

Limited Introgression between Rock-Wallabies with Extensive Chromosomal Rearrangements

Sally Potter ^{*},^{1,2} Jason G. Bragg,³ Rustamzhon Turakulov,¹ Mark D.B. Eldridge,² Janine Deakin,⁴ Mark Kirkpatrick ⁵, Richard J. Edwards ⁶ and Craig Moritz¹

¹Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

²Australian Museum Research Institute, Australian Museum, Sydney, NSW, Australia

³National Herbarium of New South Wales, The Royal Botanical Gardens and Domain Trust, Sydney, NSW, Australia

⁴Institute for Applied Ecology, University of Canberra, Bruce, ACT, Australia

⁵Department of Integrative Biology, University of Texas, Austin, TX, USA

⁶School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

*Corresponding author: E-mail: sally.potter@anu.edu.au.

Associate editor: Anne Yoder

Abstract

Chromosome rearrangements can result in the rapid evolution of hybrid incompatibilities. Robertsonian fusions, particularly those with monobrachial homology, can drive reproductive isolation amongst recently diverged taxa. The recent radiation of rock-wallabies (genus *Petrogale*) is an important model to explore the role of Robertsonian fusions in speciation. Here, we pursue that goal using an extensive sampling of populations and genomes of *Petrogale* from north-eastern Australia. In contrast to previous assessments using mitochondrial DNA or nuclear microsatellite loci, genomic data are able to separate the most closely related species and to resolve their divergence histories. Both phylogenetic and population genetic analyses indicate introgression between two species that differ by a single Robertsonian fusion. Based on the available data, there is also evidence for introgression between two species which share complex chromosomal rearrangements. However, the remaining results show no consistent signature of introgression amongst species pairs and where evident, indicate generally low introgression overall. X-linked loci have elevated divergence compared with autosomal loci indicating a potential role for genic evolution to produce reproductive isolation in concert with chromosome change. Our results highlight the value of genome scale data in evaluating the role of Robertsonian fusions and structural variation in divergence, speciation, and patterns of molecular evolution.

Key words: chromosome rearrangement, introgression, marsupial, speciation, Robertsonian fusion.

Introduction

There is a long history of theory and empirical work investigating the role of chromosome rearrangements on divergence and speciation. Links between chromosome evolution and rapid divergence and speciation have been outlined across diverse taxa (e.g., Leaché et al. 2016; de Vos et al. 2020; Wang et al. 2020). Classical models of chromosomal evolution focus on hybrid sterility resulting from the unbalanced gametes produced by hybrids that are heterozygous for chromosome rearrangements (White 1969; King 1993). Although supported in various systems (see Brown and O'Neill [2010]), this model has been challenged due to the difficulty for underdominant rearrangements to fix within populations (see Hoffman and Rieseberg [2008]; Faria and Navarro [2010]). The alternate model focusses on the suppression of recombination in heterokaryotypes within or surrounding a rearrangement, and subsequent accumulation of alleles within the region of alleles causing local adaptation and/or reproductive isolation (Noor et al. 2001; Rieseberg

2001). This model has been explored theoretically (Navarro and Barton 2003; Kirkpatrick and Barton 2006; Guerrero and Kirkpatrick 2014; Dagitlis and Kirkpatrick 2016) and has gained substantial empirical support from diverse organisms, including plants (e.g., Rieseberg et al. 1995; Rieseberg 2001) and insects (e.g., Noor et al. 2007; Manoukis et al. 2008; see also Ayala and Coluzzi [2005]). Understanding how chromosome differences between populations can suppress gene flow can be challenging. Some rearrangements may have little or no effect on fertility of hybrids (Rieseberg 2001). Thus, the number of rearrangements by which two species differ may be an inaccurate gauge of how much the rearrangements contribute to their isolation. Rearrangements can themselves act as Bateson Dobzhansky Muller Incompatibilities (Bateson 1909; Dobzhansky 1936; Muller 1942) or establish genic incompatibilities by the alleles that they carry. Alternatively, isolation might be entirely attributable to genic incompatibilities, with no contribution from the structural changes themselves (Coyne and Orr 2004). Further empirical work across diverse

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

systems is required to understand the role of structural variation in divergence and speciation; in particular, whether different chromosome rearrangements (e.g., inversion vs. fusion) have different mechanisms for driving reproductive isolation.

One type of chromosomal rearrangement, Robertsonian fusions, is particularly interesting in the context of chromosomal speciation. These occur when two acrocentric chromosomes fuse at their centromeres to form a new metacentric chromosome. A single Robertsonian fusion may have little impact on hybrid fitness because the fused and unfused chromosomes are able to pair successfully, by forming a trivalent that is able to segregate normally (e.g., in shrews and mice; reviewed in Searle [1993]; Garagna et al. [2014]; Searle et al. [2019]). Sterility and barriers to gene flow can evolve, however, when different populations fix for different fusions that involve the same chromosome arms (monobrachial homology). This situation can lead to complex chains of chromosomes in hybrid meiosis, causing segregation errors that lead to sterility (Gropp et al. 1982; Baverstock et al. 1983; Baker and Bickham 1986; Garagna et al. 2014). Longer chains of chromosomes are thought to generally show lower fertility (Searle 1993; Basset et al. 2019). Thus, establishment of such complex combinations of rearrangements is expected to cause stronger reproductive isolation.

Two prominent organismal systems used to explore the role of Robertsonian fusions in divergence and speciation are mice (genus *Mus*; Giménez et al. 2013, Garagna et al. 2014, Britton-Davidian et al. 2017) and shrews (genus *Sorex*; see Searle et al. [2019]). The insight gained from these systems has come from in depth cytogenetic research, fertility, and meiotic studies and importantly, detailed research on the molecular composition and organization of centromeric regions across contact zones as well as between species (reviewed in Garagna et al. [2014]; Searle et al. [2019]). The recent divergence, multiple parallel pairwise comparisons between populations and/or species with varied combinations of Robertsonian fusions and the ability to conduct laboratory research has revealed a lot. However, such deep experimental evidence is lacking for most organisms with extensive chromosome change. New genomic tools and evidence are expected to improve our understanding of chromosome change and speciation for diverse organisms (e.g., Potter et al. 2017; Campbell et al. 2018; Mérot et al. 2020), and for some systems have already done so (Carbone et al. 2014; Franchini et al. 2020); yet most exemplars of rapid chromosome change are still to be examined at a genome scale. To avoid conflating the contributions to reproductive isolation of genic evolution with chromosome change, we need to focus genome-scale analyses on very recently diverged species groups (Coyne and Orr 2004). Here, we highlight an approach to evaluate the role of chromosomal structure in the history of divergence and speciation, as well as molecular evolution processes in a group of species with a complex evolutionary and taxonomic history due to recent divergence and complex chromosome rearrangements.

