated from the 2n=15 populations, while two other 2n=15 populations (Bungonia, Burra) also show some separation.

SNPs

The STRUCTURE plot for all individuals shows clear differentiation of regions at K=6, separating the North-East, South-East, West and Central regions as well as distinguishing the outlying Cooma and Omeo populations (Fig. 5). These patterns were clear regardless of whether all individuals were included in the analysis or whether an even number of individuals were selected from each population. Additional differentiation among sites was evident as K values were increased (Supplementary Fig. S2). An AMOVA (Table S3) indicated significant effects of regions and sites: 32% of variance is found within sites, 17% between sites within regions, and 51% between regions.

The DAPC analysis provided a clear picture of differentiation that matched the results of the STRUCTURE analysis. When all individuals and sites were included, there was a strong separation of the Omeo populations from the other areas across the two main axes which accounted for 32 and 27% of the variation, with the other sites falling into two main groups (Fig. 6a). Omeo and Cooma could both be separated based on the third axis (accounting for 17% of the variation) from all other populations (Fig. 6b). Note also how individuals from the same site tend to clump close together even when they are all in the same region. When the Omeo and Cooma populations are excluded, patterns for the other regions become clearer (Fig. 6c), with close associations between the ACT sites (Mulligans, Hall) and the 2n=17 NSW populations (Boorowa, Wallendbeen). Based on the nuclear markers, the 2n=17 Gundagai samples also fall close to the other 2n=17 populations, unlike for the mtDNA markers. Apart from the Omeo and Cooma populations, the other 2n=15 populations fell into two groups, but most individuals could still be allocated to sites, highlighting substantial differentiation across the sample sites even when these were quite close together. The exceptions seemed to be Kambah Pool and Burra, as well as Windellama, Bungonia and Gundary, where populations fall close together (Fig. 6). The F_{ST} values between sites (Fig. S1) were variable and in many cases substantial, being around 0.8 for comparisons with Omeo and Cooma populations and varying within the range 10-20% for the other populations. These substantial differences point to populations at sites that are often unique in terms of their nuclear composition even if there is overlap in mtDNA variation.

The phylogenetic analysis confirmed the uniqueness of the populations. Both the Bayesian tree (Fig. S3) and the ML tree (Fig. S4) showed that the individuals clustered into their collection sites. This included sites where the DAPC and STRUCTURE analyses suggested some overlap, such as Kambah Pool and Burra, and also Windellama, Bungonia and Gundary.

IBD analysis

For analyses of IBD, F_{ST}-derived distance (F_{ST} / (1 – F_{ST}) was regressed against geographical distance (Fig. 7) with the relationship being highly significant ($P < 0.001$, R^2 0.7729, slope 0.0059). Comparisons with the Cooma population were notable for falling above the line established from the other population comparisons, which is consistent with the high F_{ST} values for comparisons with this population (Fig. S1). A Mantel test indicated a significant association between geographic and genetic distance ($r = 0.7660$, $P < 0.001$, 1,000,000 permutations) consistent with the IBD regression analysis.

We also ran an IBD analysis on the mtDNA data by comparing the number of nucleotide differences between populations. A Mantel test on the mtDNA data indicated a significant association between geographic distance and nucleotide differences ($r = 0.876$, $P<0.001$) in agreement with the nuclear comparison. A Mantel test also indicated a positive association between the nuclear differences among populations and the mtDNA differences ($r = 0.693$, $P<0.001$).

Discussion

Although *Keyacris scurra* is an endangered species, it appears to have successfully persisted in small areas for many decades as long as suitable habitat has remained. We have recorded them persisting at Windellama and Gundagai South cemeteries despite these environments covering areas of only a few hectares and being surrounded by farmland. Moreover, both the latter sites have populations with high levels of genomic variability, suggesting little loss over the last few decades and implying that population sizes have been substantial enough to prevent much loss of genetic variation through genetic drift, although this remains a concern for the future. On the other hand, *K. scurra* has been lost from many other small remnant areas where they were recorded in the 1950s and 1960s, most likely through inappropriate site management. For instance, White (1963) performed evolutionary studies on the cemetery site at Murrumbateman in the ACT, where we failed to find the grasshopper despite multiple attempts to locate them there. We also visited many cemeteries in Victoria where *K. scurra* had been present in the 1950s (White, 1956), but specimens could not be found. In these areas, we found that *Themeda triandra* grassland has often persisted, but we believe that site management has removed the specific habitat elements required for *K. scurra* to persist; either via the exclusion of daisies through overgrowth of *Themeda* (Stuwe and Parsons 1977), or via structural modification of

the grass sward by regular mowing to keep cemeteries neat (Clayden et al., 2018). These observations show how tenuous survival can be for threatened species in agricultural landscapes, and how much they are subject to stochasticity and the unintended consequences of small-scale management decisions.

The populations at Windellama and Bungendore had previously been subjected to a deliberate translocation by White (1957) who introduced males from other populations in an attempt to alter the chromosomal constitution of the populations to explore the potential effects of natural selection on chromosome polymorphisms. Translocations are expected to boost genetic diversity and result in hybrid populations that are genetically distinct from parental populations as noted for the threatened field cricket, *Gryllus campestris* (Witzenberger and Hochkirch, 2008) and adders (Madsen et al., 1999). However here we find that the two populations map in multivariate space with nearby populations (Fig. 6) which were 19+ km away (Fig. 1) despite being separated from that site by unsuitable farmland, suggesting that the past deliberate translocation in this case had little impact on the uniqueness of the natural population nor boosted genetic variation.

At this stage, there is little support for the need to "genetically rescue" most populations of *K. scurra* from low levels of genetic variation within populations, perhaps with the exception of Bungonia which had low variation and showed relatively high inbreeding. Genetic rescue which involves the deliberate introduction of individuals across populations to overcome the deleterious effects of mutations that have become fixed in small populations (Weeks et al., 2011, Whiteley et al., 2015), it can be useful where there is strong evidence of a decline in genetic diversity and has been proposed as a useful approach for some threatened Australian insects (Roitman et al., 2017). However, with genetic variation persisting so far even in small areas, there is likely to be limited benefit from such an exercise. Instead we suspect that it is important to maintain the remaining variation across the range of the species given that there is very strong genetic differentiation among the populations. The F_{ST} values of up to 0.8 are extremely large and imply that populations often have different alleles predominating at loci that are polymorphic even within the same chromosomal form. Both selection and genetic drift may have contributed to this high level of differentiation. Thus our high resolution genetic data mirrors the chromosomal inversion polymorphisms observed by White (1956) which are often strongly differentiated between populations even within the chromosomal races.

The value of small reserves in preserving invertebrates (Hafernik and Hafernik, 1992) and plant biodiversity (Kendall et al. 2017) has been well recognized. However, small populations from reserves may lack genetic diversity which is linked to the adaptive capacity of populations (Hoffmann et al., 2017, Willi et al., 2006). In the case of *K. scurra* populations from restricted sites like the cemeteries assessed here, the high level of diversity still remaining at these sites suggests that they may, at least for now, be

able to counter environmental changes threatening populations into the future through evolutionary responses. Thus, while declines in invertebrate populations may well compare to those seen in plant and vertebrate populations (Leidner and Neel, 2011), remedial action to counter declines could be much easier through the recreation of small habitat areas. Habitats where *K. scurra* persist are quite variable and these will need to be managed in different ways to conserve *K. scurra*. For example, fire management practices could be modified to avoid burning or at least using more controlled burns during "at risk" life stages. During drought, browsing mammals (including native species) may need to be excluded to avoid overgrazing of *Themeda*. And cemetery management groups could be consulted to ensure that suitable habitat is fenced and not regularly mowed. Management of the Omeo populations will be particularly important since these appear to be the last remaining stronghold of the species in Victoria and are genetically quite distinct.

