Range size variably predicts genetic diversity in *Gehyra* geckos

Ching Ching Lau[a*], Keith Christian[b], Jessica Fenker[a], Rebecca J. Laver[a,c], Kate O'Hara[a], Stephen Zozaya[a], Craig Moritz[a], Emily Roycroft[a*]

[a] *Research School of Biology, Australian National University, Acton, ACT, Australia*

[b] *Research Institute for the Environment and Livelihoods, Charles Darwin University, NT, Australia*

[c] *The University of the Sunshine Coast, Moreton Bay Campus, Petrie, QLD, Australia*

*Corresponding authors: chingching.lau@anu.edu.au and emily.roycroft@gmail.com

Author Contributions: CCL processed and analysed the population genomic data, with advice and assistance from ER and CM. CM and KC secured funding. RJL ran species tree inference. KOH, RJL and JF contributed to laboratory work. SMZ, KC, RJL and CM contributed to fieldwork and sampling. ER and CM supervised the project. CCL and ER wrote the manuscript with input from all authors.

Acknowledgements: The authors acknowledge the Traditional Owners and Custodians of the lands on which we work, and of the lands upon which fieldwork was conducted for this project. We thank Leo Tedeschi, Octavio Jiménez Robles, Kimberley Day, Kade Skelton and Scott Macor for assistance with fieldwork, and Paul Oliver, Conrad Hoskin and the Australian Wildlife Conservancy for contributing samples. We thank Rhiannon Schembri for assistance with DNA extractions, and Lauren Ashman and Renae Pratt for assistance with laboratory work of some samples underlying the phylogenetic inference and Nick Matzke for support to the

phylogenetic analyses. We are grateful to Parks and Wildlife Commission of the Northern Territory for assistance in fieldwork, and Cecilia Myers and Dunkeld Pastoral Corporation for access to tissues. We also thank the curators and collections staff at the Western Australian Museum, Australian Biological Tissue Collection at the South Australian Museum, Museum and Art Gallery of the Northern Territory, Museums Victoria and Queensland Museum who provided access to samples. Fieldwork and tissue collection was conducted under Western Australian DBCA Licence #FO25000067 and #FO25000338-2, Parks and Wildlife Commission of the Northern Territory permits #64816, #66691 and #69132, and Queensland Government Department of Environment and Science permit #WA0036387, and with approval from the Australian National University's Animal Ethics (A2019/15 and A2022/07), and Charles Darwin University's Animal Ethics Committee (A19005).

Funding: This project was supported by funding from the Australian Research Council via Discovery Project DP190102395.

Data Accessibility Statement: Raw sequence data, final variant call files, final alignment files and the 62-lineage *Gehyra* phylogeny have been deposited in Dryad and will be made public on manuscript acceptance. Sample numbers are listed in Supplementary Table 1.

Abstract

Genetic diversity is a fundamental population genetic parameter, and predicts adaptive capacity. Neutral theory predicts a positive correlation between population (or range) size and genetic diversity, but this can be confounded by other demographic processes. To investigate the role of range size, population fluctuation and introgression in determining genetic diversity, we generate and analyse population-level, genomic-scale SNP data from 21 species of Australian *Gehyra* geckos (769 samples) that vary in range size over three orders of magnitude. Using a best-practice approach to estimate SNP-based heterozygosity, we found a significantly positive overall correlation between range size and heterozygosity, although with a shallow slope ($R^2 = 0.30$), consistent with Lewontin's Paradox. At a clade level, we show a stronger relationship between range size and heterozygosity in the *australis* group ($R^2 = 0.74$, $p < 0.01$) than the *nana* group ($R^2 = 0.15$, n.s.). A significantly negative correlation between Tajima's D and range size in both groups, and evidence for introgression in the *nana* group, suggest a role for both population fluctuation and introgression in driving deviations from theoretical expectations. Our results provide insight into the biological and demographic processes that influence genetic diversity, in addition to neutral expectations.

Keywords: heterozygosity, introgression, Lewontin's Paradox, population genomics, RAD-seq

Introduction

Genetic diversity is the principal indicator of genetic health and adaptive capacity in natural populations, and informs our understanding of recent and historical demography of populations and species. Many studies have highlighted the importance of measuring and maintaining genetic diversity in conservation (e.g., (García-Dorado & Caballero, 2021; Kardos et al., 2021), including as a component of the United Nations Convention on Biological Diversity (Hoban et al., 2024). When genetic diversity is high, as expected for large and connected populations, there is more potential variation for selection to act upon, increasing adaptive potential and population viability in the face of environmental change (Reed & Frankham, 2003). Given the recognised global significance of genetic diversity (Exposito-Alonso et al., 2022; García-Dorado & Caballero, 2021), reliable estimates are crucial for conservation and genetic management of species.

The neutral theory of evolution (Kimura, 1983) predicts that in a population of constant size, neutral genetic diversity will correlate positively with effective population size, as formalised by the infinite sites model ($\theta = 4N_e\mu$; Kimura, 1969). We also expect that variance in heterozygosity should be correlated with heterozygosity, and therefore with population size (Kimura, 1983). However, empirical studies have demonstrated that in natural populations, the relationship between genetic diversity and population size is not always clear-cut. While many studies have found a predictive relationship between population size and genetic diversity (e.g. in plants; Hamrick et al., 1992; Newman & Pilson, 1997, and lizards Hague & Routman, 2016), there have also been many exceptions. The most well-known deviation is 'Lewontin's Paradox', the observation that the range of effective population sizes in metazoans is several orders of magnitude larger than the range of observed genetic diversity (Lewontin, 1974). This paradox has been reinforced by empirical studies at broad scales (Buffalo, 2021; Galtier &

Rousselle, 2020; Romiguier et al., 2014), and various processes have been proposed by which to explain it (Charlesworth & Jensen, 2022; Leffler et al., 2012), including fluctuating population size, population structure and selection.

Although Lewontin's Paradox highlights that genetic diversity can level off in large populations, we still expect a positive trend between genetic diversity and effective population size (Charlesworth & Jensen, 2022). Macroecological evidence and theory also support a strong relationship between range size and local abundance (Gaston et al., 1997; Holt et al., 1997). Accordingly, we predict that wide-ranging species should have higher heterozygosity than small range species (Doyle et al., 2015; Gitzendanner & Soltis, 2000). Significant deviations from this expectation could arise through recent range expansion in widespread species (Peter & Slatkin, 2013, 2015; Slatkin & Excoffier, 2012), increased diversity in small-range species due to recent introgression (Alcala et al., 2013), or differences in local population density, as could arise through habitat differences or interspecific competition (Holt et al., 1997).

Here we investigate determinants of genetic diversity in a diverse radiation of northern Australian, scansorial geckos, using population-level genomic data from 21 species or lineages across two clades of the *Gehyra* radiation: the *australis* and *nana* groups, within the phylogenetic framework of a robust coalescent species tree. Focal species vary over 1000-fold in range size and vary in preference for arboreal and rocky habitats. All species are locally abundant, with smaller species (mostly in the *nana* group) occupying smaller to large rocks, larger species occurring on large boulders and cliff faces, and two species in the *australis* group specialised for arboreal habitats (Doughty, Bauer, et al., 2018; Oliver et al., 2019). Within the *nana* group, there is also evidence for size structuring of local communities (Moritz et al.,

2018). With extensive paleoenvironmental change across Australia's north (Bowman et al., 2010; Potter et al., 2018), and many closely related, sympatric species (Doughty, Bourke, et al., 2018; Moritz et al., 2018; Oliver et al., 2019), there is potential for both unstable population size and/or recent hybridisation/introgression, making the *Gehyra* radiation an ideal test case to investigate the impact of demographic and biological processes on genetic diversity. Using a best-practice approach to estimate individual heterozygosity, we explore the roles of range size (as a proxy for population size), population expansion (indicated by Tajima's D), and introgression in determining genetic diversity.

Methods

Sampling, DNA extraction and sequencing

For population genomic analysis, we sampled a total of 775 individuals from 21 *Gehyra* species or lineages (18 described species, two undescribed species, and one additional lineage of a described species; Supplementary Table 1), spanning a variety of range sizes, from 412 to 536,949 km^2 (Supplementary Figure 1). Samples were collected from either museum collections (Supplementary Table 1), or from the field under Australian National University Animal Ethics (A2019/15 and A2022/07) and Charles Darwin University (A19005) Animal Ethics approvals. Of these 21 lineages, nine came from the *australis* group (Oliver et al., 2020) and 12 from the *nana* group (Doughty, Bourke, et al., 2018; Moritz et al., 2018), these being two major clades within the *Gehyra* radiation. For taxonomically undescribed lineages, we follow the nomenclature of Moritz et al. (2018); i.e. *Gehyra nana*1, *G. nana*2, *G. nana*4 and *G. nanamulti* (the latter a divergent form of *G. nana* with mtDNA introgressed from *G. multiporosa*). Two independent lineages of *G. occidentalis* as identified in (Oliver et al., 2017), are indicated by *G. occidentalis* KL (King Leopold Ranges) and *G. occidentalis* OR (Oscar Range).

We extracted total genomic DNA using either a high-salt extraction protocol (Sunnucks & Hales, 1996), or a Qiagen DNeasy Blood & Tissue Kit following the manufacturer's protocol. Samples were then sent to Diversity Arrays Technology (DArTseq; Canberra, Australia) for restriction enzyme-based library preparation (using PstI and HpaII enzymes; Kilian et al., 2012) and sequencing on an Illumina HiSeq2500 platform. In addition to this population-level DArTseq sampling, we also combined 119 new and existing *Gehyra* exon capture samples (Ashman et al., 2018; Moritz et al., 2018; Oliver et al., 2019), representing a total of 62 species or lineages (Supplementary Table 2), to generate a well-supported, taxonomically complete species tree for the *Gehyra* radiation. Exon capture data were generated using the custom, 1900-locus targeted capture approach described in (Moritz et al., 2018).

Bioinformatic processing and variant calling

Unexpected patterns in genetic diversity may be the result of recently recognised biases in SNP-only measures of heterozygosity, that do not account for invariant (Schmidt et al., 2021) or multi-allelic (Sopniewski & Catullo, 2024) sites, which may have affected results in a previous study of genetic diversity in northern Australian lizards (Fenker et al., 2021). To ensure methodological factors did not impact our conclusions, we leveraged two high-quality *Gehyra* reference genomes for bioinformatic processing, that allowed us to estimate individual heterozygosity using both variant and invariant sites. Raw reads generated using the DArTseq platform were processed and mapped to one of two high-quality draft reference genomes (Roycroft et al. unpublished data), *Gehyra lapistola* (*australis* group) and *G. paranana* (*nana* group) using the docker-based pipeline described at https://hub.docker.com/r/trust1/gatk (v0.4.1). Briefly, adaptor sequences and poly-G tails were first trimmed using cutadapt (M. Martin, 2011), and trimmed reads were mapped to each respective reference genome using

BWA-MEM and indexed by samtools v1.10 (Danecek et al., 2021). We used DeepVariant v1.1.0 (Poplin et al., 2018) to call variants with its pre-trained WGS model and generate a gVCF file of SNPs per individual.

Filtering sample-specific gVCF *files to estimate individual heterozygosity*

To retain invariant sites for individual heterozygosity estimation, and given that cohort genotype files retain polymorphic sites only, we first filtered each gVCF file to retain sites with a genotype quality > 20, minimum read depth > 5, and remove all sites with missing data, then used bcftools v1.10.2 "convert" to expand compressed invariant sites records (Danecek et al., 2021). Variable sites rejected by DeepVariant were excluded from individual heterozygosity calculations. Using filtered, single sample gVCF files containing both variant and invariant sites, we calculated individual heterozygosity as the proportion of heterozygous sites out of total valid sites per individual. We then calculated the among-sample variance of individual heterozygosity for each species or lineage using the *var* function in R (R Core Team 2022). We prefer to use mean observed 'individual' heterozygosity, rather than expected heterozygosity (or theta) to avoid confounding effects of non-random mating across sampled individuals. We confirmed that individual heterozygosity was not impacted by sample size be calculating average individual heterozygosity on five down-sampled sets of our best sampled species, sampling *G. koira* (n = 100, n = 75, n = 50, n = 25 and n = 10 (Supplementary Figure 2A), and we also confirmed that variance in heterozygosity was not correlated with sample size (Supplementary Figure 2B).