The rock-wallabies (genus *Petrogale*) are a group of Australian marsupials that have long been promoted as an

example of rapid chromosomal speciation caused by hybrid dysfunction (Sharman et al. 1989; King 1993; Brown and O'Neill 2010). This radiation is young, between 0.44 and 1.6 My old (Potter et al. 2012). Rock-wallabies typically live in small, isolated populations, a lifestyle that may facilitate fixation of underdominant chromosome rearrangements by genetic drift (Eldridge and Close 1993). The unusual population structure of *Petrogale*, together with their extensive chromosomal rearrangements, make this group an interesting contrast to other model systems in studies of the role of structural variation in speciation. The 17 recognized species of *Petrogale* comprise 23 distinct karyotypes, which is the greatest chromosomal diversity in any group of marsupial (Eldridge and Close 1993). Most rearrangements in *Petrogale* are Robertsonian fusions that are fixed between species, and the genus is unusual for having numerous examples of monobrachial homology (see Eldridge and Close [1993]). Inversions and centric shifts are also present (Eldridge et al. 1989, 1990, 1991, 1992; Eldridge and Close 1992, 1993), and some such rearrangements are polymorphic within species (Eldridge and Close 1997; refer to fig. 1). Given the diverse chromosomal rearrangements and little time for confounding effects of genic divergence, *Petrogale* is ripe for investigations into the effects of chromosomal rearrangements on speciation.

Six species of *Petrogale* that form the *penicillata* complex (Eldridge et al. 1991) have extensive differences in chromosome structure (fig. 1). They are distributed parapatrically, and rare chromosomal hybrids have been found at contact zones (Briscoe et al. 1982). Some species differ by simple Robertsonian fusions (F1's having chains of three chromosomes at meiosis), whereas others have more complex differences (monobrachial homology with F1's having chains of four or five). Experimental hybrid crosses result in infertile males, which show extensive mispairing in meiosis, and partially fertile F1 females (Sharman et al. 1989; Close et al. 1996; Close and Bell 1997). These results suggest that postzygotic isolation is strong.

The stronger sterility in *Petrogale* males conforms to Haldane's rule, whereby in crosses between two taxa, the heterogametic sex is more likely to be sterile or inviable (Haldane 1922). Both Haldane's rule and meiotic sex chromosome inactivation have been linked to a disproportionately large role of the X chromosome in hybrid incompatibilities (the large X-effect; Coyne and Orr 1989; Masly and Presgraves 2007). However, in mice where there is extensive evidence for the large X-effect, it is unclear if chromosome rearrangements are important to it (Larson et al. 2017). Recent analysis of X loci in *Petrogale* (Potter et al. 2017) identified greater phylogenetic resolution on the X compared with autosomes, as well as greater divergence (i.e., faster X effect; Vicoso and Charlesworth 2006). Comparisons of divergence across the genome will be important in understanding the role of sex chromosome versus autosomal evolution in reproductive isolation.

Existing molecular data have given an inconclusive picture of gene flow between species in *Petrogale*. Mitochondrial DNA (mtDNA) phylogenies show strong discordance with