Our data on the associations between genetic variation and the area of available habitat are difficult to interpret without further investigation. We found a significant negative relationship between habitat area and F_{IS}, indicating elevated breeding between related individuals in smaller sites, but no association with genetic variation as measured by heterozygosity. A naïve expectation would be that observed heterozygosity would reveal the opposite: a positive relationship with reduced heterozygosity at smaller sites. Our data did not show this trend; this unexpected relationship may be an artefact of our relatively small sample size and method of habitat measurement, but it may also have biological foundations, and reflect past expansions and contractions in the distribution and population size of *K. scurra*. For example, rapid post-glacial expansion from refugia may have led to populations with a high residual heterozygosity, but a recent history of population fragmentation may be contributing to inbreeding in some populations. The high levels of observed heterozygosity compared to expected heterozygosity in some populations also warrants further investigation, particularly in relation to inversion polymorphisms which can directly affect heterozygosity (Kennington et al 2006). Patterns in current populations may display the legacy of past events and ongoing chromosomal dynamics which could be resolved by additional genomic resources so that (for instance) comparisons of heterozygosity could be made within and outside of inverted regions and population histories could be documented from linkage data.

Why we failed to correlate apparent habitat mapping and levels of SNP variation is unclear. Apart from the grassland model, we did attempt several other approaches such as using polygons from satellite images. In all cases, the correlation remained with the same tendency. Despite its limitations, the selected vegetation model is the most accurate geographic information system we currently have for habitat description. A key element to be checked in the future is to include not only *Themeda* but also *K. scurra* host species in the vegetation model. Also, working with smaller scales could be appropriate especially when dispersal

barriers are present, like in the Windellama cemetery, or when encountering non-native areas but equally qualified as *K. scurra* habitats, such as at Goulburn and Wallenbeen. This modification would decrease the vegetation values of the first two and increase the last ones.

The substantial genetic distances separating populations raise the issue of how to conserve diversity within the species. Clearly at this stage genetic uniqueness of populations is not associated with a loss of genetic diversity as is the case of marsupials and some other invertebrates (Weeks et al., 2016). It is important to conserve current levels of diversity across the landscape and the genomic data suggests that this can be achieved with relatively small areas. Increasing the number of fragments also helps protect against fires and other catastrophes that threaten Australia's insect species more generally (Sands, 2018), and provides nearby populations for future translocation efforts. The recreation of vegetation dominated by *Themeda* and a range of daisies is tractable if the high costs of seed can be overcome (Gibson-Roy and Delpratt 2015), so that the strategic creation of insurance populations of *K. scurra* is likely possible.

Beyond their conservation merit, the ability to create populations may also permit studies of fundamental biological questions. Following on from Michael White's early work, with the benefit of modern molecular tools, there are opportunities to further understand the evolutionary dynamics of *K. scurra* populations and reconsider some key evolutionary questions that were previously considered in this system. Early work by White argued that chromosomal rearrangements which could easily be scored in this grasshopper represented examples of heterozygote advantage and adaptive fitness interactions among chromosomal forms (White, 1957, White et al., 1963), which were interpreted as chromosomal forms being at different fitness peaks in an adaptive landscape (Lewontin and White, 1960). This was queried by others who argued for the importance of weak inbreeding (Allard and Wehrhahn, 1964) and changes in the selective advantage of different chromosomal arrangements across time (Colgan and Cheney, 1980) in accounting for patterns in these arrangements. By establishing populations with different combinations of chromosomal rearrangements from the same or different populations along climate gradients where the species occurs, and tracking changes in both the frequency of the rearrangements and their genomic content, it should be possible to gain insights into the extent to which rearrangements lock up adaptive genetic combinations, enhance or retard rates of evolutionary change, and change in fitness as a consequence of environmental variation. Such issues continue to be debated in the literature where *Drosophila* inversions in particular are regarded as important in climate adaptation (Rane et al., 2015, Kapun et al., 2016). Efforts to pursue these questions with *Keyacris scurra* would be greatly enhanced by developing an assembled and annotated genome of this and related morabine species.

LITERATURE CITED

421 ALLARD, R. & WEHRHAHN, C. 1964. A theory which predicts stable equilibrium for inversion polymorphisms in the
422 grasshopper *Moraba scurra*. *Evolution*, 18**,** 129-130.
423 ALLENDORF, F. W., HOHENLOHE, P. A. & LUIKART, G. 2010. Genomics and the future of conservation genetics.
424 *Nature Reviews Genetics*, 11**,** 697-709.
425 BLACK, S. H. & VAUGHAN, M. 2009. Endangered insects. *Encyclopedia of Insects*. Elsevier.
426 BLACKITH, R. & BLACKITH, R. M. 1969. Observations on the biology of some morabine grasshoppers. *Australian*
427 *Journal of Zoology*, 17**,** 1-12.
428 CATCHEN, J., HOHENLOHE, P. A., BASSHAM, S., AMORES, A., & CRESKO, W. A. 2013. Stacks: an analysis tool set for
429 population genomics. *Molecular Ecology*, 22, 3124–3140.
430 CHEN, Y. T., TSENG, H. Y., JENG, M. L., SU, Y. C., HUANG, W. S. & LIN, C. P. 2017. Integrated species delimitation and
431 conservation implications of an endangered weevil *Pachyrhynchus sonani* (Coleoptera: Curculionidae) in
432 Green and Orchid Islands of Taiwan. *Systematic Entomology*, 42**,** 796-813.
433 CLAYDEN, A., GREEN, T., HOCKEY, J. & POWELL, M. 2018. Cutting the lawn – Natural burial and its contribution to the
434 delivery of ecosystem services in urban cemeteries. *Urban Forestry & Urban Greening*, 33**,** 99-106.
435 COLGAN, D. & CHENEY, J. 1980. The inversion polymorphism of *Keyacris scurra* and the adaptive topography.
436 *Evolution*, 34**,** 181-192.
437 CRANSTON, P. S. 2010. Insect biodiversity and conservation in Australasia. *Annual Review of Entomology*, 55**,** 55-75.
438 CRAWFORD, L. A., DESJARDINS, S. & KEYGHOBADI, N. 2011. Fine-scale genetic structure of an endangered population
439 of the Mormon metalmark butterfly (*Apodemia mormo*) revealed using AFLPs. *Conservation Genetics*, 12**,**
440 991-1001.
441 DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G.,
442 MARTH, G. T., SHERRY, S. T., MCVEAN, G., DURBIN, R., & 1000 GENOMES PROJECT ANALYSIS GROUP. 2011.
443 The variant call format and VCFtools. Bioinformatics (Oxford, England), 27(15), 2156–2158.
444 DAY, M. F. C. & RENTZ, D. C. F. 2004. Kenneth Hedley Lewis Key 1911-2002. *Historical Records of Australian Science*,
445 15**,** 65–76.

DORROUGH, J. & SCROGGIE, M.P. 2008. Plant responses to agricultural intensification. *Journal of Applied Ecology*, 45: 1274-1283.

DRAY, S. & DUFOUR, A.B. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), pp.1-20.

DUPUIS, J. R., GEIB, S. M., OSBORNE, K. H. & RUBINOFF, D. 2020. Genomics confirms surprising ecological divergence and isolation in an endangered butterfly. *Biodiversity and Conservation*, 1-25.