Cohort genotyping, filtering and cohort population genomic analysis

To obtain population-level SNP datasets for subsequent analysis, we performed cohort genotyping with DeepVariant and GLnexus v1.2.7 (Poplin et al., 2018; Yun et al., 2021) at

three different scales; 1) by species, 2) by group (for the *australis* group), and 3) by geographical regions in *nana* group, i.e. sympatric species. Prior to final cohort genotyping, we confirmed that the species/lineage identification of each individual sample was correct using principal components analysis (PCA), implemented in *SNPrelate* (Zheng et al., 2012) and *gdsfmt* (Zheng et al., 2017) in R, in additional to identification with diagnostic mtDNA markers prior to sequencing. At both scales, cohort genotyped SNPs were filtered using vcftools v0.1.16 with iterative filtering approach (O'Leary et al., 2018) to remove low quality and low coverage genotype calls using the following sequential minimum thresholds; 1) minor allele count ≥ 2, 2) minor allele frequency ≥ 0.01, 3) quality value ≥ 20, 4) minimum mean depth ≥ 10, 5) minimum depth ≥ 3, 6) missing sites < 20% across individuals, 7) only keep samples with < 50% individual missingness, 8) sites with missing data < 5% and minor allele frequency ≥ 0.01, 9) ≥ 10kb distance in the reference genome from the previous SNP. Following all filters, we retained 769 individuals across 21 lineages for final analyses. With these cohort genotyped SNPs, we estimated Tajima's D for each species using the *neutrality.stats* function in the R package 'Popgenome' (Pfeifer et al., 2014). We tested for statistical significance of Tajima's D by comparing the inferred value to both a normal and standard normal distribution of 100 simulated values using 'ms' (Hudson, 2002). Tajima's D values were considered statistically significant if the estimated value fell outside the 95% of simulated values (two-tailed).

Inferring a coalescent species tree for the Australian Gehyra radiation

To generate a well-supported, taxonomically complete species tree for the Australian *Gehyra* radiation for phylogenetically-informed analyses in this study, and to facilitate future comparative studies across *Gehyra*, we combined data from prior exon capture sequencing of 119 samples representing 62 species or lineages (Ashman et al., 2018; Moritz et al., 2018; Oliver et al., 2019). Sequence data for each locus was processed and aligned using the EAPhy

pipeline (Blom, 2015) – details provided in the Supplementary Methods. Final, filtered alignments were ranked using a combination of taxon completeness (> 95%), sequence data completeness (> 95%), locus length (250 bp < length < 500 bp), and tree-length (0.25 < length < 2, inferred in PAUP (Swofford, 2002) using a HKY model). We then selected the top ranked 100 loci for species tree analysis in Starbeast2 v0.14 (Ogilvie et al., 2017). To increase computational tractability, we followed a 'divide and conquer' approach similar to that described in (Ashman et al., 2018) to split the Australian *Gehyra* radiation into a 'skeleton' tree, and four sub-clades for analysis (see Supplementary Table 2). For each clade, we ran two independent Starbeast2 runs of 10^9 iterations, sampling every 10^5 iterations, with 20% discarded as burn-in – see Supplementary Methods. The posterior distribution of sub-clades was combined with the distribution of 'skeleton' tree backbone, to generate a combined global species tree for 62 total species/lineages, provided in the supplementary material to facilitate future comparative studies in the Australian *Gehyra* radiation. For this paper, we then pruned this global tree topology to retain the 21 species/lineages included in our downstream analyses.

Hybridisation and introgression tests

Hybridisation and/or recent introgression are expected to increase heterozygosity and genetic diversity independent of geographic range size. To explore this, we first tested for the presence of individuals with mixed ancestry or potential hybrids using sNMF (Frichot et al., 2014) using the R package LEA (Frichot & François, 2015), at three pairwise contact zones that have shown evidence for potential introgression in previous work (i.e., *Gehyra nana*1 vs. *G. nana*2, *G. nana*1 vs. *G. nanamulti* and *G. nana*1 vs. *G. nana*4). We ran sNMF with alpha = 100, seed = 100, repetitions = 10, and K = 2 indicating the two parental populations of each pairwise comparison. The *G. nana*1 vs. *G. nana*2 contact zone was run with K = 3, as preliminary results suggested additional clusters provided better model fit. We did not test for recent hybridization

in the *australis* group, as no admixture has been detected at contact zones between lineages in this group (Fenker et al., 2021; Oliver et al., 2020; S. Zozaya unpublished data), and our preliminary results did not show any evidence for mixed ancestry or elevated heterozygosity. We then compared the distribution of individual heterozygosity in *G. nana* lineages identified as having mixed ancestry (i.e. ancestry coefficient < 0.8 from any one parental population), to those with pure parental ancestry. We also used an ABBA-BABA approach to test for past introgression among non-sister taxa within each *Gehyra* clade, using the 'Dtrios' function in 'Dsuite' (Malinsky et al., 2021). As we are primarily interested in introgression at timescales that could influence population genomic metrics in extant populations, we only included taxon sets in ABBA-BABA tests that could feasibly have experienced introgression based on current geographic distributions (10 and 56 total comparisons, see Supplementary Figure 1). Tests were based on phylogenetic relationships of *Gehyra* species inferred in our coalescent species tree (see Supplementary Figure 3). Resultant p-values for each D-statistic were adjusted using the *p.adjust* function in R with a false discovery correction (Benjamini & Hochberg, 1995). We then used the python script *dtools.py* provided by Dsuite to plot the F-branch statistic for each comparison, which measures the proportion of gene flow between corresponding branches of the phylogeny (Malinsky et al., 2021; S. H. Martin et al., 2013)

Estimating range size

To estimate range size for each species or lineage, we took the area of a convex hull polygon fitted to geographic records for each species (using the *'minimum bounding geometry'* function in QGIS v. 3.20), clipped to the Australian coastline. Geographic records came from samples genotyped in previous papers (Doughty, Bauer, et al., 2018; Oliver et al., 2020), in this study, and unpublished records. Each polygon was inspected and, where necessary (in two cases; *G. arnhemica* & *G. australis*), manually adjusted. For three species (*G. calcitectus*, *G. ipsa, G.*

pluraporosa) with extremely small and rock-restricted distributions, range sizes were estimated by manually drawing a polygon around the respective rocky ranges using Google Earth. Resultant maps for each species and details of modified or manually estimated ranges are provided in Supplementary Figure 1.

Testing predictors of heterozygosity

To statistically investigate the significance of range size and population expansion/contraction (as measured by Tajima's D) as predictors of mean individual heterozygosity, we used both linear and phylogenetic linear modelling approaches, using the 'lm' function and package *phylom* (Tung Ho & Ané, 2014) respectively. Although heterozygosity is not anticipated to be a phylogenetically inherited trait, we applied both linear and phylogenetic linear models to address any non-independence (see Fig. S7). We assessed model fit using the coefficient of determination (R-squared) and p-values using marginal t-tests. We used a Welch's one-sided t-test to compare heterozygosity between the *australis* and *nana* groups, using the *t.test* function in R with alternative hypothesis set as 'greater'. To additionally explore whether larger range species have higher intra-lineage variance in heterozygosity, we also tested for a correlation between range size and variance in heterozygosity.

Results

Genomic-scale data across the Australian Gehyra radiation

We successfully generated genome-wide, reduced representation data for 21 species and lineages in the *Gehyra* radiation, with sequenced regions spanning all putative autosomes in the draft genomes of *Gehyra lapistola* (*australis* group) and *G. paranana* (*nana* group). Following initial filtering, we retained 769 individuals for final analyses, with an average of

2.14 million callable sites (variant + invariant) and average depth of coverage of 10.2X across all samples, and an average of 40 individuals per species (see Supplementary Table 3 for per-species sequencing summary statistics). Per-species cohort genotyped SNPs (i.e. only variant sites) ranged from 329 – 5631 filtered SNPs (see Supplementary Table 4). Filtered SNP cohorts across multiple species for ABBA-BABA tests retained 3,557 SNPs (*australis* group), 1,934 SNPs for sympatric *nana* species 'group 1' and 1,299 SNPs for sympatric *nana* species 'group 2'. The species-lineage identification of all retained individuals was confirmed via PCA (Supplementary Figure 4). We additionally used existing or new exon capture data from 119 samples, spanning 62 species and lineages in the *Gehyra* radiation (Supplementary Table 2), to facilitate the estimation of a robust coalescent species tree.

Inter- and intra-specific variation in heterozygosity

Our results highlight significant variation in individual heterozygosity across the *Gehyra* phylogeny, both between and within species (Figure 1). Overall, individual heterozygosity was significantly higher in the *nana* group than in the *australis* group ($p = 0.023$, Figure 1B), with *Gehyra nana*1 and *G. nanamulti* showing the highest heterozygosity, and *G. ipsa* and *G. calcitectus* showing the lowest heterozygosity (Figure 1). In several cases, our results show high within-species variation in individual heterozygosity (Figure 1), with the three species displaying the highest variance in individual heterozygosity including the widespread species *G. kimberleyi*, the rock specialist *G. occidentalis* KL, and the small-range G. *pluraporosa*.

Range size predicts heterozygosity, variance in heterozygosity, and Tajima's D

Across the entire dataset, including all 21 lineages, linear models show a significant positive correlation ($R^2 = 0.30$, $p < 0.01$) between range size of each species and average individual heterozygosity (Fig. 2A). When the two clades were modelled separately, the *australis* group

showed a strong and significant positive correlation (R^2 = 0.74, Fig. 2B), while in the *nana* group the relationship was not significant (Fig. 2C). Across all species, range size significantly predicted variance in average individual heterozygosity (Fig. 2D), with a consistent slope when the *australis* (Fig. 2E) and *nana* (Fig. 2F) clades were considered separately. Our models also support a significantly negative correlation between range size and Tajima's D (Fig. 2 G–I), with larger-range species showing more negative Tajima's D values. In six cases, Tajima's D was significantly negative in widespread species (Supplementary Table 4), indicating a significant excess of low frequency alleles, consistent with recent population expansion. This pattern was consistent and significant in the combined 21-species dataset, and in both clades, with the strongest negative slope in the *australis* group (Fig. 2H). In all cases, linear and phylogenetic linear models show comparable trends (Fig. 2, Supplementary Table 5), suggesting our results are robust to potential phylogenetic non-independence, despite overall higher heterozygosity in the *nana* group than in the *australis* group (Fig. 1).

Contrasting patterns of hybridisation and introgression in two Gehyra clades

Using sNMF, we assigned ancestry coefficients to individuals from three contact zones within the *nana* group (i.e., *nana*1 vs. *nana*2, *nana*1 vs. *nana*4, *nana*1 vs. *nanamulti*) to investigate potential elevated heterozygosity as a result of recent hybridization. Most samples were identified as pure parental origin based on sNMF ancestry coefficients (Supplementary Figure 5), with no indication of F1 hybrids. Individuals that were assigned mixed ancestry (i.e. with < 0.8 ancestry coefficient from any one parental population) did not show elevated heterozygosity compared to pure parental individuals (Supplementary Figure 5). Using an ABBA-BABA approach to test for introgression between non-sister lineage pairs that are now in contact, we found no significant evidence for introgression in the *australis* group (Supplementary Table 6). In the *nana* group, three lineage pairs showed significant (adjusted

$p < 0.01$) evidence for past introgression (Fig. 3, Supplementary Table 7). Using the *f*-branch statistic, the signal of introgression appeared to be localised at three topological points in the phylogeny; 1) between *nana*2 and *nana*1, 2) between *nana*2 and the ancestor of *nana*1 + *nanamulti*, and 3) between *granulum* and *kimberleyi* (Fig. 3). Thus, it is possible that some of the elevated heterozygosity in these taxa (Fig. 1) derives from historical introgression.

Discussion

Range size predicts genetic diversity, with clade-specific demographic idiosyncrasies

Given that individual heterozygosity is a measure of genetic diversity, and range size is a proxy of effective population size, a positive relationship between these factors is expected under neutral theory (Kimura 1969, Charlesworth & Jensen 2022). While we see strong and significant support for this relationship in the *australis* group, the relationship between heterozygosity and range size is not significant in the *nana* group. The pattern we observe in the *nana* group is consistent with Lewontin's Paradox, in which species with large populations do not have correspondingly high genetic diversity, i.e. a 'leveling off' of genetic diversity in large populations. Many possible explanations for Lewontin's Paradox have been proposed (Charlesworth & Jensen, 2022; Filatov, 2019), and a key category among these implicates the potential role of demographic factors (e.g. recent population expansion). Several wide-ranging species in our data, in both the *nana* (*nana*1, *nana*2, *nana*4 and *nanamulti*) and *australis* groups (*gemina* and *koira*) have significant signatures of population expansion (Supplementary Table 4) and species with the largest range sizes are more likely to have signatures of population expansion (negative Tajima's D, Fig. 2G). Population expansion can result in low diversity despite large population size, and has thus been used to explain deviations from expected

relationships between range size and genetic diversity in a variety of taxa (e.g. Henn et al., 2012; Westbury et al., 2019).