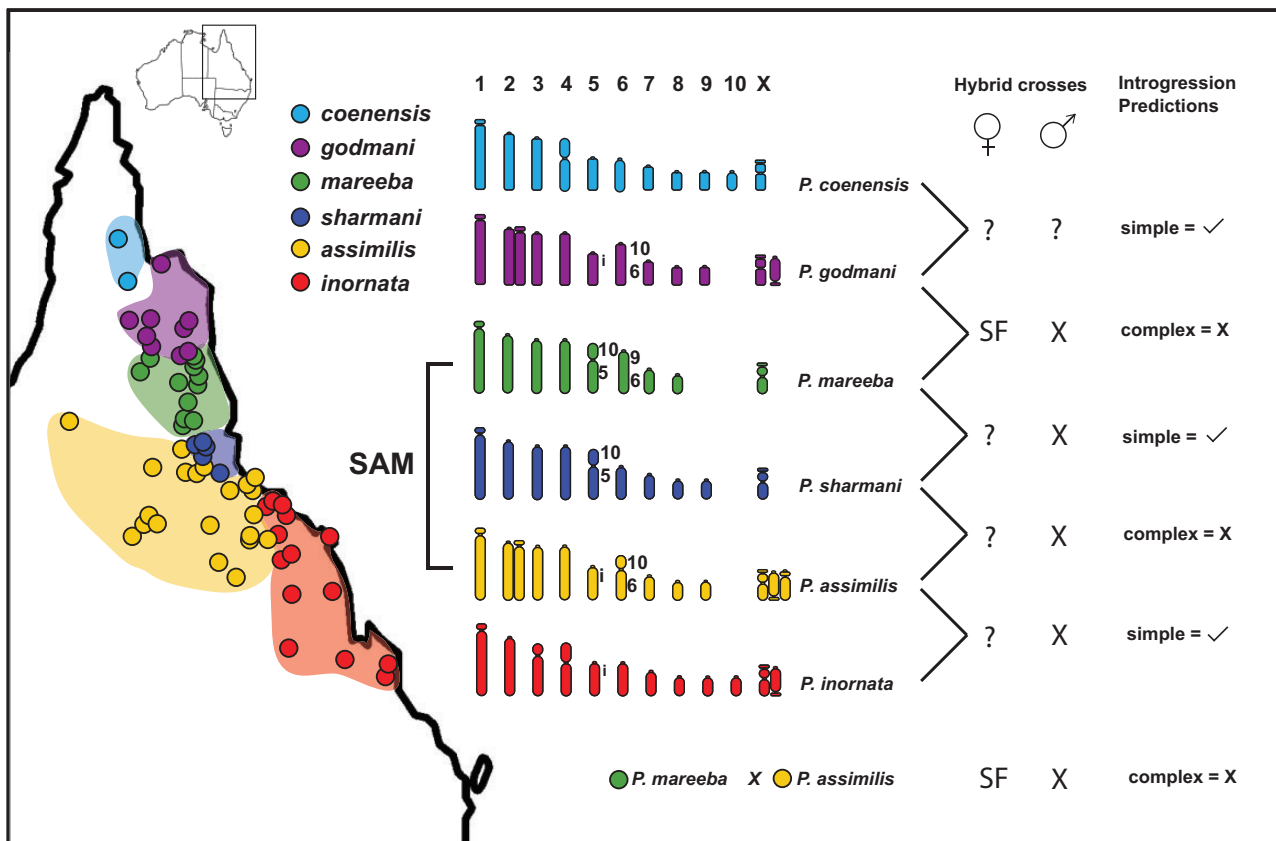


Fig. 1. Karyotypes of six species of rock-wallaby (*Petrogale*) from north-eastern Australia which form part of the *penicillata* complex, including a map of sample locations and distributions for each species. The color of individuals on the map is species specific and linked to the karyotypes. Robertsonian fusions are highlighted by chromosome numbers, inversions are denoted as (i), and polymorphic karyotypes are shown for chromosome 2 and the X chromosome. The SAM (*P. sharmani*, *P. assimilis*, *P. mareeba*) complex is outlined. The outcomes of experimental hybrid crosses are outlined for both males and females: X = infertile; SF = subfertile; ? = unknown. Based on simple versus complex heterozygote formation between species pairs, we have predicted whether we expect to see introgression or not. For simple heterozygotes which form trivalents during meiosis, we expect introgression could be present as chromosomes could segregate properly during meiosis of hybrids. For complex heterozygotes which form chains of four or five chromosomes, we expect to see no introgression.

species boundaries as defined by karyotypes and hybrid sterility (Potter et al. 2015, 2017). This discord could result from introgression or from incomplete lineage sorting. Coalescent analysis of microsatellites inferred high levels of introgression, even among the most chromosomally divergent taxa (Potter et al. 2015). Allozymes, mtDNA, microsatellites (Briscoe et al. 1982; Bee and Close 1993; Potter et al. 2015), and (most recently) a sample of some 2,000 nuclear exons (Potter et al. 2017) have only partially resolved the evolutionary relationships of these six species.

Proper evaluation of the competing hypotheses for the roles of chromosome rearrangements in speciation demands a knowledge of the history of divergence among species. But these relationships are challenging to estimate when there is extensive discordance between gene trees resulting from introgression and/or incomplete lineage sorting (Maddison 1997; Edwards 2009; Linkem et al. 2016). In some cases, introgression is so extensive that only a small fraction of the genome shares an evolutionary history that is consistent with the species phylogeny (Fontaine et al. 2015; Edelman et al. 2019; McGee et al. 2020). However, approaches are now available to test for introgression separate from incomplete

lineage sorting, using both population and phylogenomic approaches (e.g., Pickrell and Pritchard 2012; Pease and Hahn 2015; Wen et al. 2018). How taxa are sampled can also affect the ability to resolve recent population divergence with gene flow; sampling of multiple geographically dispersed individuals per taxon should result in improved parameter estimates compared with just one or two individuals (Robinson et al. 2014; McLaughlin and Winker 2020; e.g., Teng et al. 2017; Linck et al. 2019).

Here, using population genomic and phylogenomic approaches, we resolve divergence histories among these closely related taxa and test for introgression while accounting for incomplete lineage sorting. We expect that introgression will be most strongly reduced between species with complex monobrachial homology (chains of four or more chromosomes in hybrid meiosis) compared with hybrids between species with simple chromosomal differences (fig. 1). We found varied evidence for introgression between species pairs, with the most extensive support found between two species which share a simple Robertsonian fusion (*P. assimilis* and *P. inornata*). There was evidence of introgression between other species pairs where complex heterozygotes would form:

P. godmani and *P. mareeba* which have a known contact zone and F1 hybrids; and less compelling evidence between *P. assimilis* and *P. sharmani*. As we expected based on infertility of male hybrids and potential links to Haldane's rule, we found the X chromosome to have greater divergence than loci from autosomes.

Results

Genomic Data Sets

We used the draft genome assembly for *Petrogale penicillata* to serve as a reference for mapping reads and calling single-nucleotide polymorphisms (SNPs; see Materials and Methods). We generated three data sets, two exon sequence data sets for the six species in the *penicillata* group and one SNP data set for five species (excluding *P. coenensis*). The SNP data set we refer to as "DART" based on the genome reduction approach used (see Materials and Methods). It genotyped 22,724 SNPs from 77 individuals with a mean of 15 per taxon (including four known F1 *P. godmani* × *P. mareeba* hybrids). The second data set is referred to as "phased exons" because it consists of phased haplotype sequences (627,699 bp) from exon capture experiments (see [supplementary table S1, Supplementary Material](#) online, for sampling). It resolved haplotypes at 1,215 loci for 67 individuals (mean of 15 per taxon). The third data set is referred to as "unphased exons" and is based on sequence data (843,619 bp) from exon capture experiments, and includes 1,617 loci on those same individuals. Except where noted, all the analyses reported here are based on the phased exons. A set of 583 loci were successfully mapped to the tammar wallaby (*Notamacropus eugenii*) genome ([supplementary table S2, Supplementary Material](#) online). These were sorted into three genomic categories: X-linked loci ($n = 45$); loci on rearranged autosomes, that is, those involved in fusions (chromosomes 5, 6, 9, 10; $n = 108$); and loci on nonrearranged autosomes not involved in fusions (chromosomes 1, 2, 7, 8; $n = 229$). Unfortunately, we are not able to identify which SNPs fall within rearranged parts of the chromosomes.