GEORGES, A., GRUBER, B., PAULY, G. B., WHITE, D., ADAMS, M., YOUNG, M. J., KILIAN, A., ZHANG, X., SHAFFER, H. B. & UNMACK, P. J. 2018. Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short−necked turtles (Chelidae: Emydura) of eastern Australia. *Molecular Ecology*, 27, 5195-5213.

GIBSON-ROY, P. & DELPRATT, J., 2015. The restoration of native grasslands.:'Land of Sweeping Plains: Managing and Restoring the Native Grasslands of South-eastern Australia'.(Eds N. Williams, A. Marshall and J. Morgan.) pp. 331−387.

GLANZNIG, A. 1995 Native vegetation clearance, habitat loss and biodiversity decline: an overview of recent native vegetation clearance in Australia and its implications for biodiversity. *Biodiversity Series No. 6.* Canberra, ACT: Dept. of the Environment, Sport and Territories.

GRODWOHL, J. B. 2017. Natural selection, adaptive topographies and the problem of statistical inference: the *Moraba scurra* controversy under the microscope. *Journal of the History of Biology*, 50, 753-796.

GRUBER, B., UNMACK, P. J., BERRY, O. F. & GEORGES, A. 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources,* 18, 691-699.

HAFERNIK & HAFERNIK, J. 1992. Threats to invertebrate biodiversity: implications for conservation strategies. In: Fiedler P.L., Jain S.K. (eds) *Conservation Biology*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-6426-9_7.

HOFFMANN, A. A., SGRÒ, C. M. & KRISTENSEN, T. N. 2017. Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution*, 32, 506-517.

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

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*, p.e281.

KAPUN, M., FABIAN, D. K., GOUDET, J. & FLATT, T. 2016. Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 33**,** 1317-1336.

KENDAL, D., ZEEMAN, B.J., IKIN, K., LUNT, I.D., MCDONNELL, M.J., FARRAR, A., PEARCE, L.M. & MORGAN, J.W. 2017. The importance of small urban reserves for plant conservation. *Biological Conservation*, 213, 146-153.

KENNINGTON, W.J., PARTRIDGE, L. & HOFFMANN, A.A. 2006. Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion In (3R) Payne in *Drosophila melanogaster* are indicative of coadaptation. *Genetics*, 172, 1655-1663.

KEY, K. 1977. The genera and species of the tribe Morabini (Orthoptera: Eumastacidae: Morabinae). *Australian Journal of Zoology*, 25**,** 499-565.

KILIAN, A., WENZL, P., HUTTNER, E., CARLING, J., XIA, L., BLOIS, H., CAIG, V., HELLER-USZYNSKA, K., JACCOUD, D. & HOPPER, C. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics*. Springer.

KIM, K. C. 1993. Biodiversity, conservation and inventory: why insects matter. *Biodiversity & Conservation*, 2**,** 191-214.

KOPELMAN, N. M., MAYZEL, J., JAKOBSSON, M., ROSENBERG, N. A. & MAYROSE, I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15**,** 1179-1191.

LEACHÉ, A. D., BANBURY, B. L., FELSENSTEIN, J., DE OCA, A. N.-M. & STAMATAKIS, A. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 64**,** 1032-1047.

LEIDNER, A. K. & NEEL, M. C. 2011. Taxonomic and geographic patterns of decline for threatened and endangered species in the United States. *Conservation Biology*, 25**,** 716-725.

497 LEWONTIN, R. & WHITE, M. 1960. Interaction between inversion polymorphisms of two chromosome pairs in the
498 grasshopper, *Moraba scurra*. *Evolution*, 116-129.
499 LINCK, E. & BATTEY, C. 2019. Minor allele frequency thresholds strongly affect population structure inference with
500 genomic data sets. *Molecular Ecology Resources*, 19, 639-647.
501 MADSEN, T., SHINE, R., OLSSON, M. & WITTZELL, H. 1999. Restoration of an inbred adder population. *Nature*, 402,
502 34-35.
503 NEW, T. R. & SANDS, D. P. 2004. Management of threatened insect species in Australia, with particular reference to
504 butterflies. *Australian Journal of Entomology*, 43, 258-270.
505 PALACIOS-GIMENEZ, O. M. & CABRAL-DE-MELLO, D. C. 2015. Repetitive DNA chromosomal organization in the
506 cricket *Cycloptiloides americanus*: a case of the unusual X(1)X(2)0 sex chromosome system in Orthoptera.
507 *Molecular Genetics and Genomics*, 290, 623-631.
508 PARADIS, E. & SCHLIEP, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.
509 *Bioinformatics*, 35(3), pp.526-528.
510 PRITCHARD, J.K., STEPHENS, M. & DONNELLY, P., 2000. Inference of population structure using multilocus genotype
511 data. *Genetics*, 155, 945-959.
512 RAMBAUT, A., DRUMMOND, A.J., XIE, D., BAELE, G. & SUCHARD, M.A., 2018. Posterior summarization in Bayesian
513 phylogenetics using Tracer 1.7. *Systematic Biology*, 67, 901.
514 RANE, R. V., RAKO, L., KAPUN, M., LEE, S. F. & HOFFMANN, A. A. 2015. Genomic evidence for role of inversion 3 RP of
515 *Drosophila melanogaster* in facilitating climate change adaptation. *Molecular Ecology*, 24, 2423-2432.
516 ROITMAN, M., GARDNER, M. G., NEW, T. R., NGUYEN, T. T., ROYCROFT, E. J., SUNNUCKS, P., YEN, A. L. & HARRISSON,
517 K. A. 2017. Assessing the scope for genetic rescue of an endangered butterfly: the case of the Eltham copper.
518 *Insect Conservation and Diversity*, 10, 399-414.
519 RONQUIST, F. & HUELSENBECK, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models.
520 *Bioinformatics*, 19, 572-1574.

ROTHERAY, E., LEPAIS, O., NATER, A., KRÜTZEN, M., GREMINGER, M., GOULSON, D. & BUSSIÈRE, L. 2012. Genetic variation and population decline of an endangered hoverfly *Blera fallax* (Diptera: Syrphidae). *Conservation Genetics*, 13**,** 1283-1291.

ROWELL, A. & CRAWFORD, I. 1995. *A survey of the morabine grasshopper Keyacris scurra (Rehn) in the ACT*, CSIRO.

SANDS, D. P. 2018. Important issues facing insect conservation in Australia: now and into the future. *Austral Entomology*, 57**,** 150-172.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30**,** 1312-1313.

STUWE, J. & PARSONS, R.F. 1977. *Themeda australis* grasslands on the Basalt Plains, Victoria: floristics and management effects. *Australian Journal of Ecology*, 2,467-476.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729

THOMAS, J. & MORRIS, M. 1994. Patterns, mechanisms and rates of extinction among invertebrates in the United Kingdom. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344**,** 47-54.

VOGLER, A. P., DESALLE, R., ASSMANN, T., KNISLEY, C. B. & SCHULTZ, T. D. 1993. Molecular population genetics of the endangered tiger beetle *Cicindela dorsalis* (Coleoptera: Cicindelidae). *Annals of the Entomological Society of America*, 86**,** 142-152.

FOLMER, O. BLACK, M. HOEH, W. LUTZ, R. & VRIJENHOEK, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3**,** 294-9.

WAGNER, D. L. & VAN DRIESCHE, R. G. 2010. Threats posed to rare or endangered insects by invasions of nonnative species. *Annual Review of Entomology*, 55**,** 547-568.

WANG, X., FANG, X., YANG, P., JIANG, X., JIANG, F., ZHAO, D., LI, B., CUI, F., WEI, J. & MA, C. 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, 5**,** 2957.