Potential effects of varying population density

Both the significantly higher overall heterozygosity of *nana* group lineages relative to the *australis* group, and the weaker relationship between heterozygosity and range size in the *nana* group, may relate to differences in population densities. Of particular note is the lack of divergence in mean heterozygosity between some sister-taxa with contrasting range sizes (e.g. *Gehyra girloorloo* vs. *G. kimberleyi*; and *G. occidentalis* KL vs. OR, Fig. 1). While a strong relationship between range size and local abundance is generally expected (Gaston et al., 1997; Holt et al., 1997), interspecies interactions at local scales may disrupt this relationship, and in turn disrupt the relationship between range size and genetic diversity. This could be especially true for lineages of *G. nana*, which can be excluded from larger rock faces in the presence of larger-bodied species (CM and SZ, pers. obs.). Such interspecific interactions, reflected in size-structuring of local communities within the *nana* group (Moritz et al., 2018), could decouple local abundance and range size, and so weaken the expected relationship between range size and genetic diversity. In contrast, rock-dwelling taxa within the *australis* group are rarely sympatric and so have a stronger relationship between heterozygosity and range size.

The relationship between variation in heterozygosity and geography

Our results demonstrate a significant positive correlation between range size and variance in heterozygosity. For widespread species, there is greater potential for differences in local population size and so genetic diversity across the landscape, especially across the often-complex geographies of northern Australia. Greater variance in heterozygosity in more genetically diverse populations is also a theoretical expectation (Kimura, 1969, 1983), as

genetically diverse populations are more likely to be geographically structured. While individual heterozygosity, even of single samples, provides a minimum estimate of genetic diversity (e.g. (Roycroft et al., 2021), this result highlights that estimates of individual heterozygosity from a single or geographically restricted set of samples may be more likely to underestimate genetic diversity in widespread species. While the slope of the relationship between range size and variance in heterozygosity was strong, there are cases where the variance in individual heterozygosity is higher than expected given range size in both range-restricted (e.g. *G. calcitectus*, *G. pluraporosa*) and widespread (*G. kimberleyi*, *G. occidentalis* KL) species. This is potentially driven by fine-scale population structure and differences in local population size. For example, despite being a short-range endemic, *G. calcitectus* occurs on disconnected limestones that may have no opportunity for present-day gene flow. In contrast, lower than expected variance may be explained by high connectivity or recent range expansion. For example, *G. gemina* has the largest range size of all species in our data, but falls below the trendline for variance in heterozygosity given its range size (Fig. 2D). Combined with evidence for a significantly negative Tajima's D (Supplementary Table 4), this suggests that *G. gemina* may have experienced a recent population expansion.

The role of introgression in determining present-day heterozygosity

During periods of secondary contact, gene flow or introgression, genetic diversity can increase unexpectedly, and be subsequently maintained over thousands of generations (Alcala et al., 2013). Our finding of no evidence for hybridization, backcrossing or introgression in the *australis* group, but significant evidence of ancient introgression in the *nana* group, could both contribute to higher overall heterozygosity, and also explain the weaker relationship between range size and heterozygosity in the *nana* group. While introgression (Fig. 3) does not seem to be correlated with a notable increase in the present-day standing heterozygosity in the four

implicated terminal lineages, ancient lineage fusion or introgression may have impacted baseline heterozygosity in the *nana* group as a whole. For example, *G. granulum* and *G. kimberleyi* do not show particularly high heterozygosity given their range sizes, compared to other species in the *nana* group (Fig. 1, Fig. 2). We hypothesise that introgression events we identified using an ABBA-BABA approach in the *nana* group are ancient and that a recurrent history of introgression has impacted genetic diversity across multiple lineages through the history of this clade. This has likely contributed to the overall higher genetic diversity in the *nana* group compared to the *australis* group. Given opportunity for hybridization and introgression at multiple zones of geographic contact (Supplementary Figure 1), a lack of significant evidence for introgression in the history of the *australis* group indicates the potential for strong barriers to geneflow or reproductive isolation in this clade.

Insights into Lewontin's Paradox

Together with other recent comparative analyses (Fonseca et al., 2023; Pelletier et al., 2018), our results support the explanatory power of neutral theory in predicting broad-scale patterns of genetic diversity, but also highlight the potential role for demographic processes to contribute to deviations from expected patterns. Various studies that address Lewontin's Paradox – the lack of collinearity of heterozygosity and population size – combine data across ecologically disparate taxa where current and past demographies may disrupt the expected relationship (Buffalo, 2021; Corbett-Detig et al., 2015; Galtier & Rousselle, 2020; Leffler et al., 2012; Romiguier et al., 2014). Here, for an ecologically similar set of taxa distributed across the same biogeographic region, we do find an overall correlation between genetic diversity and range size. However, contrasting patterns between two independent but closely related clades reveal how effects of past introgression, interspecific interactions and variable responses to

449 past environmental change can together disrupt the relationship between population or range
450 size and genetic diversity in natural populations.
451
452 **Reference List**
453

454 Alcala, N., Streit, D., Goudet, J., & Vuilleumier, S. (2013). Peak and Persistent Excess of
455 Genetic Diversity Following an Abrupt Migration Increase. *Genetics*, *193*(3), 953–971.
456 https://doi.org/10.1534/genetics.112.147785
457 Ashman, L. G., Bragg, J. G., Doughty, P., Hutchinson, M. N., Bank, S., Matzke, N. J., Oliver,
458 P., & Moritz, C. (2018). Diversification across biomes in a continental lizard radiation.
459 *Evolution*, *72*(8), 1553–1569. https://doi.org/10.1111/evo.13541
460 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and
461 powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*,
462 *57*, 289–300.
463 Blom, M. P. K. (2015). EAPhy: A Flexible Tool for High-throughput Quality Filtering of
464 Exon-alignments and Data Processing for Phylogenetic Methods. *PLoS Currents*.
465 https://doi.org/10.1371/currents.tol.75134257bd389c04bc1d26d42aa9089f
466 Bowman, D. M. J. S., Brown, G. K., Braby, M. F., Brown, J. R., Cook, L. G., Crisp, M. D.,
467 Ford, F., Haberle, S., Hughes, J., Isagi, Y., Joseph, L., McBride, J., Nelson, G., &
468 Ladiges, P. Y. (2010). Biogeography of the Australian monsoon tropics. *Journal of*
469 *Biogeography*, *37*(2), 201–216. https://doi.org/10.1111/j.1365-2699.2009.02210.x
470 Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population
471 size suggests natural selection cannot explain Lewontin's Paradox. *ELife*, *10*.
472 https://doi.org/10.7554/eLife.67509
473 Charlesworth, B., & Jensen, J. D. (2022). How Can We Resolve Lewontin's Paradox?
474 *Genome Biology and Evolution*, *14*(7). https://doi.org/10.1093/gbe/evac096
475 Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural Selection Constrains
476 Neutral Diversity across A Wide Range of Species. *PLOS Biology*, *13*(4), e1002112.
477 https://doi.org/10.1371/journal.pbio.1002112
478 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham,
479 A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of
480 SAMtools and BCFtools. *GigaScience*, *10*(2).
481 https://doi.org/10.1093/gigascience/giab008
482 Doughty, P., Bauer, A. M., Pepper, M., & Keogh, J. S. (2018). Spots before the eyes: revision
483 of the saxicoline geckos of the Gehyra punctata (Squamata: Gekkonidae) species
484 complex in the Pilbara region of Western Australia. *Records of the Western Australian*
485 *Museum*, *33*(1), 1. https://doi.org/10.18195/issn.0312-3162.33(1).2018.001-050
486 Doughty, P., Bourke, G., Tedeschi, L. G., Pratt, R. C., Oliver, P. M., Palmer, R. A., &
487 Moritz, C. (2018). Species delimitation in the Gehyra nana (Squamata: Gekkonidae)
488 complex: Cryptic and divergent morphological evolution in the Australian Monsoonal
489 Tropics, with the description of four new species. *Zootaxa*, *4403*(2), 201–244.
490 https://doi.org/10.11646/zootaxa.4403.2.1
491 Doyle, J. M., Hacking, C. C., Willoughby, J. R., Sundaram, M., & DeWoody, J. A. (2015).
492 Mammalian genetic diversity as a function of habitat, body size, trophic class, and

493 conservation status. *Journal of Mammalogy*, *96*(3), 564–572.
494 https://doi.org/10.1093/jmammal/gyv061
495 Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C.,
496 Lang, P. L. M., Leventhal, L., Nogues-bravo, D., Pagowski, V., Ruffley, M., Spence, J.
497 P., Arana, S. E. T., Weiß, C. L., & Zess, E. (2022). Genetic diversity loss in the
498 Anthropocene. *Science*, *377*(6613), 1431–1435.
499 Fenker, J., Tedeschi, L. G., Melville, J., & Moritz, C. (2021). Predictors of phylogeographic
500 structure among codistributed taxa across the complex Australian monsoonal tropics.
501 *Molecular Ecology*, *30*(17), 4276–4291. https://doi.org/10.1111/mec.16057
502 Filatov, D. A. (2019). Extreme Lewontin's Paradox in Ubiquitous Marine Phytoplankton
503 Species. *Molecular Biology and Evolution*, *36*(1), 4–14.
504 https://doi.org/10.1093/molbev/msy195
505 Fonseca, E. M., Pelletier, T. A., Decker, S. K., Parsons, D. J., & Carstens, B. C. (2023).
506 Pleistocene glaciations caused the latitudinal gradient of within-species genetic
507 diversity. *Evolution Letters*, *7*(5), 331–338. https://doi.org/10.1093/evlett/qrad030
508 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological
509 association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.
510 https://doi.org/10.1111/2041-210X.12382
511 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and
512 Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–983.
513 https://doi.org/10.1534/genetics.113.160572
514 Galtier, N., & Rousselle, M. (2020). How Much Does *Ne* Vary Among Species? *Genetics*,
515 *216*(2), 559–572. https://doi.org/10.1534/genetics.120.303622
516 García-Dorado, A., & Caballero, A. (2021). Neutral genetic diversity as a useful tool for
517 conservation biology. In *Conservation Genetics* (Vol. 22, Issue 4, pp. 541–545).
518 Springer Science and Business Media B.V. https://doi.org/10.1007/s10592-021-01384-9
519 Gaston, K. J., Blackburn, T. M., & Lawton, J. H. (1997). Interspecific Abundance-Range
520 Size Relationships: An Appraisal of Mechanisms. *The Journal of Animal Ecology*,
521 *66*(4), 579. https://doi.org/10.2307/5951
522 Gitzendanner, M. A., & Soltis, P. S. (2000). Patterns of genetic variation in rare and
523 widespread plant congeners. *American Journal of Botany*, *87*(6), 783–792.
524 https://doi.org/10.2307/2656886
525 Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A
526 test with sympatric lizard species. *Heredity*, *116*(1), 92–98.
527 https://doi.org/10.1038/hdy.2015.76
528 Hamrick, J. L., Godt, M. J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels
529 of genetic diversity in woody plant species. *New Forests*, *6*(1–4), 95–124.
530 https://doi.org/10.1007/BF00120641
531 Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion.
532 *Proceedings of the National Academy of Sciences*, *109*(44), 17758–17764.
533 https://doi.org/10.1073/pnas.1212380109
534 Hoban, S., da Silva, J. M., Hughes, A., Hunter, M. E., Kalamujić Stroil, B., Laikre, L.,
535 Mastretta-Yanes, A., Millette, K., Paz-Vinas, I., Bustos, L. R., Shaw, R. E., Vernesi, C.,
536 Funk, C., Grueber, C., Kershaw, F., MacDonald, A., Meek, M., Mittan, C., O'Brien, D.,
537 … Segelbacher, G. (2024). Too simple, too complex, or just right? Advantages,
538 challenges, and guidance for indicators of genetic diversity. *BioScience*, *74*(4), 269–280.
539 https://doi.org/10.1093/biosci/biae006
540 Holt, R. D., Lawton, J. H., Gaston, K. J., & Blackburn, T. M. (1997). On the Relationship
541 between Range Size and Local Abundance: Back to Basics. *Oikos*, *78*(1), 183.
542 https://doi.org/10.2307/3545815

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338. https://doi.org/10.1093/bioinformatics/18.2.337

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Chris Funk, W. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(48). https://doi.org/10.1073/pnas.2104642118

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* (pp. 67–89). https://doi.org/10.1007/978-1-61779-870-2_5

Kimura, M. (1969). The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics*, *61*(4), 893–903. https://doi.org/10.1093/genetics/61.4.893

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLoS Biology*, *10*(9), e1001388. https://doi.org/10.1371/journal.pbio.1001388

Lewontin, R. (1974). *The Genetic Basis of Evolutionary Change* . Columbia University Press.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast *D* -statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Moritz, C. C., Pratt, R. C., Bank, S., Bourke, G., Bragg, J. G., Doughty, P., Keogh, J. S., Laver, R. J., Potter, S., Teasdale, L. C., Tedeschi, L. G., & Oliver, P. M. (2018). Cryptic lineage diversity, body size divergence, and sympatry in a species complex of Australian lizards (Gehyra). *Evolution*, *72*(1), 54–66. https://doi.org/10.1111/evo.13380