Population Structure and Phylogenetic Relationships

Principal coordinate analysis (PCoA) based on the DART data set clearly separate out each of *P. godmani* and *P. inornata* from the clade of *P. sharmani*, *P. assimilis*, and *P. mareeba*, hereon referred to as the *sharmani*–*assimilis*–*mareeba* (SAM) clade ([fig. 2a](#)). The three SAM species could not be further subdivided based on this analysis. Four F1 hybrids between *P. godmani* and *P. mareeba* that were identified previously by karyotype form a cluster intermediate between those two species. By contrast, discriminant analysis of principal components ([Jombart 2008, Jombart et al. 2010](#)) correctly assigned individuals to the three species. This result is based on 19 PCs (63.1% of conserved variance) and two discriminant eigenvalues ([fig. 2b](#)).

Bayesian clustering of individuals using STRUCTURE ([Pritchard et al. 2000](#)) also failed to resolve the SAM clade. With the full DART data set, $K = 4$ clusters were identified. Despite this, only *P. inornata* and *P. godmani* are separated,

and the closely related SAM clade forms a third cluster ([fig. 2a](#)). When these three species were analyzed alone, two clusters ($K = 2$) were identified but with different outcomes across replicate runs. With $K = 3$, we again fail to separate the three species, and the groupings are identical to those with $K = 2$ ([fig. 2b](#)).

We estimated the phylogenetic relationships amongst species, applying SVDquartets and a species tree approach with the phased exon data set at both species- and individual levels. In the species-tree analysis, there is strong support for *P. godmani* and *P. coenensis* forming a clade, and *P. inornata* is most likely the sister lineage to the SAM clade, which itself is strongly supported ([fig. 2c](#)). These patterns also emerge in examining divergence histories of individuals, although the relationships amongst the three species in the SAM clade are unresolved, and most relationships have low bootstrap support (<50%; [supplementary fig. S1, Supplementary Material](#) online). These results are consistent with, but improve on, previous estimates of the phylogeny based on exon sequences with just two individuals per species ([Potter et al. 2017](#)). The *SplitsTree* network ([Bryant and Moulton 2002; Huson and Bryant 2006](#)) correctly assigns individuals to species, but with a few exceptions in the SAM clade ([fig. 2d](#)). The network indicates introgression among species, or incomplete lineage sorting, manifest as webs in the network.

We analyzed divergence, measured as d_{XY} , between the four most closely related species (*P. inornata* and the SAM clade). Consistent with our expectations, we find that divergence is significantly greater on the X than on the autosomes ([fig. 3a](#)). Inconsistent with expectations, we find greater divergence in nonrearranged than in rearranged chromosomes ([fig. 3a](#)). Restricting the analysis to just the three species in the SAM clade shows the same trends and significant differences, although with lower d_{XY} values ([fig. 3b](#) and [supplementary fig. S2](#) and [table S3, Supplementary Material](#) online).

Introgression

We used three approaches to explore introgression ([table 1](#)). The first is based on *TreeMix*, an algorithm that uses SNP frequency data to evaluate population splitting, drift, and introgression ([Pickrell and Pritchard 2012](#)). Our DART data supported a model of one or two introgression events over alternatives with up to four events. Using *P. godmani* as the outgroup based on the phylogeny ([fig. 2c](#)) and *ad hoc* statistics in OptM ([Fitak 2021](#)), a model with one introgression event fit our data best, and accounted for 99.76% of the variance in the data ([supplementary fig. S3](#) and [table S4, Supplementary Material](#) online). This model indicates introgression from the lineage leading to *P. godmani* into *P. mareeba* ([fig. 4](#)). A model with a second introgression event from *P. assimilis* into *P. inornata* fit the data better ([fig. 4](#)), but the improvement was not statistically significant ([supplementary fig. S3](#) and [table S4, Supplementary Material](#) online).

In a second approach to introgression, we analyzed the DART data with the *D*-statistics (ABBA-BABA tests) using *Dsuite* ([Malinsky et al. 2021](#)) which allows for population sampling by using allele frequency data. We tested for introgression between all possible pairs of the three species in the