WANG, L.G., LAM, T.T.Y., XU, S., DAI, Z., ZHOU, L., FENG, T., GUO, P., DUNN, C.W., JONES, B.R., BRADLEY, T. & ZHU, H. 2020. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution*, *37*, 599-603.

WEEKS, A. R., SGRO, C. M., YOUNG, A. G., FRANKHAM, R., MITCHELL, N. J., MILLER, K. A., BYRNE, M., COATES, D. J., ELDRIDGE, M. D., SUNNUCKS, P., BREED, M. F., JAMES, E. A. & HOFFMANN, A. A. 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications*, 4**,** 709-725.

WEEKS, A. R., STOKLOSA, J. & HOFFMANN, A. A. 2016. Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in Zoology*, 13**,** 31.

WHITE, M. 1957. Cytogenetics of the grasshopper *Moraba scurra*. 1. Meiosis of interracial and Interpopulation hybrids. *Australian Journal of Zoology*, 5**,** 285-304.

WHITE, M. J., ANDREW, L. E. & LEWONTIN, R. C. 1963. Cytogenetics of grasshopper *moraba scurra* .7. Geographic-variation of adaptive properties of inversions. *Evolution*, 17**,** 147-&.

WHITE, M. J. D. 1956. Adaptive chromosomal polymorphism in an Australian grasshopper. *Evolution*, 10**,** 298-313.

WHITELEY, A. R., FITZPATRICK, S. W., FUNK, W. C. & TALLMON, D. A. 2015. Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30**,** 42-9.

WICKHAM, H. 2016. *ggplot2: elegant graphics for data analysis*. Springer.

WILLI, Y., BURSKIRK, J. V. & HOFFMANN, A. A. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics* 37**,** 433-458.

WITZENBERGER, K. & HOCHKIRCH, A. 2008. Genetic consequences of animal translocations: A case study using the field cricket, *Gryllus campestris* L. *Biological Conservation*, 141**,** 3059-3068.

YU G, SMITH D, ZHU H, GUAN Y & LAM T. T. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8, 28-36.

Figure legends

571 **Figure 1**. Map of sites surveyed for molecular variation. These sites encompass most of the current known fragmented
572 distribution of *Keyacris scurra*. Singletons from two additional sites were included in the molecular survey: a site close to
573 Gundagai ("Gundagai Cemetery") and a site close to Windellama ("Windellama North").

574 **Figure 2.** Box plots for individual heterozygosity by location (for selected SNPs, all individuals prefiltered to MAC = 1 from data
575 where population with an even number of individuals were sampled, but then computed for all individuals from a population).
576 Note that populations are ordered to match the STRUCTURE analysis below. Gundagai Cemetery and Windellama North are not
577 included here because they are represented by singletons.

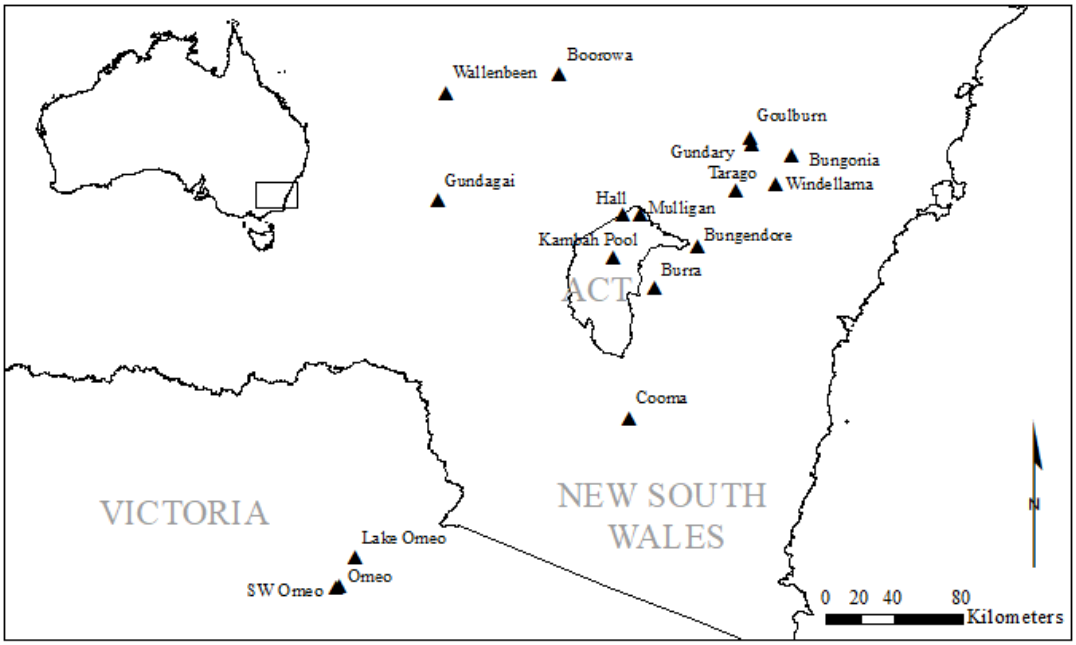
578 **Figure 3.** Association between a likelihood model of intact native grassland (Sinclair and White, unpublished) at different spatial
579 scales ("buffers") and observed heterozygosity (left column) or F_{IS} (right column). Dots reflect individual sites and are presented
580 with correlation coefficients (*r)* and *P* values.

581 **Figure 4**. Variation in the COI gene sequence across *K. scurra* as depicted by a network diagram. The numbers of nucleotide
582 changes are indicated in brackets. The size of the coloured areas reflects the number of haplotypes and branch lengths reflect
583 the number of nucleotide changes.

584 **Figure 5.** STRUCTURE plots for (a) all individuals and (b) even populations at K=6 (best supported by modified Evanno method).

585 **Figure 6.** DAPC of *Keyacris* scurra individuals with Omeo and Cooma included along the two main linear discriminant (LD) axes
586 (Fig. 6a) and the first and third axes (Fig. 6b) and when these populations are excluded (Fig. 6c). (N = 158, 45 PCs, RMSE 0.043
587 when all sites included, N=131, 30 PCs, RMSE = 0.029 when Cooma and Omeo excluded).

588 **Figure 7.** Correlation of geographical distance with F_{ST}-derived distance between populations