Newman, D., & Pilson, D. (1997). INCREASED PROBABILITY OF EXTINCTION DUE TO DECREASED GENETIC EFFECTIVE POPULATION SIZE: EXPERIMENTAL POPULATIONS OF *CLARKIA PULCHELLA*. *Evolution*, *51*(2), 354–362. https://doi.org/10.1111/j.1558-5646.1997.tb02422.x

Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, *34*(8), 2101–2114. https://doi.org/10.1093/molbev/msx126

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*(16), 3193–3206. https://doi.org/10.1111/mec.14792

Oliver, P. M., Ashman, L. G., Bank, S., Laver, R. J., Pratt, R. C., Tedeschi, L. G., & Moritz, C. C. (2019). On and off the rocks: Persistence and ecological diversification in a tropical Australian lizard radiation. *BMC Evolutionary Biology*, *19*(1), 1–15. https://doi.org/10.1186/s12862-019-1408-1

Oliver, P. M., Laver, R. J., Martins, F. D. M., Renae, C., Hunjan, S., & Moritz, C. C. (2017). A novel hotspot of vertebrate endemism and an evolutionary refugium in tropical Australia. *Diversity and Distributions*, *23*, 53–66. https://doi.org/10.1111/ddi.12506

Oliver, P. M., Prasetya, A. M., Tedeschi, L. G., Fenker, J., Ellis, R. J., Doughty, P., & Moritz, C. (2020). Crypsis and convergence: Integrative taxonomic revision of the Gehyra australis group (Squamata: Gekkonidae) from northern Australia. *PeerJ*, *2020*(1), 1–56. https://doi.org/10.7717/peerj.7971

Pelletier, T. A., Carstens, B. C., Tank, D. C., Sullivan, J., & Espíndola, A. (2018). Predicting plant conservation priorities on a global scale. *Proceedings of the National Academy of Sciences*, *115*(51), 13027–13032. https://doi.org/10.1073/pnas.1804098115

Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, *67*(11), 3274–3289. https://doi.org/10.1111/evo.12202

Peter, B. M., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution*, *69*(3), 721–734. https://doi.org/10.1111/evo.12609

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983–987. https://doi.org/10.1038/nbt.4235

Potter, S., Xue, A. T., Bragg, J. G., Rosauer, D. F., Roycroft, E. J., & Moritz, C. (2018). Pleistocene climatic changes drive diversification across a tropical savanna. *Molecular Ecology*, *27*(2), 520–532. https://doi.org/10.1111/mec.14441

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology*, *17*(1), 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. A.-T., Weinert, L. A., Belkhir, K., Bierne, N., … Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, *515*(7526), 261–263. https://doi.org/10.1038/nature13685

Roycroft, E., MacDonald, A. J., Moritz, C., Moussalli, A., Miguez, R. P., & Rowe, K. C. (2021). Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(27), e2021390118. https://doi.org/10.1073/pnas.2021390118

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

Slatkin, M., & Excoffier, L. (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics*, *191*(1), 171–181. https://doi.org/10.1534/genetics.112.139022

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

Sunnucks, P., & Hales, D. F. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae).

Molecular Biology and Evolution, *13*(3), 510–524. https://doi.org/10.1093/oxfordjournals.molbev.a025612

Swofford, D. (2002). *PAUP*: Phylogenetic analysis using parsimony (* and other methods), ver.4.0a163* (4.0a163). Sinauer Associates.

Westbury, M. V., Petersen, B., Garde, E., Heide-Jørgensen, M. P., & Lorenzen, E. D. (2019). Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size. *IScience*, *15*, 592–599. https://doi.org/10.1016/j.isci.2019.03.023

Yun, T., Li, H., Chang, P.-C., Lin, M. F., Carroll, A., & McLean, C. Y. (2021). Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics*, *36*(24), 5582–5589. https://doi.org/10.1093/bioinformatics/btaa1081

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., & Levine, D. (2017). SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*, *33*(15), 2251–2257. https://doi.org/10.1093/bioinformatics/btx145

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Main Text Figures

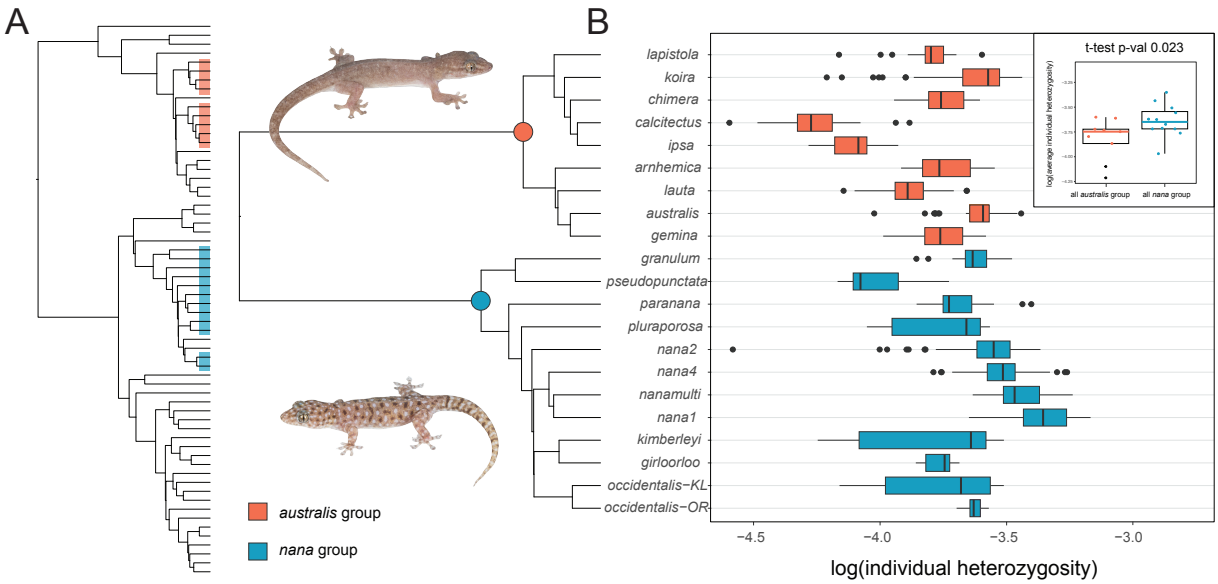
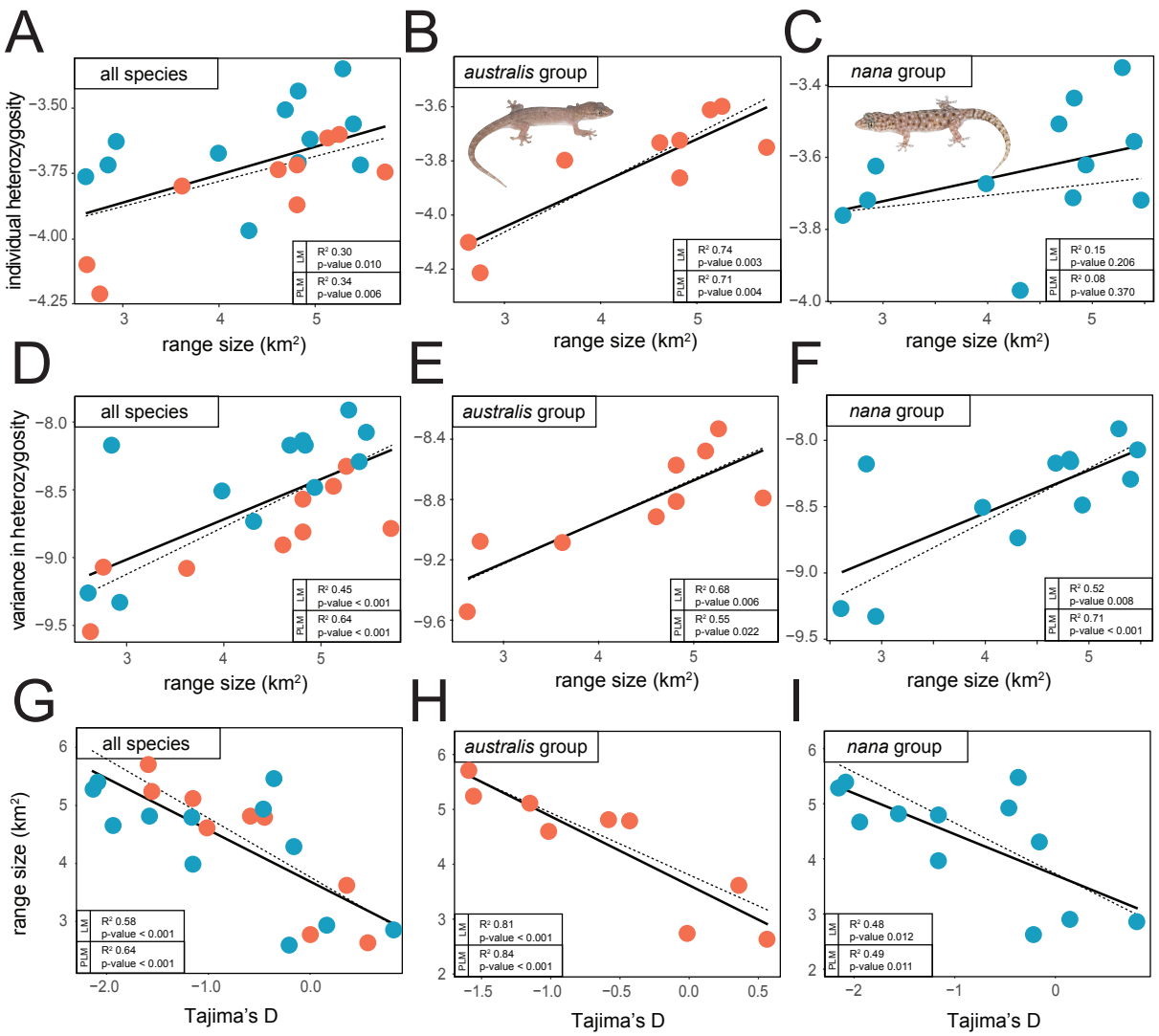


Figure 1 Genetic diversity in two *Gehyra* focal groups. A) Coalescent species tree for the entire Australian *Gehyra* radiation inferred in Starbeast2, and B) the distribution of individual heterozygosity in the *australis* and *nana* groups. Inset shows the significant difference in average individual heterozygosity in all *australis* group lineages, compared to all *nana* group lineages. Photos of *G. lapistola* (top) and *G. paranana* (bottom) by Scott Macor.



684

685 **Figure 2** Predicting genetic diversity in *Gehyra* geckos. Correlations between range size and

686 individual heterozygosity (A – C), range size and variance in individual heterozygosity (D –

687 F), and Tajima's D and range size (G – I). Solid trend line indicates standard linear model,

688 dashed trend line indicates phylogenetic linear model. All axes except Tajima's D values are

689 log-scaled.

690

691

692

693

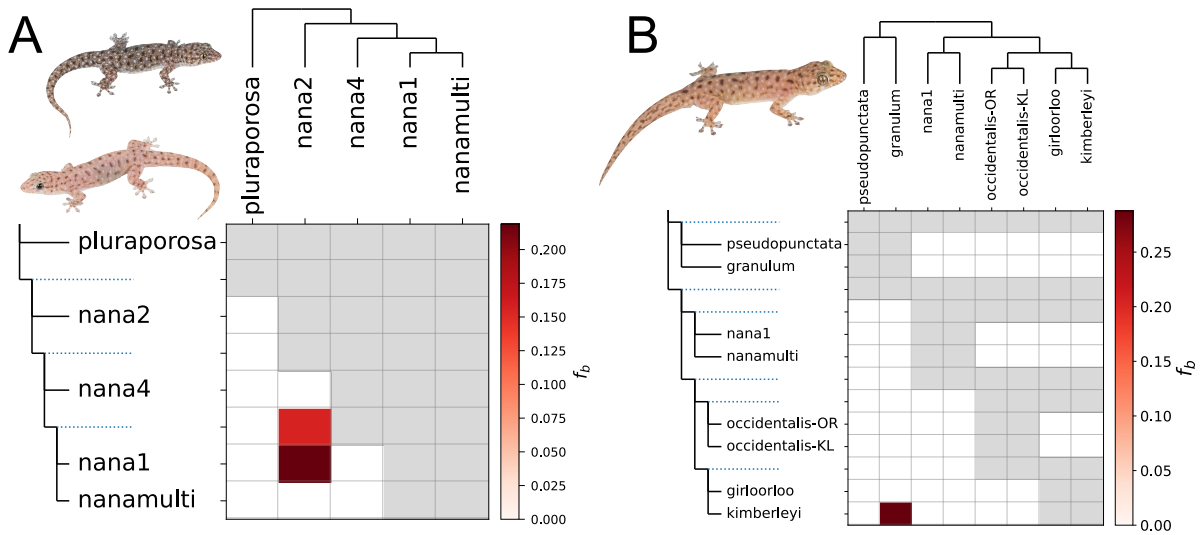


Figure 3 Evidence for historical introgression in the *nana* clade. Pairwise f-branch statistic (f_b) as a measure of excess allele sharing between species or ancestral branches, compared to their respective sister branches, in A) four species in the *nana* complex, with *pluraporosa* as an outgroup, and B) eight species in broader *nana* clade. In contrast, no pairwise comparisons in the *australis* clade showed significant evidence for introgression, see Supplementary X. Photos of *Gehyra nana*2, *G. nana*1 and *G. granulum* by Scott Macor.