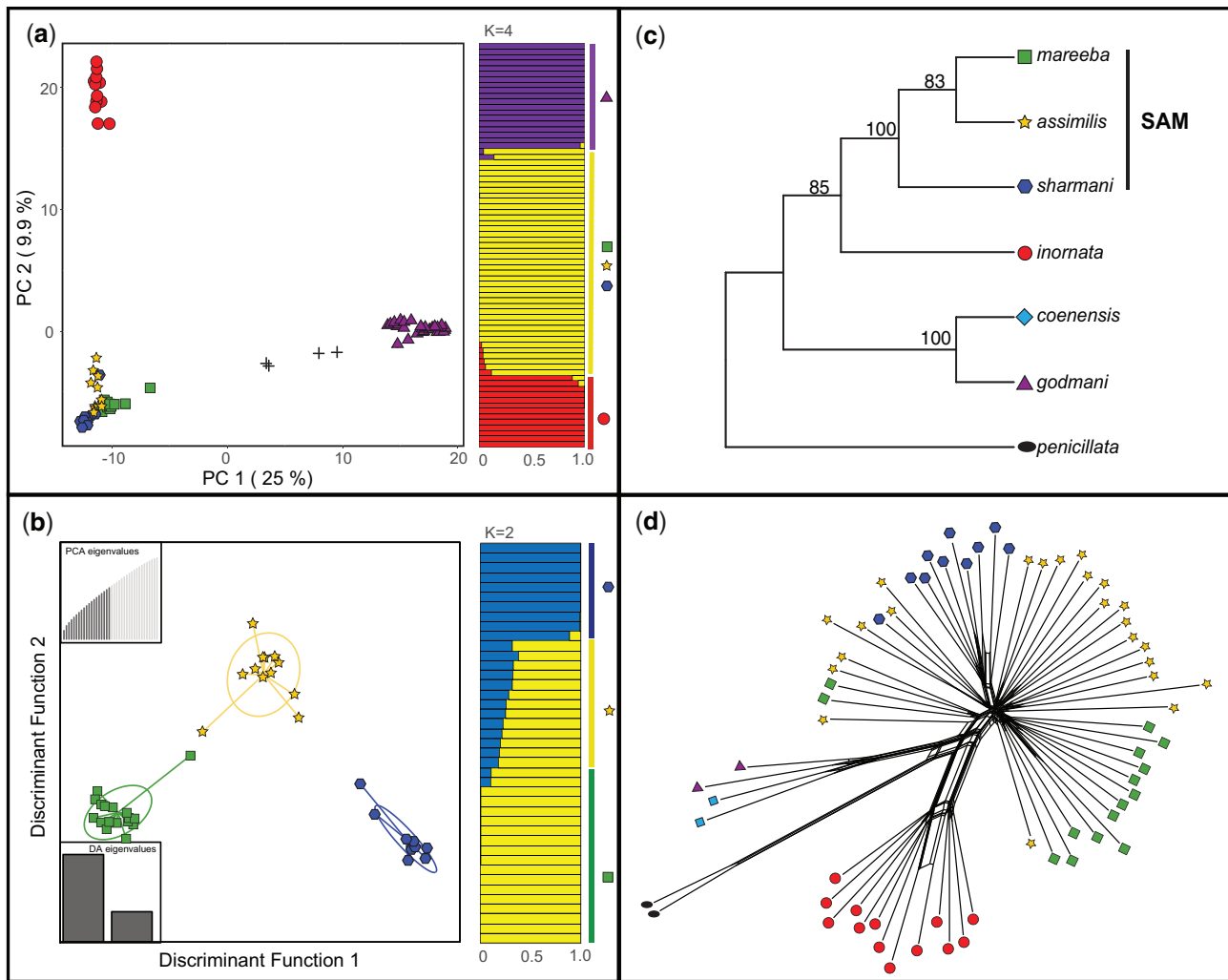


Fig. 2. Summary of population genomic (*a, b*) and phylogenomic (*c, d*) results based on the DArT and phased exon data sets, respectively. (*a*) PCoA results for the five *penicillata* complex taxa examined in this study (not including *P. penicillata* used as an outgroup). Species are colored individually and the symbols match species identification from (*c*). The PCoA (left) includes four known F1 hybrids (designated by +) between *P. godmani* and *P. mareeba*. STRUCTURE results for these same SNPs and individuals (right) are grouped into three clusters, with *P. assimilis*, *P. mareeba*, and *P. sharmani* (SAM) forming one group despite $K = 4$ clusters supported. (*b*) DAPC result for the SAM group based on 19 principal components (PCAs) and two discriminant functions (DAs), colored by species (left). STRUCTURE results for just the SAM group indicate two populations and are colored by cluster (right). The lines on the right represent the species boundaries. (*c*) Phylogenetic species tree of the *penicillata* complex based on the multispecies coalescent analysis in SVDquartets. Species colors and symbols match those in the population analyses (*a, b*) and the bootstrap support values are located on branches. (*d*) Network for individuals estimated using a distance-based approach in *SplitsTree* (Neighbor-Net). The regions of the network where multiple branches connect individuals and groups indicate reticulation.

SAM clade using *P. inornata* as the outgroup. We found significant evidence of introgression between *P. assimilis* and *P. sharmani* ($P < 0.05$, $Z > 3$). We also tested for introgression between parapatric *P. assimilis* and *P. inornata*, with *P. godmani* as the outgroup. Significant introgression was detected here as well ($P < 0.05$, $Z > 3$). We note that the D -statistic is five times greater in the second species pair than in the first (0.02 vs. 0.11; [supplementary table S5, Supplementary Material](#) online). Other species pairs also showed $P < 0.05$, suggestive of introgression. But in those cases, $Z < 3$, which we interpreted as insufficient to determine that introgression had occurred (see [Zheng and Janke \[2018\]](#)).

Our third approach to introgression used MIGRATE ([Beerli and Palczewski 2010](#), [Beerli et al. 2019](#)). This is a Bayesian

coalescent-based algorithm that estimates the mutation-scaled effective population size and introgression rates assuming a star phylogeny (see Materials and Methods). Here, we used the unphased exon data set as this program takes full account of ambiguity-coded sequence data. Estimated population sizes were very low for all species ($\Theta = 4N_e \mu = 3 \times 10^{-5}$), while introgression rates between some pairs of taxa were substantial ($M = m/\mu$ up to 2.5×10^4). The product of $\Theta/4$ and M estimates the absolute rate of introgression; here, the number of heterospecific genomes entering the population per generation. This rate was greatest for introgression from *P. assimilis* into *P. inornata* (0.19 genomes per generation), as well as from *P. assimilis* into *P. sharmani* and *P. mareeba* (0.12 and 0.13 genomes per generation,

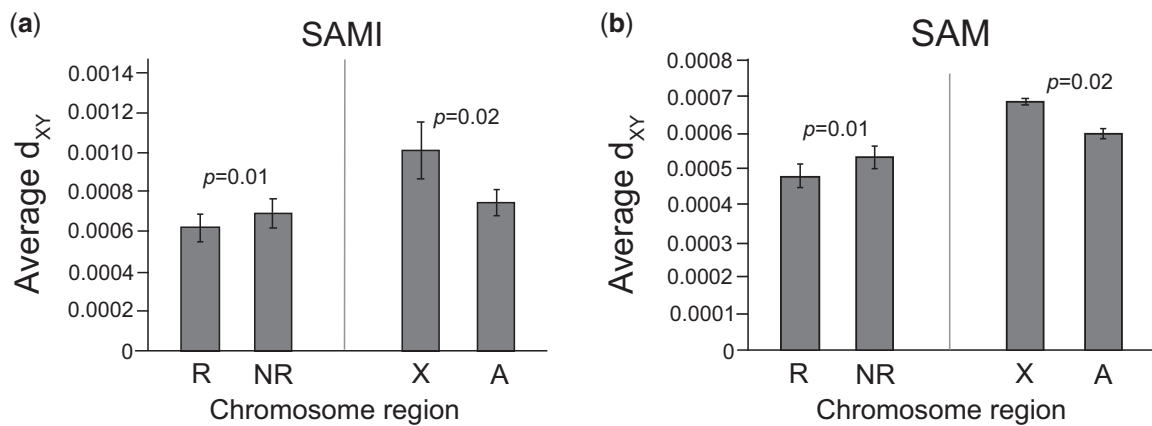


FIG. 3. (a) Average pairwise divergence (d_{XY}) between *P. inornata* and *P. sharmani*, *P. assimilis*, *P. mareeba* (SAM) species for comparisons between rearranged (R: 5,6,9,10) and nonrearranged (NR: 1,2,7,8) chromosomes, and between X and autosomal (A) loci. Tests for significant differences between the average d_{XY} were calculated using *t*-tests and significant values ($P < 0.05$) are marked with an asterisk (*). (b) Average d_{XY} within the SAM group only, for R versus NR and X versus A.