589

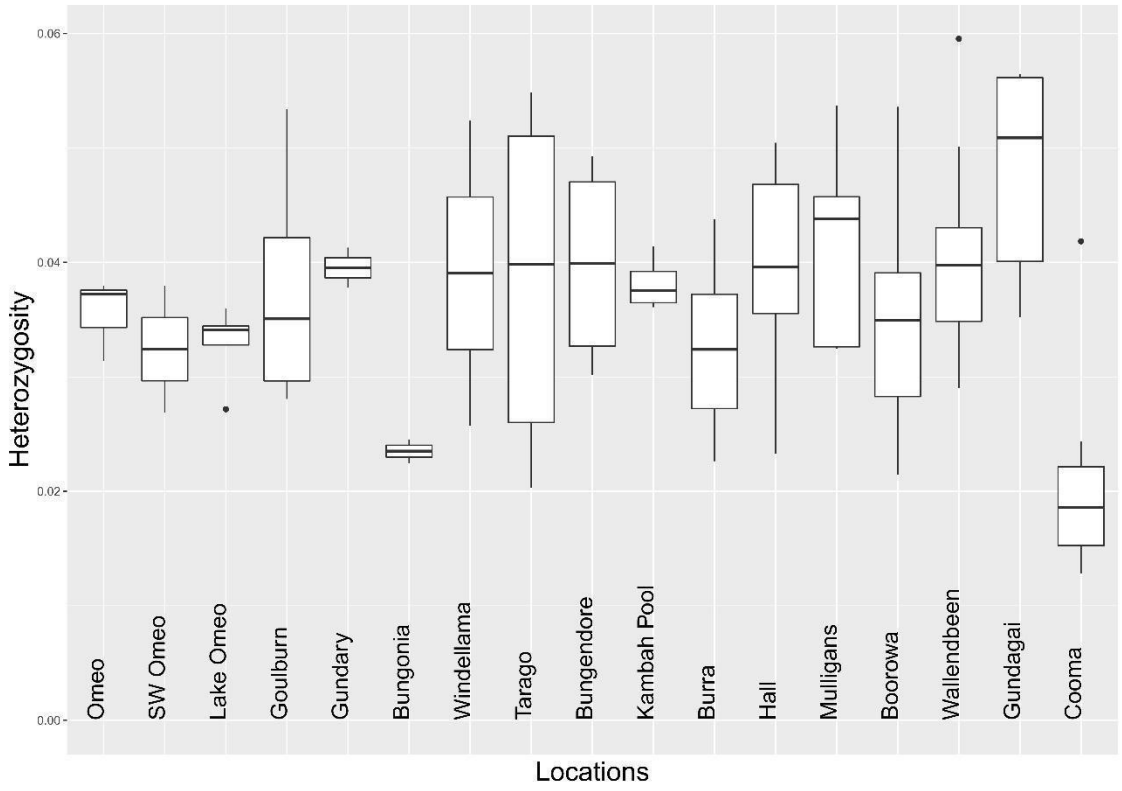
590 Fig. 1 Map of sites surveyed for molecular variation. These sites encompass most of the current known fragmented distribution

591 of *Keyacris scurra*. Singletons from two additional sites were included in the molecular survey: a site close to Gundagai

592 ("Gundagai Cemetery") and a site close to Windellama ("Windellama North").

593

25

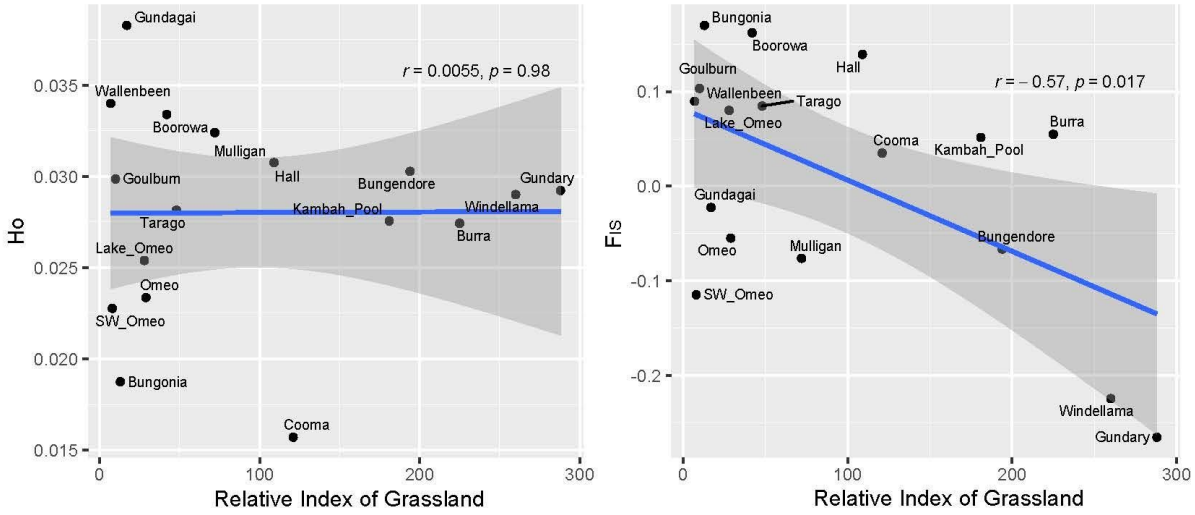


594

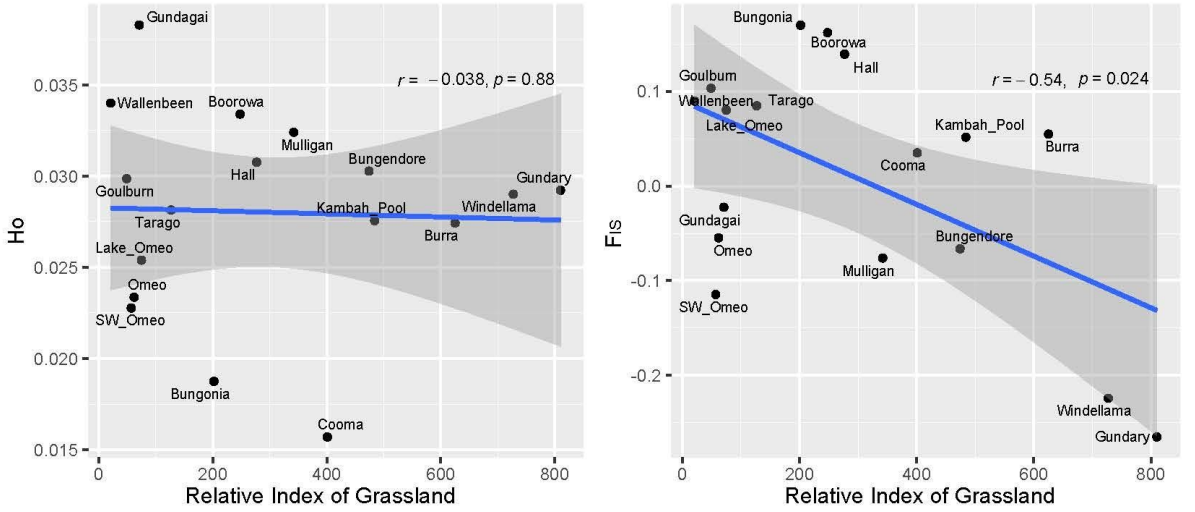
595 Fig. 2. Box plots for Individual heterozygosity by location (for selected SNPs, all individuals prefiltered to MAC = 1 from data
596 where population with an even number of individuals were sampled, but then computed for all individuals from a population).
597 Note that populations are ordered to match the STRUCTURE analysis below. Gundagai Cemetery and Windellama North are not
598 included here because they are represented by singletons.