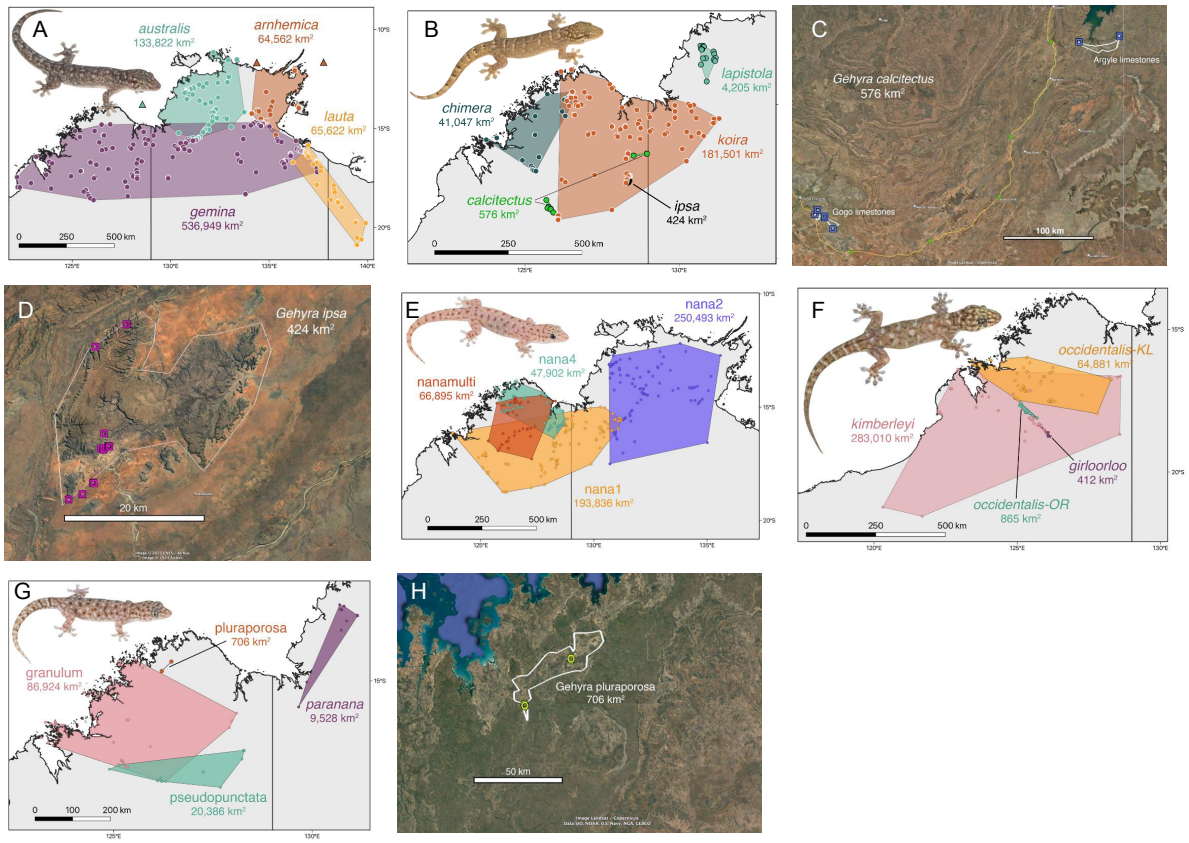
Supplementary Methods

Inferring a coalescent species tree for the Australian Gehyra radiation – additional detail

Where possible, the final alignment included two representative individuals (one haplotype per individual) per species/lineage. In cases where only one sample was available for any lineage and the sample had higher than 20x average coverage, both the h0 and h1 haplotypes from the EAPhy pipeline were included in final alignments to ensure two representatives for coalescent analyses. Each locus alignment was also manually inspected in Geneious 6.1.8 (https://www.geneious.com) and edited to remove alignment errors. Loci with evidence of paralogy, low-quality individuals (high rates of Ns due to low coverage), chimeras and contaminated individuals (haplotypes of incorrect species, often associated with chimeras) were removed.

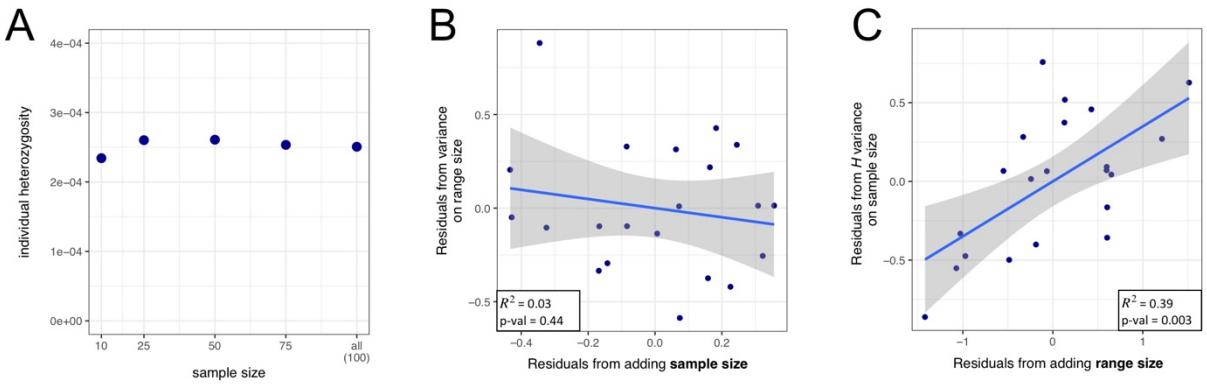
Sub-clades for the 'divide-and-conquer' approach were defined as monophyletic groups containing no more than 20 lineages, with 100% bootstrap support in a preliminary RAxML analysis (i.e., *australis* group, n = 20; *xenopus* group, n = 4, *nana* group, n = 14; *variegata* group, n = 20). The skeleton tree was then made up of two representative taxa to define the base of each sub-clade, and any other poorly supported, spurious or evolutionarily unique lineages that did not fall within the above sub-clades (see Supplementary Table 2). For the skeleton tree and each sub-clade, we ran two independent Starbeast2 runs of 10^9 iterations, sampling every 10^5 iterations, with 20% discarded as burn-in. Each run used a linked strict clock model across all exons with 1/X prior. Each locus had its own partition with substitution model set to HKY+G, constant population model and Birth-Death speciation model. Tree-root height was set to 1.0, to facilitate downstream integration of sub-clade trees. In one case (the *australis* group, the largest sub-clade), we ran four independent runs to ensure adequate ESS values (> 200) for all parameters (assessed using Tracer V1.7).

Supplementary Figures and Tables

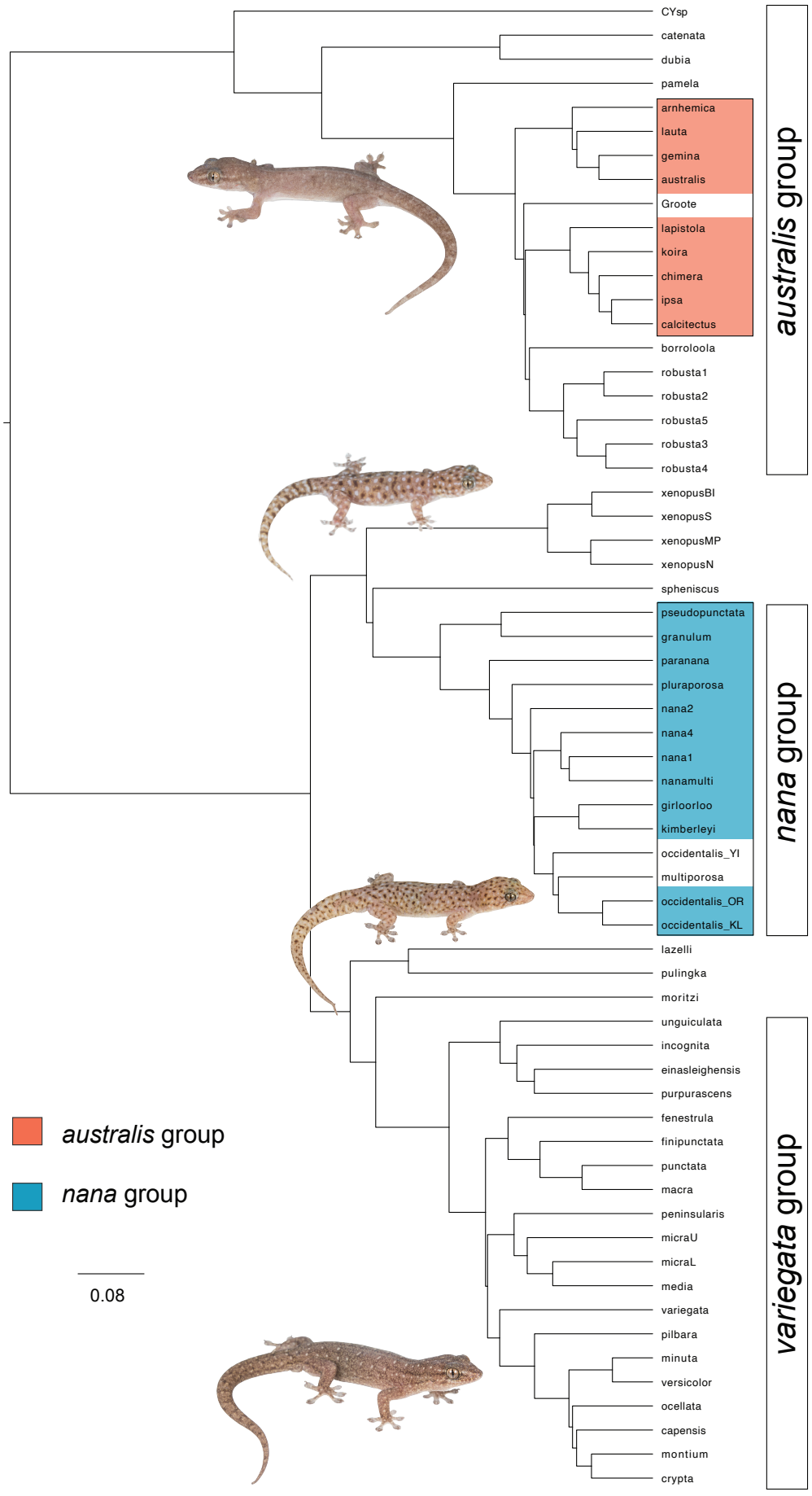


Supplementary Figure 1 Geographic records and estimated range size of *Gehyra* species included in this study. A) *G. arnhemica* (pictured), *G. australis*, *G. gemina*, and *G. lauta*. Triangular points represent points manually added to extend the polygons of *G. arnhemica* and *G. australis* across areas where they have been observed to occur despite having no samples or specimens from those areas. B) *G. calictectus*, *G. chimera*, *G. ipsa* (pictured), *G. koira*, and *G. lapistola*. Polygons to estimate range size were drawn manually for *G. calcitectus* and *G. ipsa*, both of which have extremely narrow distributions on specific rock types. C) *G. calcitectus* distribution. Polygons were drawn around the approximate boundaries of the two limestone formations, with the range size estimate being the sum of the two polygons. D) *G. ipsa* distribution. The polygon was drawn around the approximate boundary of the Purnululu massif, which is accessible by road only on its western half. E) *G. nana* complex, which is comprised of four deeply divergent and morphologically similar candidate species: *nana1* (pictured), *nana2*, *nana4*, and *nanamulti*. F) *G. girloorloo*, *G. kimberleyi*, and two deeply divergent candidate species within the *G. occidentalis* complex*: occidentalis-KL* (pictured) and