Table 1. Summary of Introgression Results from the Three Analyses: *TreeMix*, ABBA-BABA, and MIGRATE.

Pairwise Comparison (P1/P2)	Expected No. of Chromosomes in Meiotic Chain	<i>TreeMix</i>	ABBA-BABA			MIGRATE	
			D-Statistic	Z Score	P Value	P1 > P2	P2 > P1
<i>P. mareeba</i> / <i>P. godmani</i>	5	<i>P. godmani</i> > <i>P. mareeba</i>	—	—	—	—	—
<i>P. mareeba</i> / <i>P. assimilis</i>	5	—	0.0176	2.134	0.0164	0.0006	0.1346
<i>P. mareeba</i> / <i>P. sharmani</i>	3	—	0.0001	0.018	0.493	0.0162	0.0024
<i>P. assimilis</i> / <i>P. sharmani</i>	4	—	0.0234	3.446	0.0003	0.1188	0.0004
<i>P. assimilis</i> / <i>P. inornata</i>	3	<i>P. assimilis</i> > <i>P. inornata</i>	0.1146	15.292	0.0	0.1881	0.0004

NOTE.—Pairwise comparisons are listed as Parental 1 (P1) and Parental 2 (P2) taxa with MIGRATE results indicating directionality of migration between these taxa. Expected chains of arms in meiosis for chromosomal heterozygotes between the pairs of species are listed, with simple heterozygotes representing trivalent chain formations and complex heterozygotes representing chains of four or five. *TreeMix* results highlight significant introgression events and the direction of gene flow. ABBA-BABA results outline the *D*-statistic value, the Z score, and the *P* value. MIGRATE results summarize the absolute rate of introgression (mutation-scaled migration rate $M \times$ effective population size Θ /ploidy of data [$4N_e$]).

respectively) (supplementary table S6, Supplementary Material online).

We estimated introgression separately for each of the three genomic categories: X-linked, rearranged autosomes, and nonrearranged autosomes. For the X chromosome, we used all 45 loci, and a subsampled similar number of loci for the nonrearranged and rearranged autosomes ($n = 50$) to allow for similar power across the three categories. The per-generation introgression rates between all pairs of species were low and similar across the three categories ($M = 0.004$ – 0.12), again with highest values from *P. assimilis* into other species ($M = 0.05$ – 0.12 , supplementary table S6, Supplementary Material online). These results are not significantly different from zero and do not represent strong evidence of ongoing introgression (one migrant per generation).

In sum, there is consistent evidence of introgression between *P. assimilis* and *P. inornata* based on all three analyses. However, only the MIGRATE analysis estimates introgression in terms of genomes per generation and these were below the rates needed to homogenize neutral variation. Both MIGRATE and *TreeMix* results suggest introgression is unidirectional, from *P. assimilis* into *P. inornata*. The concordance of introgression results for this species pair agrees with our

expectation as these two species differ by only a single Robertsonian fusion. Evidence of introgression between other species pairs, however, is mixed. The ABBA-BABA and MIGRATE analyses suggest there has been some introgression between *P. assimilis* and *P. sharmani*, which have different fusions with monobrachial homology; however, this result was not supported by *TreeMix* results. Last, we detected introgression from an ancestor of *P. godmani* into *P. mareeba* using *TreeMix*, but data were not available to test this with the other two methods.

Discussion

Resolving the evolutionary histories of divergence and gene flow amongst very closely related species is important to determine effects of chromosome rearrangements on isolation, but is also expected to be challenging. Studies have shown different impacts of Robertsonian fusions on levels of introgression, largely associated with the complexity of chromosomal rearrangement (e.g., simple vs. complex heterozygotes; see Garagna et al. [2014]; Yannic et al. [2019]). Our previous work based on microsatellites inferred introgression even amongst species with complex rearrangements (Potter et al.

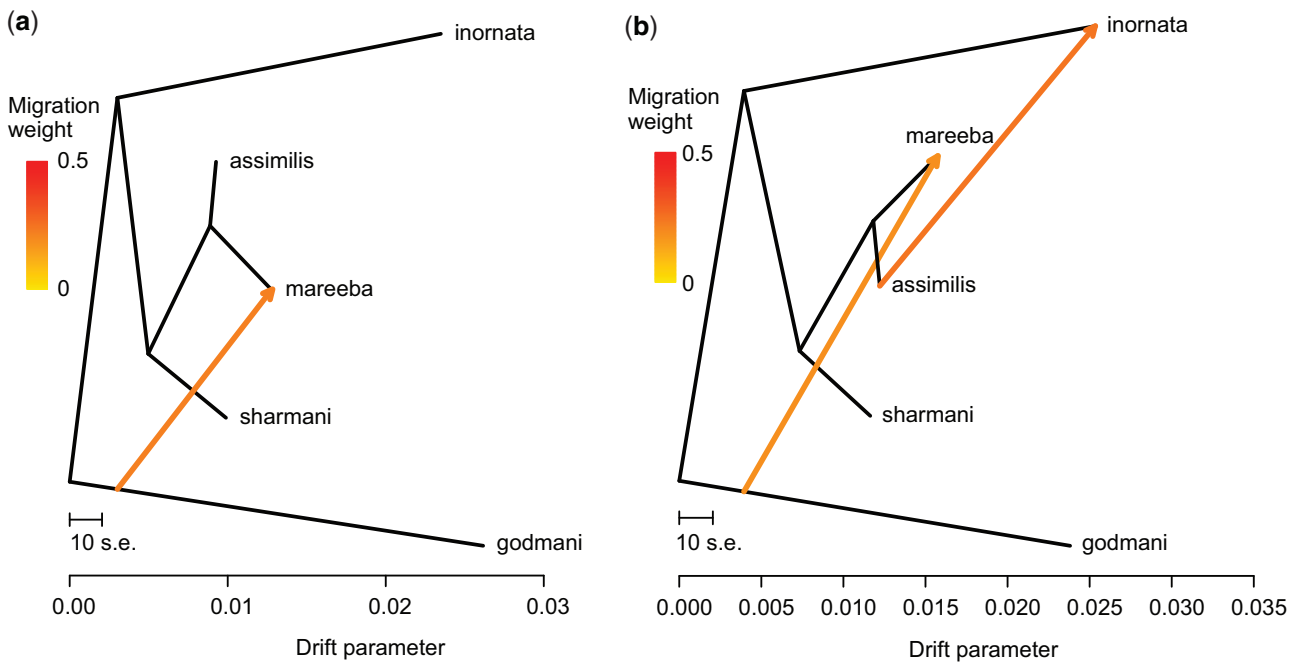


FIG. 4. (a) *TreeMix* results which support a single introgression event from the ancestor of *P. godmani* into *P. mareeba*. The migration rate is intermediate for this introgression event. (b) *TreeMix* results from a model of two introgression events, building off of the $m = 1$ model (a) and including an introgression event with high migration from *P. assimilis* into *P. inornata*.