599

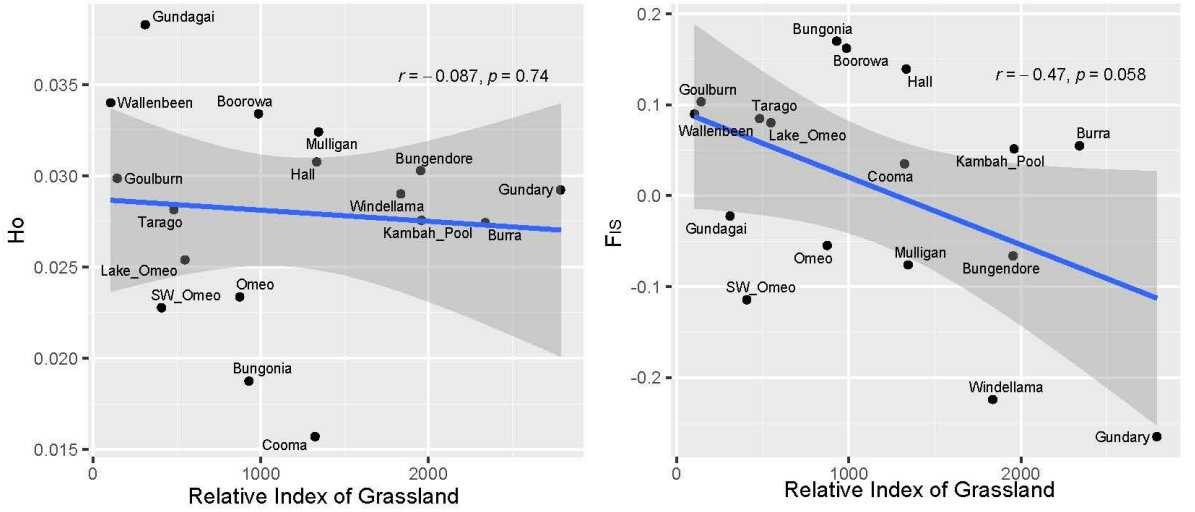
Buffer 25m

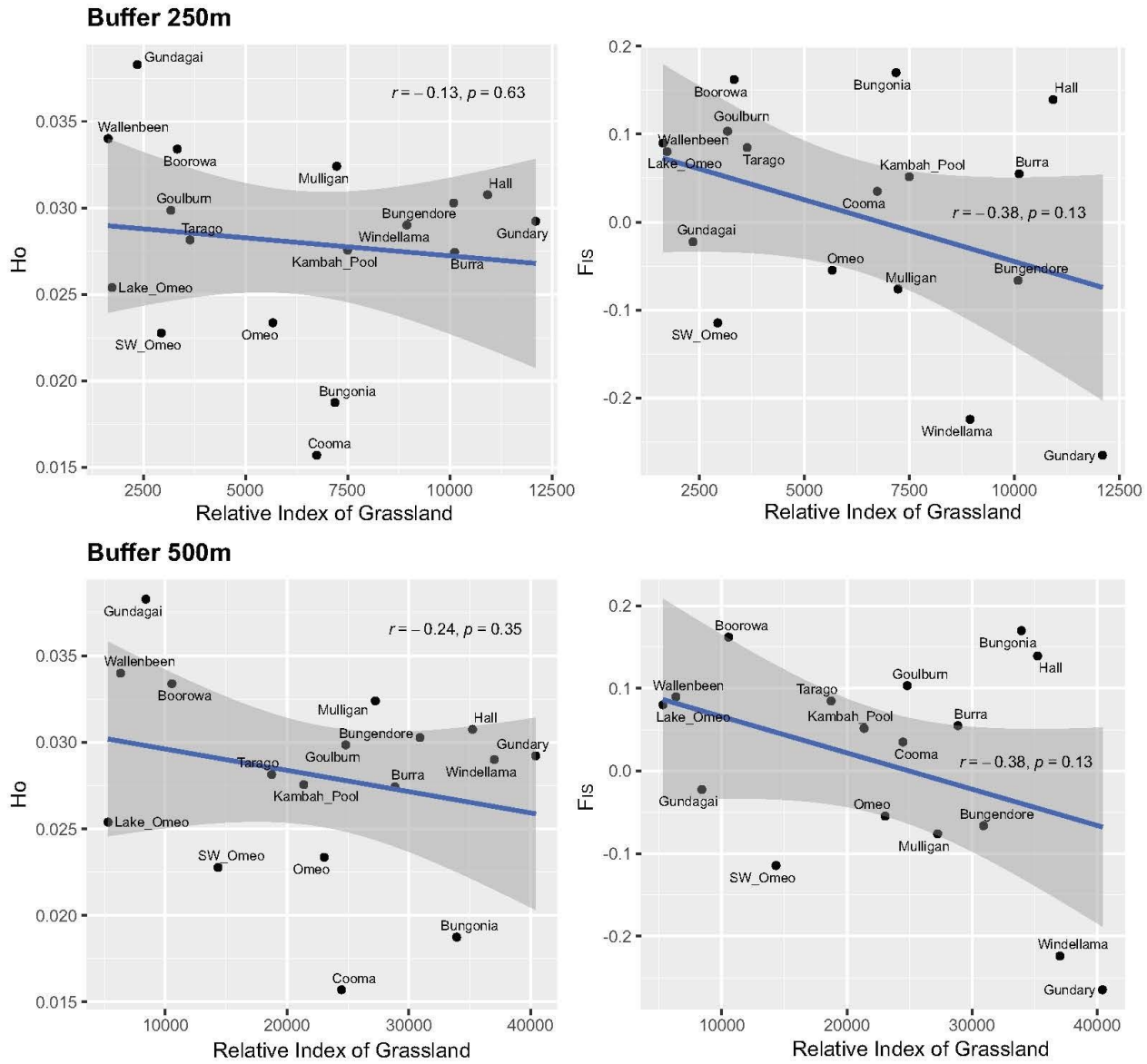


Buffer 50m



Buffer 100m





601

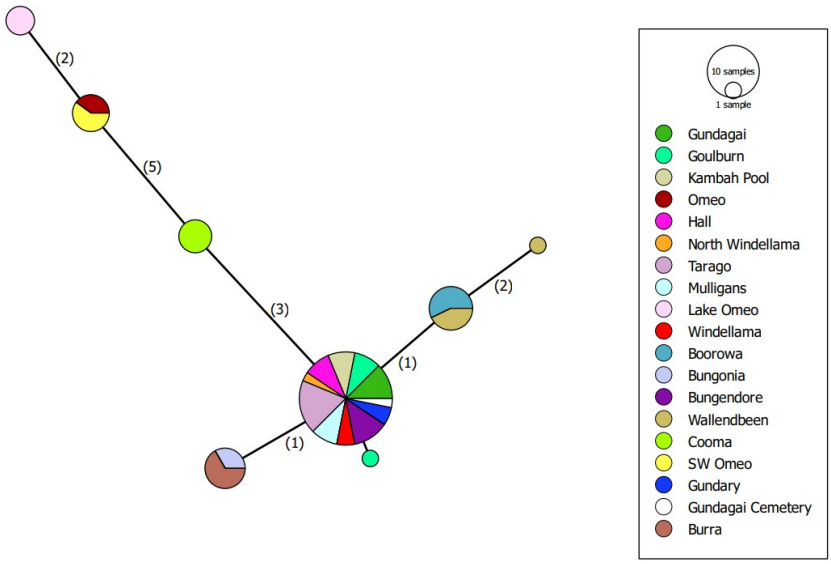
602

603 Fig. 3. Association between a likelihood model of intact native grassland (Sinclair and White, unpublished) at different spatial

604 scales ("buffers") and observed heterozygosity (Ho) (left column) or F_{IS} (right column). Dots reflect individual sites and are

605 presented with correlation coefficients (*r*) and *P* values.

606



607

608 Fig. 4. Variation in the COI gene sequence across *K. scurra* as depicted by a network diagram. The numbers of nucleotide
609 changes are indicated in brackets. The size of the coloured areas reflects the number of haplotypes and branch lengths reflect
610 the number of nucleotide changes.