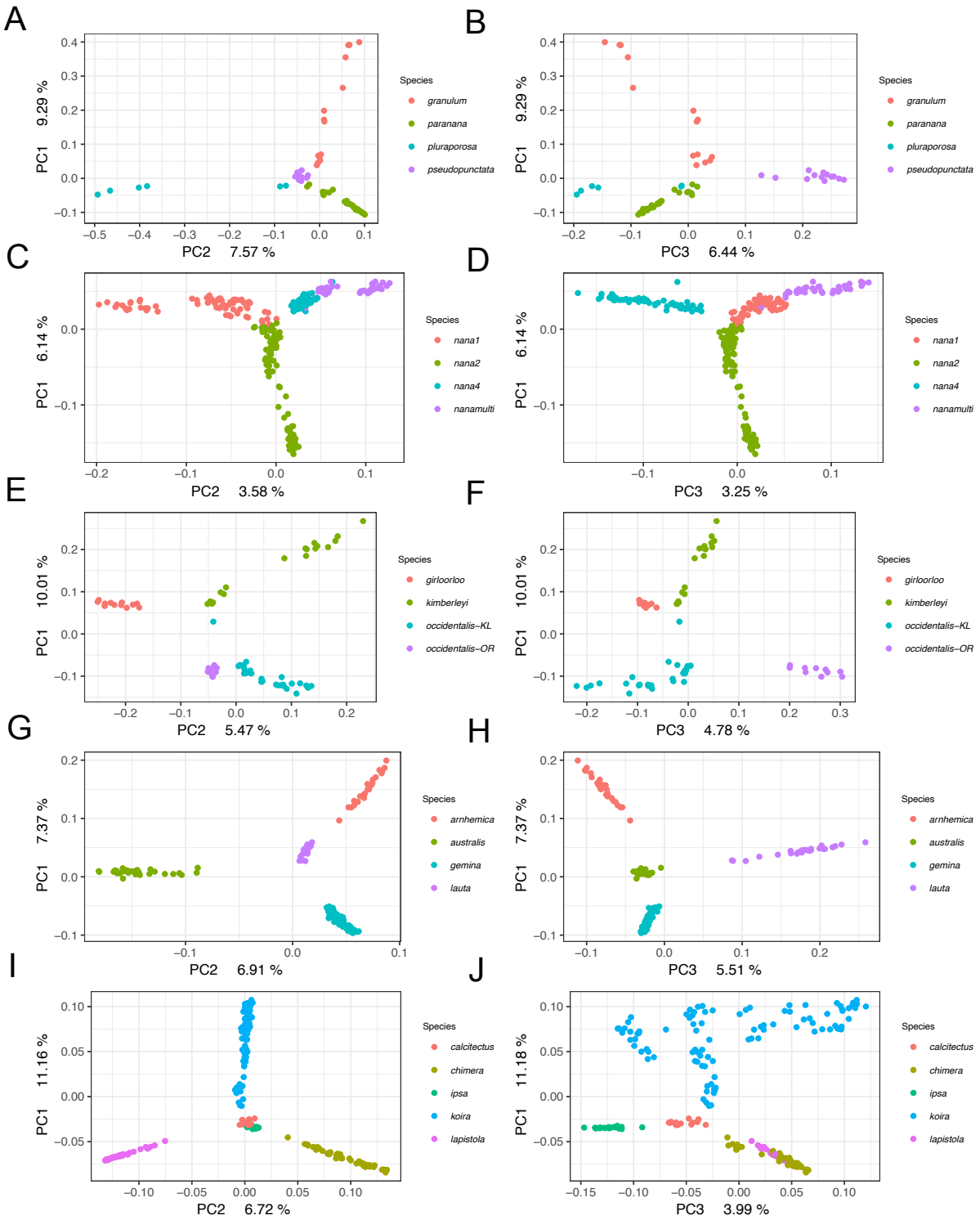
occidentalis-OR. G) *G. granulum*, *G. paranana* (pictured), *G. pluraporosa*, and *G. pseudopunctata*. The polygon for *G. pluraporosa* was drawn manually. H) *G. pluraporosa* distribution. The species is known from only two localities within the King Edward River sandstone on Theda Station. The polygon here has been drawn around the approximate margins of that sandstone formation, excluding some nearby and seemingly well-connected formations where intense searches have not detected this otherwise easy to find species (S. Zozaya pers. obs.). All photos by Scott Macor.



Supplementary Figure 2 Sample size does not impact heterozygosity or variance in heterozygosity. A) Sequentially down-sampled subsets of *Gehyra koira* between n = 100 and n = 10, showing no significant impact of sample size of average individual heterozygosity. B) Given a model lm(range size ~ *H* variance), how well does adding sample size explain the unexplained residuals, compared to C) given lm(N ~ *H* variance), how well does adding range size explain the unexplained residuals. N = sample size, *H* = individual heterozygosity. All variables in the models are log-scaled.

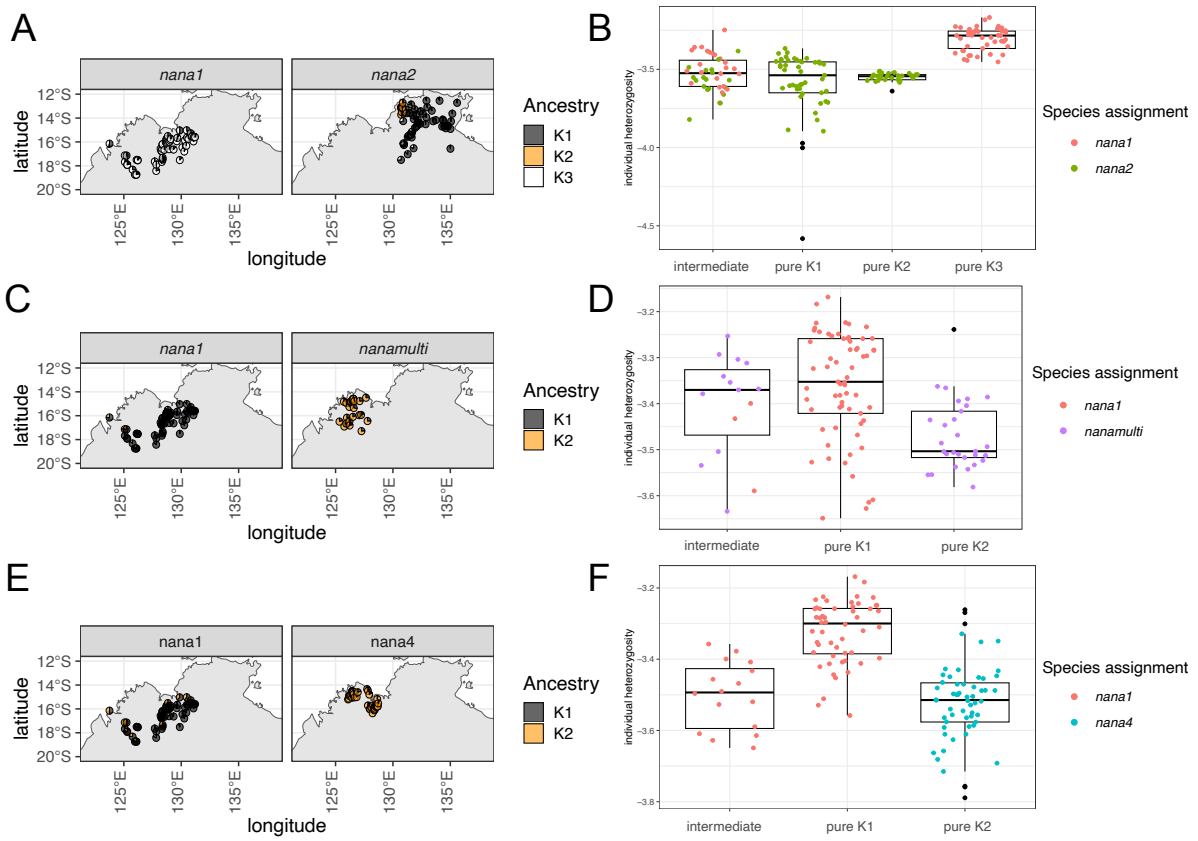

CYsp
catenata
dubia
pamela
arnhemica
lauta
gemina
australis
Groote
lapistola
koira
chimera
ipsa
calcitectus
borroloola
robusta1
robusta2
robusta5
robusta3
robusta4
xenopusBI
xenopusS
xenopusMP
xenopusN
spheniscus
pseudopunctata
granulum
paranana
pluraporosa
nana2
nana4
nana1
nanamulti
girloorloo
kimberleyi
occidentalis_YI
multiporosa
occidentalis_OR
occidentalis_KL
lazelli
pulingka
moritzi
unguiculata
incognita
einasleighensis
purpurascens
fenestrula
finipunctata
punctata
macra
peninsularis
micraU
micraL
media
variegata
pilbara
minuta
versicolor
ocellata
capensis
montium
crypta
australis group
nana group
variegata group
australis group
nana group
0.08

Supplementary Figure 3 Coalescent species tree for 62 species and lineages in the Australian *Gehyra* radiation, estimated from 100 exon-capture loci using a divide-and-conquer approach in Starbeast2. Branch lengths are in proportional coalescent units. Species included in population genomic analysis are highlighted (red = *australis* group, blue = *nana* group). Photos by Scott Macor (top to bottom; *Gehyra lapistola, G. paranana, G. moritzi* and *G. crypta*).



Supplementary Figure 4 Principal components analysis (PCA) showing the assignment of each sample into lineages. Species are grouped by clades based on phylogenetic relationships, A) and B) *granulum, paranana, pluroporosa* and *pseudopunctata*, C) and D) *nana1, nana2, nana4* and *nanamulti*, E) and F) *girloorloo, kimberleyi, occidentalis-KL* and *occidentalis-OR*, G) and H) *arnhemica, australis, gemina* and *lauta*, I) and J), *calcitectus, chimera, ipsa, koira*

and *lapistola*. For each horizonal pair of plots, the left plot shows PC1 vs PC2, and the right shows PC1 vs PC3, as some samples from distinct species overlap in PC1 and PC2. This is likely an artefact of close and more distantly related lineages being analysed together in each PCA plot. All samples were also barcoded with ND2 to verify lineage assignment (data not shown).



Supplementary Figure 5 Heterozygosity is not elevated in individuals of mixed ancestry at three pairwise contact zones. A) Ancestry coefficients inferred in sNMF for *Gehyra nana*1 and *G. nana*2, B) with no significant difference in heterozygosity between individuals assigned an intermediate ancestry. C) Ancestry coefficients inferred in sNMF for *Gehyra nana*1 and *G. nanamulti*, D) with no significant difference in heterozygosity between individuals assigned an intermediate ancestry. E) Ancestry coefficients inferred in sNMF for *Gehyra nana*1 and *G. nana*4, D) with no significant difference in heterozygosity between individuals assigned an intermediate ancestry. Intermediate individuals are those that have < 0.8 ancestry coefficient for any one parental population. Black points on boxplots indicate outliers. K = population clusters based on genetic ancestry.

Supplementary Table 1 Metadata for specimens used in this study for population-level genomics.

See https://figshare.com/s/580e3c70a2a6cfe1dd88

Supplementary Table 2 Metadata for exon capture samples used to generate the coalescent species tree, including sampling strategy for species included in the 'divide-and-conquer' approach.

See https://figshare.com/s/580e3c70a2a6cfe1dd88

Supplementary Table 3 Post-filtering coverage, callable sites and individual heterozygosity across 769 individuals from 21 species and lineages included for population-scale DArT-seq data generation and analysis.