2015). Here, however, the much larger genomic data sets provide new and more robust insights.

For the first time, we were able to demonstrate genetic differentiation between the three species in the SAM clade (*P. assimilis*, *P. mareeba*, and *P. sharmani*) and determined the divergence histories between all six species in the notoriously difficult *penicillata* group (fig. 2c). These results validate some previous conclusions based on just two individuals per species (Potter et al. 2017). In contrast to Potter et al. (2017), this current work finds moderate support for *P. assimilis* and *P. mareeba* as sister species. We do, however, also find evidence of reticulate evolutionary processes (incomplete lineage sorting or introgression) between species (fig. 2d and table 1). The new phylogeny suggests that the history of structural changes is more complex than was previously appreciated. For example, five to ten fusions have fixed independently in the parapatric *P. mareeba* and *P. sharmani* or there has been a reversal in *P. assimilis* (fig. 1). Given the low rates of introgression that we found, it is likely that incomplete lineage sorting rather than introgression is driving most of the uncertainty in phylogenetic relationships.

Introgression between Species

Across the *penicillata* group, strong evidence for introgression is restricted to just some species pairs (table 1). Where it has occurred, it primarily involves the geographically widespread and centrally located *P. assimilis*. The clearest evidence for introgression is between *P. assimilis* and *P. inornata*, which differ by a single fusion between chromosomes 6 and 10, and centric shifts on chromosomes 3 and 4. This supports our prediction that introgression should be less impeded between species whose hybrids form simple heterozygotes (trivalents).

The finding is also consistent with introgression seen in other taxa where hybrids show proper segregation of chromosomes during meiosis (e.g., *Mus musculus domesticus*, *Sorex araneus*; see Garagna et al. [2014]; Searle et al. [2019], respectively). Introgression between *P. assimilis* and *P. inornata* has previously been reported from mtDNA and microsatellite loci (Potter et al. 2015), as well as at their contact zone based on allozymes and mtDNA (Briscoe et al. 1982; Bee and Close 1993).

Our results also show some evidence of introgression between *P. godmani* and *P. mareeba*, based on *TreeMix* and PCoA results (fig. 2; table 1). The PCoA demonstrates the mixed ancestry of F1 hybrids midway between these two species, consistent with previous reports of mtDNA leakage and known karyotypic hybrids (Briscoe et al. 1982; Sharman et al. 1989; Bee and Close 1993; Close and Bell 1997; Potter et al. 2015). Evidence of introgression between these two species was also found from microsatellite markers (Potter et al. 2015). These species have complex chromosomal differences: *P. godmani* has fusions between chromosomes 6 and 10, and an inversion on chromosome 5; *P. mareeba* has fusions between chromosomes 6 and 9, and between chromosomes 5 and 10. In hybrid meiosis, these form a chain of five chromosomes. The results from these two species contradict our prediction that complex rearrangements should inhibit introgression. Similarly, complexity of rearrangements has not seemed to effect gene flow between races of the common shrew (*S. araneus*) (Horn et al. 2012). Empirical systems where more fine-scale mapping of loci to rearranged and nonrearranged loci are suggesting introgression, when present, is localized to nonrearranged regions of the genome (e.g., Giménez et al. 2013). Unfortunately, we do not know

which of the SNPs used in our analyses lie within rearranged parts of the genome, so we cannot test the hypothesis that there will be lower introgression in those regions.

We find limited evidence of introgression between members of the SAM clade. The lack of substantial introgression is not surprising given that they differ in multiple structural rearrangements. Meiosis in an F1 hybrid between *P. assimilis* and *P. mareeba* would involve a chain of five chromosomes, whereas a hybrid between *P. assimilis* and *P. sharmani* hybrid would generate a chain of four chromosomes. For this last pair, there is indication of introgression from ABBA-BABA tests on SNPs and MIGRATE analyses of exons, but that conclusion is not supported by the *TreeMix* analysis of SNPs (table 1). Previous research has suggested limited introgression at the contact zone between *P. assimilis* and *P. sharmani* based on mtDNA, but incomplete lineage sorting could not be ruled out (Bee and Close 1993). Introgression between these species has been found based on microsatellite loci (Potter et al. 2015). Conversely, *P. mareeba* and *P. sharmani* show no evidence of introgression even though they differ by only a single fusion. However, this fusion of chromosomes 6 and 9 also involves a subsequent centric shift resulting in an acrocentric chromosome (Eldridge et al. 1989; Metcalfe et al. 1997), which could influence crossover of segments of chromosome 9 and be more complex than a simple Robertsonian fusion.

Higher levels of introgression among these taxa were previously inferred using coalescent analyses of microsatellite loci (Potter et al. 2015) and that result could be due in part to homoplasy in the microsatellite data (Balloux et al. 2000; Queney et al. 2001). Identical allele sizes could be present due to different mutational histories (convergent evolution), not shared ancestry, and have the potential to increase estimates of introgression and reduce estimates of divergence. A study in shrews (*S. araneus*) found extreme underrepresentation of population structure across a chromosomal hybrid zone and cautioned that studies based on autosomal microsatellite loci could overestimate the extent of gene flow (Balloux et al. 2000). We therefore caution the interpretation of gene flow based solely on microsatellites. Introgression will be better estimated using SNP or sequence data, particularly for empirical systems where gene exchange could be limited as a result of chromosomal rearrangements. Our findings here for the SAM group suggest if introgression is present, it could be very localized in the genome so that whole genome sequencing will be required to obtain robust and detailed results.