611

612

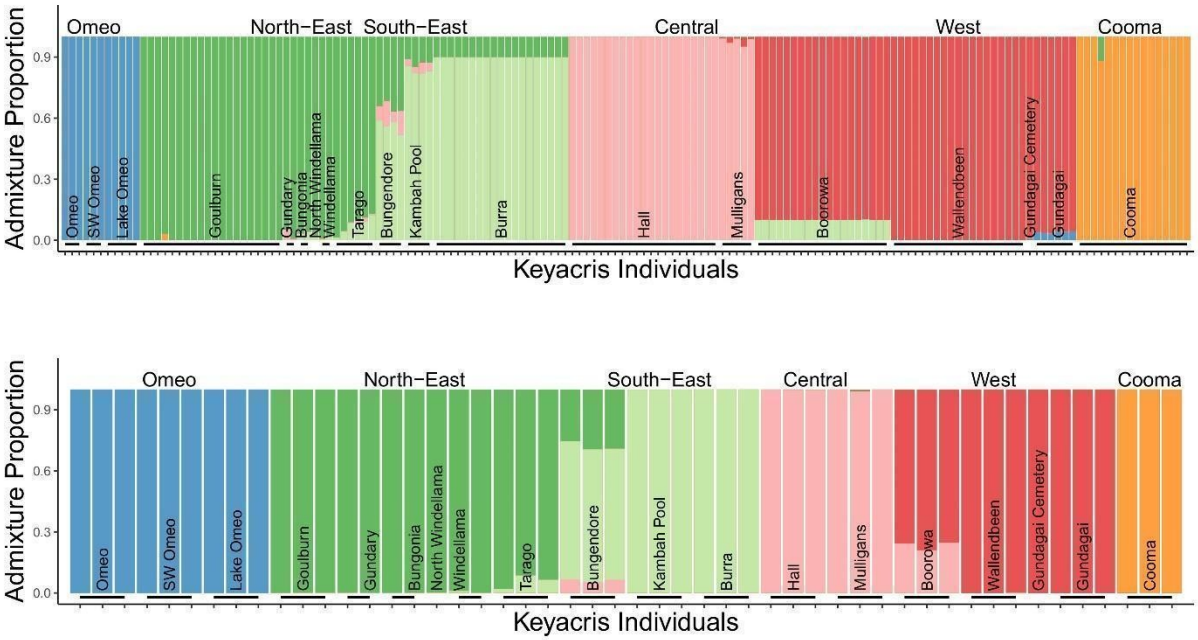
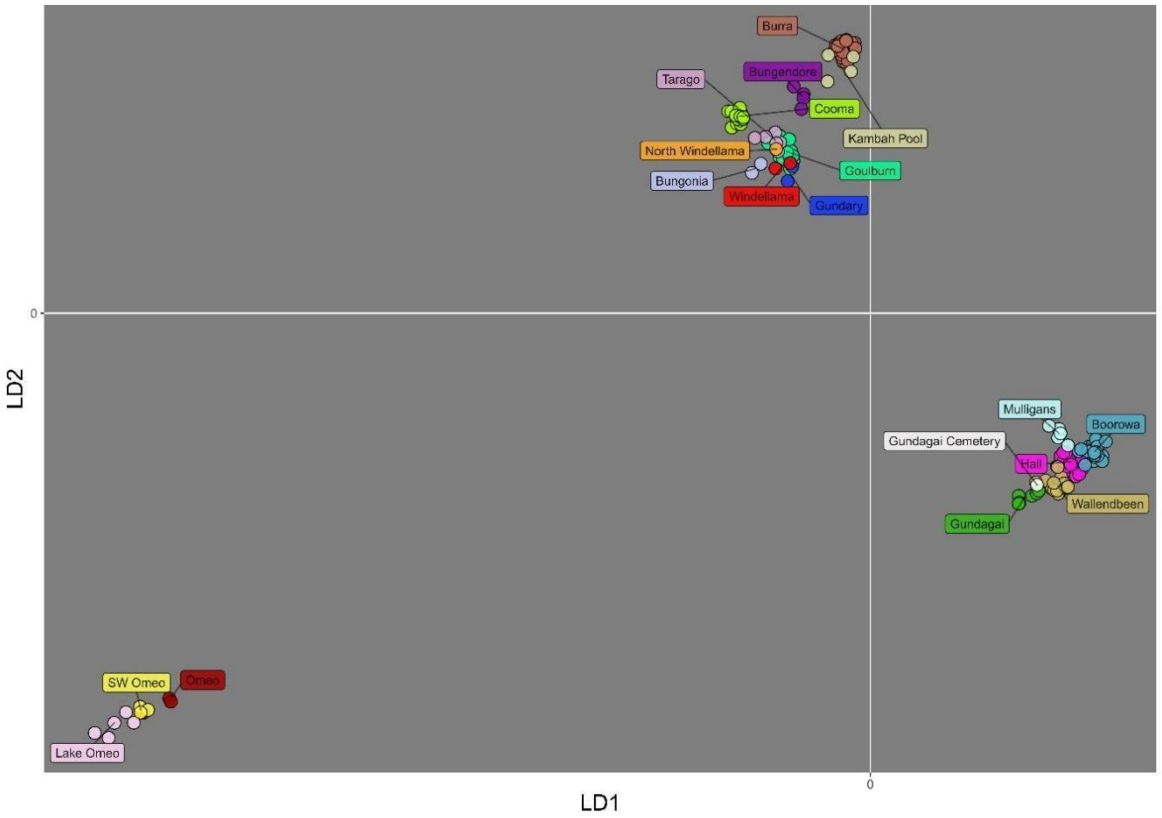
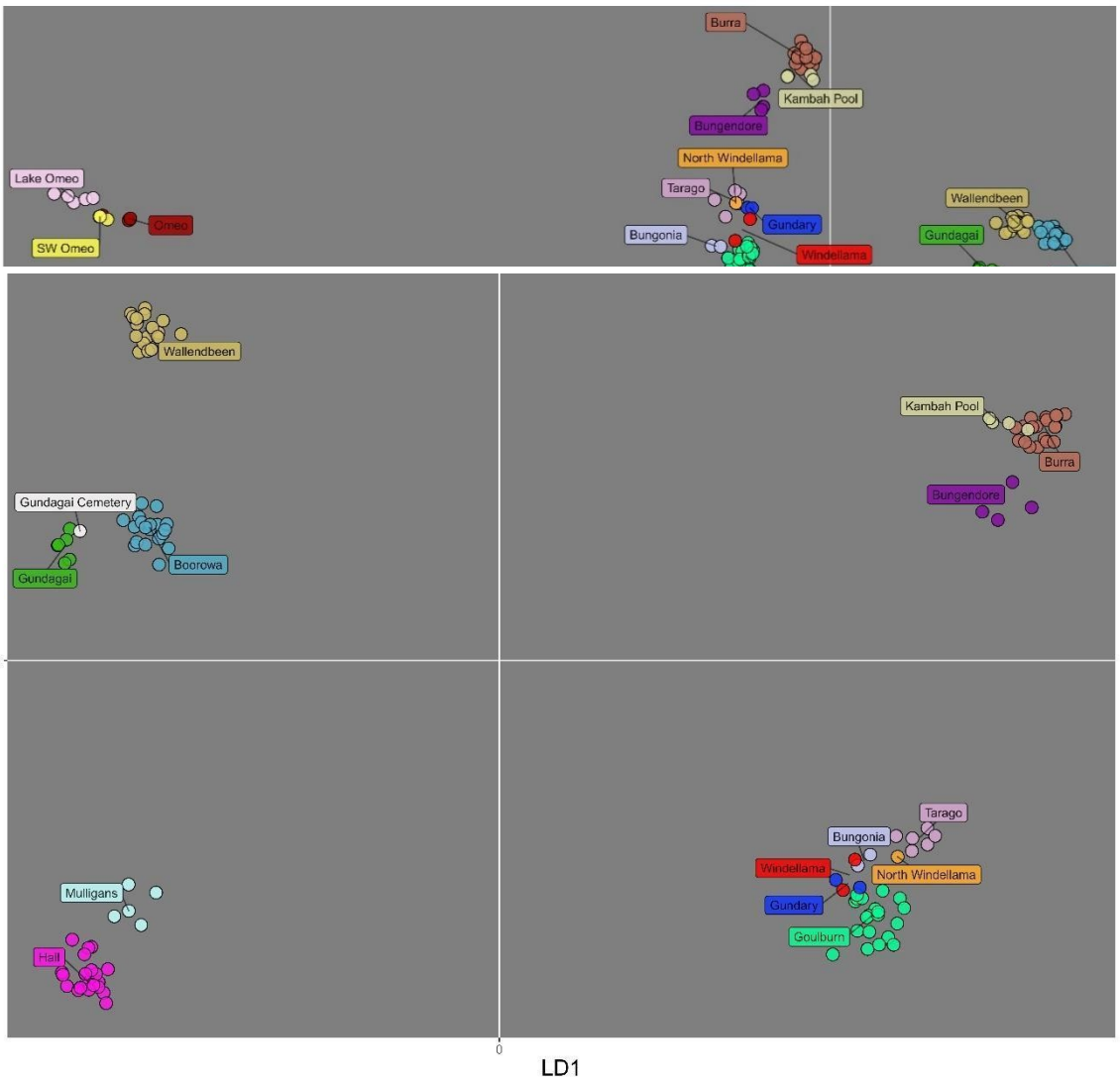