Species (n samples)	Average coverage (exclude sites with < 3 cov.)	Coverage range (exclude sites with < 3 cov.)	Average callable sites (after filtering and vcf expansion)	Average homozygous + invariable sites (after filtering and vcf expansion)	Average heterozygous sites (after filtering and vcf expansion)	Average individual heterozygosity (excluding sites rejected by DeepVar)
arnhemica (n = 24)	11.15	7.69 – 14.94	1893620	1893275	345.29	0.00019
australis (n = 34)	10.45	6.79 – 16.85	2031131	2030628	503.44	0.000245
calcitectus (n = 15)	12.58	9.61 – 14.49	2520328	2520171	156.53	6.12E-05
chimera (n = 46)	10.14	6.05 – 14.65	2555554	2555069	484.57	0.000184
gemina (n = 66)	10.61	6.12 – 13.03	2162013	2161640	372.45	0.000179
girloorloo (n = 12)	7.45	6.57 – 8.65	1587067	1586790	276.92	0.000173
granulum (n = 14)	8.95	6.72 – 10.78	1640657	1640262	395.00	0.000239
ipsa (n = 20)	12.20	8.46 – 15.90	2504637	2504439	198.10	7.97E-05
kimberleyi (n = 18)	8.29	7.29 – 12.40	1646088	1645768	320.06	0.000191
koira (n = 100)	13.93	9.73 – 23.57	2875103	2874384	719.10	0.000241
lapistola (n = 41)	10.76	9.42 – 14.24	3064887	3064395	492.02	0.00016
lauta (n = 24)	11.37	7.24 – 15.75	2142802	2142504	297.16	0.000133
*nana*1 (n = 68)	9.05	6.67 – 11.93	1834702	1833877	824.41	0.000446
*nana*2 (n = 96)	9.58	6.66 – 13.06	1979918	1979358	560.48	0.000278
*nana*4 (n = 58)	8.88	5.59 – 13.69	1713925	1713375	549.79	0.000311
nanamulti (n = 41)	8.73	6.70 – 11.79	1644083	1643458	624.98	0.000367
occis-KL (n = 29)	9.86	7.17 – 11.61	2136125	2135722	402.93	0.000194
occis-OR (n = 10)	9.70	8.58 – 10.58	1867144	1866698	446.00	0.000237

paranana (n = 34)	9.74	6.78 – 15.67	2693385	2692820	564.82	0.000212
pluraporosa (n = 6)	9.59	9.09 – 10.06	2095907	2095495	411.50	0.000191
pseudopunctata (n = 13)	10.38	8.95 – 14.47	2239661	2239428	232.38	0.000107

Supplementary Table 4 Tajima's D and significance across 21 species and lineages included for population-scale DArT-seq data generation and analysis. Bold values indicate significance ($p < 0.05$). P-values and upper/lower bounds are shown both using a normal distribution and a standard normal distribution for comparison.

species	clade	#final filtered SNPs in species cohorts	Tajima's D	upper bound (normal dist.)	lower bound (normal dist.)	p-value (normal dist.)	z-score	p-value (st. normal dist.)	upper bound (st. normal dist.)	lower bound (st. normal dist.)
arnhemica	*australis*	1247	-0.438	1.932	-1.587	0.18	-0.444	0.361	1.69	-1.78
australis	*australis*	1726	-1.156	1.271	-1.66	0.065	-1.314	0.168	1.5	-1.68
calcitectus	*australis*	329	-0.014	1.719	-1.974	0.194	0.151	0.394	1.61	-1.91
chimera	*australis*	2280	-1.019	2.04	-1.81	0.101	-0.984	0.246	1.83	-1.96
gemina	*australis*	2538	-1.591	1.794	-1.591	**0.042**	-1.612	0.109	1.81	-1.92
ipsa	*australis*	640	0.568	1.961	-1.558	0.156	0.606	0.332	1.79	-1.75
koira	*australis*	5631	-1.562	1.808	-1.459	**0.035**	-1.705	0.093	1.86	-1.8
lapistola	*australis*	1704	0.36	1.888	-1.651	0.164	0.48	0.355	1.72	-1.89
lauta	*australis*	1281	-0.577	1.882	-1.789	0.174	-0.442	0.362	1.58	-1.94
girloorloo	*nana*	390	-0.216	1.237	-1.958	0.236	0.031	0.399	1.37	-1.85
granulum	*nana*	484	-0.463	1.695	-2.059	0.18	-0.461	0.359	1.66	-1.78
kimberleyi	*nana*	605	-0.364	1.496	-2.189	0.22	-0.216	0.39	1.46	-1.82
nana1	*nana*	3175	-2.145	1.295	-1.961	**0.002**	-2.523	**0.017**	1.5	-1.69
nana2	*nana*	3056	-2.089	1.962	-1.686	**0.013**	-2.147	**0.04**	1.79	-1.91
nana4	*nana*	1901	-1.941	2.03	-1.592	**0.014**	-2.028	0.051	1.51	-1.88
nanamulti	*nana*	2411	-1.567	1.927	-1.697	**0.045**	-1.554	0.119	1.67	-1.94
occidentalis-KL	*nana*	1375	-1.165	1.715	-1.817	0.07	-1.29	0.174	1.83	-1.78
occidentalis-OR	*nana*	744	0.152	1.456	-1.717	0.204	0.326	0.378	1.55	-1.8
paranana	*nana*	1514	-1.155	1.683	-1.659	0.088	-1.156	0.204	1.46	-1.83
pluraporosa	*nana*	507	0.818	1.597	-1.859	0.122	0.828	0.283	1.88	-1.8
pseudopunctata	*nana*	399	-0.165	1.388	-1.894	0.199	0.034	0.399	1.56	-1.95

Supplementary Table 5 Results from linear and phylogenetic linear models testing the relationship between range size, heterozygosity, variance in heterozygosity and Tajima's D.

Response ~ Predictor		Linear model					Phylogenetic linear model				
		Estimate	Std Error	t-value	p-value	R^2	Estimate	Std Error	t-value	p-value	R^2
(Ho ~ Range size)	**australis group**										
	(Intercept)	-4.53	0.163	-27.837	< 0.001		-4.606	0.214	-21.532	< 0.001	
	final_area	0.162	0.036	4.478	0.003	0.741	0.181	0.044	4.15	0.004	0.711
	nana group										
	(Intercept)	-3.91	0.206	-18.952	< 0.001		-3.834	0.173	-22.15	< 0.001	
	final_area	0.063	0.046	1.354	0.206	0.155	0.032	0.034	0.939	0.37	0.081
	australis + nana groups										
	(Intercept)	-4.183	0.168	-24.966	< 0.001		-4.164	0.269	-15.505	< 0.001	
	final_area	0.107	0.037	2.857	0.01	0.3	0.096	0.031	3.127	0.006	0.34
(Range size ~ Tajima's D)	**australis group**										
	(Intercept)	3.617	0.222	16.267	< 0.001		3.806	0.316	12.061	< 0.001	
	tajD	-1.256	0.231	-5.43	0.001	0.808	-1.131	0.184	-6.131	< 0.001	0.843
	nana group										
	(Intercept)	3.706	0.304	12.173	< 0.001		3.733	0.623	5.995	< 0.001	
	tajD	-0.741	0.242	-3.063	0.012	0.484	-0.919	0.295	-3.11	0.011	0.492
	australis + nana groups										
	(Intercept)	3.689	0.2	18.445	< 0.001		3.766	1.054	3.575	0.002	
	tajD	-0.89	0.175	-5.072	< 0.001	0.575	-1.016	0.177	-5.754	< 0.001	0.635
(Ho ~ Tajima's D)	**australis group**										
	(Intercept)	-3.935	0.066	-59.245	< 0.001		-3.902	0.126	-31.031	< 0.001	
	tajD	-0.189	0.069	-2.735	0.029	0.517	-0.179	0.073	-2.442	0.045	0.46
	nana group										
	(Intercept)	-3.738	0.049	-76.259	< 0.001		-3.743	0.085	-44.007	< 0.001	
	tajD	-0.117	0.039	-3.001	0.013	0.474	-0.075	0.04	-1.85	0.094	0.255
	australis + nana groups										
	(Intercept)	-3.831	0.045	-86.037	< 0.001		-3.818	0.233	-16.394	< 0.001	
	tajD	-0.153	0.039	-3.927	0.001	0.448	-0.123	0.039	-3.147	0.005	0.343
(Variance ~ Range size)	**australis group**										
	(Intercept)	-10.042	0.317	-31.701	< 0.001		-10.074	0.472	-21.331	< 0.001	
	final_area	0.274	0.07	3.887	0.006	0.683	0.282	0.096	2.928	0.022	0.551
	nana group										
	(Intercept)	-9.838	0.435	-22.634	< 0.001		-10.207	0.408	-25	< 0.001	
	final_area	0.322	0.098	3.302	0.008	0.522	0.399	0.081	4.951	0.001	0.71
	australis + nana groups										
	(Intercept)	-9.905	0.34	-29.156	< 0.001		-10.169	0.528	-19.277	< 0.001	
	final_area	0.297	0.076	3.905	0.001	0.445	0.349	0.06	5.786	< 0.001	0.638

Supplementary Table 6 D-suite introgression statistics for the *australis* group based on 3,557 cohort-genotyped SNPs, showing no significant evidence for introgression.

P1	P2	P3	D-statistic	Z-score (from adj. p-value)	adj. p-value	f4-ratio	BBAA	ABBA	BABA	Original Z-score	Original p-value
arnhemica	*australis*	*calcitectus*	0.113	0.655	0.512	0.041	44.619	5.815	4.639	0.948	0.343
arnhemica	*australis*	*chimera*	0.214	1.466	0.143	0.064	47.640	6.432	4.168	2.100	0.036
australis	*gemina*	*arnhemica*	0.022	0.295	0.768	0.024	23.483	14.623	14.007	0.423	0.672
arnhemica	*australis*	*ipsa*	0.224	1.296	0.195	0.051	49.001	5.513	3.498	1.913	0.056
arnhemica	*australis*	*koira*	0.116	0.801	0.423	0.020	43.971	6.460	5.116	1.249	0.212
australis	*lauta*	*arnhemica*	0.106	1.332	0.183	0.128	18.626	17.015	13.740	1.969	0.049
chimera	*calcitectus*	*arnhemica*	0.151	0.812	0.417	0.041	16.095	10.438	7.702	1.322	0.186
arnhemica	*gemina*	*calcitectus*	0.080	0.539	0.590	0.029	45.514	5.670	4.825	0.752	0.452
ipsa	*calcitectus*	*arnhemica*	0.270	1.466	0.143	0.068	16.616	11.189	6.429	2.154	0.031
calcitectus	*koira*	*arnhemica*	0.093	0.619	0.536	0.029	17.831	11.064	9.176	0.855	0.392
arnhemica	*lauta*	*calcitectus*	0.272	1.812	0.070	0.087	46.721	5.859	3.356	2.876	0.004
arnhemica	*gemina*	*chimera*	0.107	0.734	0.463	0.034	49.080	6.191	4.991	1.078	0.281
ipsa	*chimera*	*arnhemica*	0.104	0.539	0.590	0.029	18.403	10.795	8.764	0.750	0.453
chimera	*koira*	*arnhemica*	0.195	1.273	0.203	0.068	15.005	14.159	9.542	1.869	0.062
arnhemica	*lauta*	*chimera*	0.139	0.787	0.431	0.038	50.947	5.561	4.201	1.159	0.246
arnhemica	*gemina*	*ipsa*	0.017	0.112	0.911	0.004	51.014	5.066	4.894	0.161	0.872
arnhemica	*gemina*	*koira*	0.015	0.112	0.911	0.003	45.343	6.052	5.871	0.138	0.891
gemina	*lauta*	*arnhemica*	0.085	0.925	0.355	0.107	19.903	16.883	14.226	1.478	0.140
ipsa	*koira*	*arnhemica*	0.331	1.812	0.070	0.094	18.201	13.368	6.720	2.721	0.007
arnhemica	*lauta*	*ipsa*	0.230	1.466	0.143	0.052	52.225	5.507	3.448	2.111	0.035
arnhemica	*lauta*	*koira*	0.034	0.228	0.820	0.005	46.630	4.819	4.500	0.342	0.732
chimera	*calcitectus*	*australis*	0.082	0.512	0.608	0.023	15.256	10.775	9.134	0.710	0.478
gemina	*australis*	*calcitectus*	0.037	0.318	0.751	0.012	53.062	4.697	4.366	0.463	0.643
ipsa	*calcitectus*	*australis*	0.208	1.466	0.143	0.054	15.634	11.383	7.463	2.126	0.033
calcitectus	*koira*	*australis*	0.104	0.787	0.431	0.030	16.283	10.850	8.804	1.165	0.244
australis	*lauta*	*calcitectus*	0.163	0.801	0.423	0.048	47.205	4.738	3.411	1.275	0.202
gemina	*australis*	*chimera*	0.127	0.902	0.367	0.031	55.391	4.706	3.643	1.437	0.151
ipsa	*chimera*	*australis*	0.115	0.628	0.530	0.031	16.436	11.092	8.812	0.898	0.369
chimera	*koira*	*australis*	0.159	1.157	0.247	0.052	13.014	13.512	9.815	1.704	0.088
lauta	*australis*	*chimera*	0.107	0.706	0.480	0.026	50.756	4.662	3.758	1.036	0.300
gemina	*australis*	*ipsa*	0.233	1.812	0.070	0.047	57.841	4.877	3.033	2.689	0.007
gemina	*australis*	*koira*	0.123	1.061	0.289	0.017	52.318	5.331	4.167	1.606	0.108
australis	*gemina*	*lauta*	0.042	0.655	0.512	0.073	20.598	17.415	16.007	0.939	0.348
ipsa	*koira*	*australis*	0.287	1.812	0.070	0.081	16.898	13.410	7.433	2.981	0.003
australis	*lauta*	*ipsa*	0.006	0.049	0.961	0.001	52.313	3.983	3.939	0.049	0.961
lauta	*australis*	*koira*	0.119	0.762	0.446	0.015	47.213	4.817	3.791	1.120	0.263
chimera	*calcitectus*	*gemina*	0.119	0.690	0.490	0.032	15.995	11.182	8.810	1.005	0.315

ipsa	*calcitectus*	*chimera*	0.054	0.425	0.671	0.045	11.597	11.828	10.607	0.596	0.551
chimera	*calcitectus*	*koira*	0.242	1.466	0.143	0.087	12.204	15.202	9.280	2.201	0.028
chimera	*calcitectus*	*lauta*	0.192	1.157	0.247	0.058	15.194	12.040	8.168	1.718	0.086
ipsa	*calcitectus*	*gemina*	0.283	1.882	0.060	0.071	16.901	12.319	6.887	3.272	0.001
calcitectus	*koira*	*gemina*	0.060	0.440	0.660	0.017	17.422	10.835	9.611	0.627	0.530
gemina	*lauta*	*calcitectus*	0.181	1.181	0.238	0.060	49.367	5.412	3.754	1.772	0.076
ipsa	*calcitectus*	*koira*	0.077	0.619	0.536	0.030	10.295	13.523	11.589	0.862	0.389
ipsa	*calcitectus*	*lauta*	0.257	1.785	0.074	0.077	15.645	12.722	7.518	2.602	0.009
koira	*calcitectus*	*lauta*	0.015	0.112	0.911	0.005	16.302	10.150	9.855	0.132	0.895
ipsa	*chimera*	*gemina*	0.147	0.787	0.431	0.040	18.324	11.917	8.857	1.187	0.235
chimera	*koira*	*gemina*	0.151	0.794	0.427	0.049	14.411	13.745	10.148	1.223	0.221
gemina	*lauta*	*chimera*	0.017	0.112	0.911	0.005	53.094	4.687	4.528	0.159	0.874
chimera	*ipsa*	*koira*	0.161	0.848	0.396	0.059	12.498	14.339	10.353	1.373	0.170
ipsa	*chimera*	*lauta*	0.065	0.358	0.720	0.020	17.145	10.897	9.565	0.518	0.605
chimera	*koira*	*lauta*	0.153	0.801	0.423	0.053	13.983	13.456	9.880	1.259	0.208
ipsa	*koira*	*gemina*	0.313	1.812	0.070	0.086	18.601	13.949	7.292	2.674	0.008
gemina	*lauta*	*ipsa*	0.242	1.466	0.143	0.048	54.589	4.851	2.964	2.269	0.023
gemina	*lauta*	*koira*	0.015	0.112	0.911	0.002	49.382	4.552	4.413	0.157	0.875
ipsa	*koira*	*lauta*	0.235	1.466	0.143	0.072	17.397	12.883	7.976	2.139	0.032

Supplementary Table 7 D-suite introgression statistics for a subset of lineages in the *nana* group that occur in contemporary contact, showing evidence for introgression in some pairs ($p < 0.01$). Table A) is based on 1,934 SNPs from *nana* 'group1', corresponding to Figure 3A, and table B) is based on 1,299 SNPs from *nana* 'group2', corresponding to Figure 3B. Significant results indicated in bold.

A)

P1	P2	P3	D-statistic	Z-score (corrected)	p-value (corrected)	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
nana4	*nana1*	*nana2*	0.544	7.924	**0.000**	8.402	0.000	0.342	4.550	10.738	3.168
nanamulti	*nana1*	*nana2*	0.340	3.984	**0.000**	4.351	0.000	0.219	5.713	8.128	4.004
nana2	*nana1*	*pluraporosa*	0.123	0.780	0.435	1.124	0.261	0.085	11.621	2.866	2.236
nanamulti	*nana1*	*nana4*	0.021	0.194	0.846	0.217	0.828	0.029	8.287	4.514	4.332
nana1	*nana4*	*pluraporosa*	0.088	0.553	0.580	0.732	0.464	0.092	5.941	3.832	3.211
nanamulti	*nana1*	*pluraporosa*	0.097	0.646	0.518	0.910	0.363	0.070	9.463	2.956	2.434
nana4	*nanamulti*	*nana2*	0.312	2.857	**0.004**	3.220	0.001	0.156	5.006	7.251	3.805
nana2	*nana4*	*pluraporosa*	0.177	0.780	0.435	1.261	0.207	0.168	4.883	4.151	2.901
nana2	*nanamulti*	*pluraporosa*	0.022	0.194	0.846	0.194	0.846	0.015	7.858	2.531	2.423
nanamulti	*nana4*	*pluraporosa*	0.159	0.780	0.435	1.147	0.251	0.154	6.100	4.169	3.027

B)

P1	P2	P3	D-statistic	Z-score (corrected)	p-value (corrected)	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
girloorloo	*kimberleyi*	*granulum*	0.525	3.293	**0.001**	4.136	0.000	0.288	1.422	2.236	0.696
girloorloo	*nana1*	*granulum*	0.463	3.049	**0.002**	3.769	0.000	0.311	2.874	2.630	0.965
girloorloo	*nanamulti*	*granulum*	0.554	3.334	**0.001**	4.324	0.000	0.412	3.268	3.090	0.886
girloorloo	*occidentalis-KL*	*granulum*	0.558	3.284	**0.001**	4.034	0.000	0.309	1.294	2.309	0.655
girloorloo	*occidentalis-OR*	*granulum*	0.448	2.490	0.013	3.051	0.002	0.275	1.972	2.381	0.907
granulum	*pseudopunctata*	*girloorloo*	0.417	2.304	0.021	2.776	0.006	0.196	2.978	1.368	0.563
kimberleyi	*girloorloo*	*nana1*	0.025	0.061	0.951	0.190	0.849	0.014	1.470	2.652	2.522
kimberleyi	*girloorloo*	*nanamulti*	0.107	0.492	0.622	0.746	0.456	0.044	1.329	2.985	2.409
girloorloo	*kimberleyi*	*occidentalis-KL*	0.163	0.760	0.448	1.012	0.312	0.048	1.428	1.864	1.341

kimberleyi	*girloorloo*	*occidentalis-OR*	0.173	0.775	0.439	1.042	0.298	0.045	1.701	2.041	1.438
girloorloo	*kimberleyi*	*pseudopunctata*	0.256	1.076	0.282	1.357	0.175	0.129	1.080	1.956	1.159
nana1	*nanamulti*	*girloorloo*	0.147	1.114	0.265	1.451	0.147	0.225	7.651	1.848	1.374
girloorloo	*occidentalis-KL*	*nana1*	0.393	2.304	0.021	2.765	0.006	0.303	0.955	5.102	2.225
girloorloo	*occidentalis-OR*	*nana1*	0.425	2.443	0.015	2.978	0.003	0.309	1.146	4.928	1.989
girloorloo	*nana1*	*pseudopunctata*	0.357	2.443	0.015	2.955	0.003	0.233	2.398	2.733	1.295
girloorloo	*occidentalis-KL*	*nanamulti*	0.329	1.651	0.099	2.013	0.044	0.202	0.860	5.162	2.603
girloorloo	*occidentalis-OR*	*nanamulti*	0.345	1.689	0.091	2.099	0.036	0.189	0.951	4.651	2.267
girloorloo	*nanamulti*	*pseudopunctata*	0.551	2.975	**0.003**	3.651	0.000	0.418	2.625	3.626	1.049
occidentalis-KL	*occidentalis-OR*	*girloorloo*	0.207	1.076	0.282	1.353	0.176	0.123	9.632	1.245	0.817
girloorloo	*occidentalis-KL*	*pseudopunctata*	0.429	2.683	**0.007**	3.305	0.001	0.288	1.011	2.950	1.178
girloorloo	*occidentalis-OR*	*pseudopunctata*	0.223	1.076	0.282	1.365	0.172	0.137	1.729	2.314	1.469
kimberleyi	*nana1*	*granulum*	0.041	0.162	0.872	0.346	0.730	0.032	1.713	1.600	1.475
kimberleyi	*nanamulti*	*granulum*	0.223	1.465	0.143	1.836	0.066	0.172	1.426	1.823	1.160
kimberleyi	*occidentalis-KL*	*granulum*	0.045	0.162	0.872	0.343	0.732	0.029	0.817	1.310	1.196
occidentalis-OR	*kimberleyi*	*granulum*	0.016	0.054	0.957	0.121	0.904	0.017	0.984	2.061	1.995
granulum	*pseudopunctata*	*kimberleyi*	0.031	0.065	0.948	0.215	0.830	0.013	1.858	1.045	0.983
nana1	*nanamulti*	*granulum*	0.157	1.210	0.226	1.567	0.117	0.143	7.970	1.989	1.450
occidentalis-KL	*nana1*	*granulum*	0.004	0.044	0.965	0.044	0.965	0.003	4.587	1.466	1.455
occidentalis-OR	*nana1*	*granulum*	0.050	0.162	0.872	0.426	0.670	0.050	5.204	2.022	1.831
granulum	*pseudopunctata*	*nana1*	0.176	1.210	0.226	1.574	0.115	0.060	2.108	1.936	1.357
occidentalis-KL	*nanamulti*	*granulum*	0.191	1.202	0.229	1.543	0.123	0.149	4.454	1.717	1.167
occidentalis-OR	*nanamulti*	*granulum*	0.182	1.076	0.282	1.370	0.171	0.191	4.929	2.367	1.637
granulum	*pseudopunctata*	*nanamulti*	0.323	1.732	0.083	2.154	0.031	0.093	1.446	2.413	1.234
occidentalis-OR	*occidentalis-KL*	*granulum*	0.058	0.175	0.861	0.459	0.646	0.047	9.448	1.648	1.468
granulum	*pseudopunctata*	*occidentalis-KL*	0.307	1.838	0.066	2.304	0.021	0.089	1.806	1.969	1.045
granulum	*pseudopunctata*	*occidentalis-OR*	0.049	0.162	0.872	0.357	0.721	0.014	2.657	1.892	1.716
nana1	*nanamulti*	*kimberleyi*	0.009	0.054	0.957	0.097	0.922	0.006	7.923	1.544	1.516
kimberleyi	*occidentalis-KL*	*nana1*	0.441	2.701	**0.007**	3.374	0.001	0.313	1.287	4.911	1.903
kimberleyi	*occidentalis-OR*	*nana1*	0.371	2.392	0.017	2.887	0.004	0.319	1.284	5.668	2.599
kimberleyi	*nana1*	*pseudopunctata*	0.191	1.038	0.299	1.303	0.193	0.124	1.533	1.998	1.357
kimberleyi	*occidentalis-KL*	*nanamulti*	0.491	2.596	**0.009**	3.171	0.002	0.238	0.978	4.758	1.623

kimberleyi	*occidentalis-OR*	*nanamulti*	0.373	1.734	0.083	2.176	0.030	0.224	1.141	5.443	2.484
kimberleyi	*nanamulti*	*pseudopunctata*	0.511	2.683	**0.007**	3.279	0.001	0.343	1.055	2.632	0.852
occidentalis-OR	*occidentalis-KL*	*kimberleyi*	0.314	1.668	0.095	2.045	0.041	0.120	9.751	1.459	0.761
kimberleyi	*occidentalis-KL*	*pseudopunctata*	0.277	1.668	0.095	2.055	0.040	0.188	0.832	2.249	1.273
kimberleyi	*occidentalis-OR*	*pseudopunctata*	0.011	0.044	0.965	0.059	0.953	0.009	0.974	2.161	2.113
nana1	*nanamulti*	*occidentalis-KL*	0.034	0.162	0.872	0.380	0.704	0.023	5.628	2.383	2.228
nanamulti	*nana1*	*occidentalis-OR*	0.016	0.054	0.957	0.161	0.872	0.009	5.959	2.621	2.540
nana1	*nanamulti*	*pseudopunctata*	0.302	1.812	0.070	2.261	0.024	0.243	7.260	2.458	1.318
occidentalis-KL	*occidentalis-OR*	*nana1*	0.012	0.054	0.957	0.097	0.923	0.009	7.202	2.596	2.534
nana1	*occidentalis-KL*	*pseudopunctata*	0.101	0.666	0.505	0.913	0.361	0.071	4.035	1.827	1.493
occidentalis-OR	*nana1*	*pseudopunctata*	0.158	0.882	0.378	1.151	0.250	0.112	4.777	2.173	1.580
occidentalis-OR	*occidentalis-KL*	*nanamulti*	0.036	0.065	0.948	0.234	0.815	0.017	7.024	2.511	2.337
occidentalis-KL	*nanamulti*	*pseudopunctata*	0.230	1.210	0.226	1.595	0.111	0.186	3.713	2.155	1.350
occidentalis-OR	*nanamulti*	*pseudopunctata*	0.436	2.204	0.028	2.637	0.008	0.326	4.235	2.852	1.119
occidentalis-OR	*occidentalis-KL*	*pseudopunctata*	0.364	2.204	0.027	2.658	0.008	0.174	8.615	1.739	0.812