Fig 5. STRUCTURE plots for (a) all individuals and (b) even populations at K=6 (best supported by modified Evanno method).

(a)


Burra
Bungendore
Tarago
Cooma
Kambah Pool
North Windellama
Goulburn
Bungonia
Windellama
Gundary
Mulligans
Boorowa
Gundagai Cemetery
Hall
Wallendbeen
Gundagai
SW Omeo
Omeo
Lake Omeo
LD2
LD1
0
0

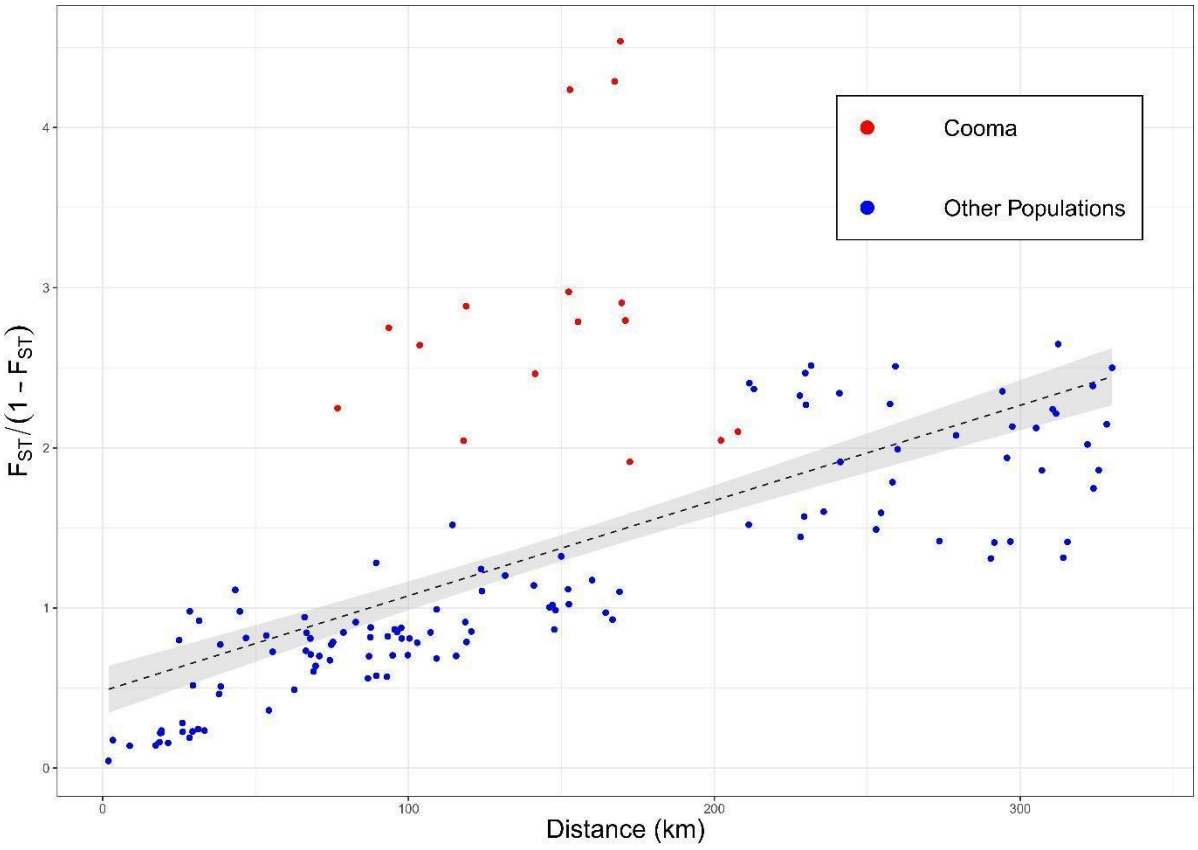
(b)


Burra
Kambah Pool
Bungendore
North Windellama
Tarago
Gundary
Bungonia
Windellama
Lake Omeo
Omeo
SW Omeo
Wallendbeen
Gundagai
Wallendbeen
Kambah Pool
Burra
Bungendore
Gundagai Cemetery
Gundagai
Boorowa
Tarago
Bungonia
Windellama
North Windellama
Gundary
Goulburn
Mulligans
Hall
LD2
LD1
0
0

(c)

Fig 6. DAPC of *Keyacris* scurra individuals with Omeo and Cooma included along the two main linear discriminant (LD) axes (Fig. 6a) and the first and third axes (Fig. 6b) and when these populations are excluded (Fig. 6c). (N = 158, 45 PCs, RMSE 0.043 when all sites included, N=131, 30 PCs, RMSE = 0.029 when Cooma and Omeo excluded).

652



653

654

655 Fig. 7. Correlation of geographical distance with F_{ST}-derived distance between populations.

Table 1. Populations and sample sizes included in genetic analysis of geographic variation. Note that all sites except for two sites where singletons were sampled (Windellama North, Gundagai Cemetery) were included in the analysis of genetic variation within localities.

Population	N	Latitude	Longitude	Environment	Chromosomal race (2n, males)*	Collection rate (hoppers/ person/min)
Boorowa	19	-34.439	148.729	Open woodland	17	0.33
Bungendore	4	-35.342	149.429	Cemetery	15	0.06
Bungonia	2	-34.863	149.942	Cemetery	15	0.10
Burra	19	-35.552	149.229	Open woodland	15	0.50
Cooma	16	-36.234	149.093	Open woodland	15	0.28
Goulburn	20	-34.772	149.731	Open woodland	15	1.10
Gundagai South[1]	6	-35.096	148.091	Cemetery	17	0.32
Gundagai Cemetery	1	-35.052	148.112	Cemetery	17	0.05
Gundary	2	-34.802	149.738	Open woodland	15	0.02
Hall	21	-35.173	149.058	Open woodland	15	0.84
Kambah Pool	4	-35.395	149.012	Grassland	15	0.11
Lake Omeo	5	-36.964	147.657	Grassland verge	15	0.25
Mulligans	4	-35.166	149.155	Open woodland	15	0.17
Omeo	3	-37.107	147.580	Open woodland	15	0.37
SW Omeo	3	-37.115	147.561	Grassy verge	15	0.26
Tarago	6	-35.046	149.654	Grassy verge	15	0.53
Wallenbeen	19	-34.405	148.258	Grassy railway cutting	17	0.33
Windellama	2	-35.014	149.863	Cemetery	15	0.07
Windellama North	1	-34.976	149.900	Grassy verge	15	-

*Based on White (1956)

[1]We refer to this site as Gundagai in the analysis below

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.