Variation in Introgression and Divergence

We find that the X chromosome shows greater divergence (measured as d_{XY}) than do the autosomes in comparisons between *P. assimilis*, *P. inornata*, *P. mareeba*, and *P. sharmani*. Similar results have been reported from many other taxa (faster X effect), for which several hypotheses have been proposed (Presgraves 2018). In many species, the X chromosome is also enriched for genes involved in hybrid incompatibilities (Coyne and Orr 2004), which can inhibit introgression of the X across species boundaries. This outcome has been seen, for

example, in *Anopheles* mosquitoes (Fontaine et al. 2015; Thawornwattana et al. 2018). Surprisingly, in our system, loci mapped to chromosomes involved in rearrangements (5, 6, 9, 10) show significantly less divergence than do loci mapped to nonrearranged chromosomes (1, 2, 7, 8), opposite to expectations if restriction of gene flow is greater for fused chromosomes (e.g., via localized suppression of recombination). The number of loci we could accurately assign to chromosomes was limited though. In addition, cytogenetic banding results may have underestimated rearrangements in putatively nonrearranged chromosomes (e.g., Eldridge and Close 1993). De novo genome assemblies which can resolve spatial organization (e.g., Hi-C approaches—Belton et al. 2012; Kong and Zhang 2019), have been shown to identify novel rearrangements (Fan et al. 2019, Damas et al. 2021). Further whole genome sequencing with de novo approaches will improve our understanding of rearrangements and whether there is localized introgression (Mérot et al. 2020). This will enable more robust testing of hypotheses surrounding the faster X effect and divergence histories in relation to rearrangements.

The introgression detected between *P. assimilis* and *P. inornata*, between which male hybrids are sterile, could suggest a role for genic divergence rather than chromosome rearrangements driving speciation, or a combination of the two in play. In both house mice (*Mus musculus domesticus*) and common shrews (*S. araneus*) simple Robertsonian fusion heterozygotes (trivalents) generally progress normally through meiosis (Garagna et al. 2014; Borodin et al. 2019). We find conflicting evidence for the prediction that introgression should be reduced between species that differ in larger numbers of rearrangements. We see some suggestion of introgression between two species pairs that differ by multiple rearrangements (*P. godmani* and *P. mareeba*, *P. assimilis* and *P. sharmani*, respectively). It is possible that introgression predated the chromosomal rearrangements, or occurred during the early stages of divergence when only a single or more simple rearrangements separated these species. However, given the evidently low level of introgression among species within the more recently evolved SAM clade, it is likely that complex combinations of chromosome rearrangements do in fact play an important (if not sole) role in divergence and speciation in this genus.

Unlike some of the better studied systems (e.g., *Mus* and *Sorex*), where data can be collected across hybrid zone transects, the disjunct populations of *Petrogale* make continuous sampling approaches unfeasible. We note that the highly fragmented population structure and small effective population size of rock-wallabies are mixed fortunes for studies of chromosomal speciation. On the one hand, those features generate strong drift and promote the fixation of potentially underdominant rearrangements. On the other hand, they result in extremely low levels of molecular polymorphism, which makes tests of hypotheses challenging. The data here are consistent with the recombination suppression model, as introgression may be ongoing outside of rearranged regions. The data are also compatible with the hybrid dysfunction model, as introgression could have occurred before the

chromosome rearrangements became fixed. Last, it is possible that the rearrangements are not important to genetic isolation, and instead mediated by genic divergence. Even in the better studied systems it is still difficult to distinguish between these hypotheses (Giménez et al. 2013; Garagna et al. 2014).

Conclusions

Individual Robertsonian fusions are thought to cause little disruption of meiosis and can fix via meiotic drive or genetic drift (Sites and Moritz 1987; Chmátal et al. 2014; Garagna et al. 2014). When species differ in multiple fusions that share chromosome arms, however, more severe meiotic disruption is expected (Baker and Bickham 1986). Due to their diverse karyotypes, rock-wallabies have long been identified as an interesting model group to explore chromosomal speciation (King 1993; Brown and O'Neill 2010).

Our results suggest that fusions have played a significant role in reducing genome-wide introgression within *Petrogale*. Consistent evidence of introgression is present between two species which share a single Robertsonian fusion. Introgression was supported for one pair of species which share complex chromosomal rearrangements. For the remaining species, there is limited evidence of introgression, and it appears to be at low levels indicative of infrequent events. Like the house mouse, speciation in *Petrogale* could reflect both hybrid dysfunction and recombination suppression associated with chromosome change, and these in concert with genic divergence, especially on the X chromosome. Comparisons and integration with evidence from model systems of *Mus* and *Sorex*, where detailed insight come from hybrid zone research and laboratory crosses (e.g., Garagna et al. 2014; Searle et al. 2019), will be important in drawing insight about the role of chromosomal evolution in other diverse systems, including *Petrogale*. Further, a much more nuanced test of the roles played by structural and genic variation in the evolution of reproductive isolation will become possible with the arrival of whole genome sequences of this and other taxa with rapidly evolving chromosomes.

Materials and Methods

Data Sets

Three different data sets were analyzed for this study with individuals overlapping where possible. The first data set consisted of SNPs generated from a complexity reduction approach ($n = 77$), the second data set consisted of full-length phased exon sequences based on an exon capture approach ($n = 67$) and the third data set consisted of unphased exon sequences from this same exon capture approach (supplementary table S1, Supplementary Material online; see below for methods).

Sampling and DNA Extraction

DNA was extracted from ear and liver biopsies stored at the Australian Museum using a salting-out method (Sunnucks and Hales 1996). A total of 88 samples, identified to species by karyotyping (Sharman et al. 1989; Bee and Close 1993), were analyzed across five species including *Petrogale assimilis*

($n = 23$), *P. godmani* ($n = 19$), *P. inornata* ($n = 14$), *P. mareeba* ($n = 17$), *P. sharmani* ($n = 11$). Sampling was geographically widespread across the known distribution of each species (fig. 1). Based on previous mtDNA sequencing (Potter et al. 2015), we subsampled individuals from across the distribution to encompass the mitochondrial diversity as well as geographic spread within each species. In addition, individuals from *P. coenensis* ($n = 2$) and *P. penicillata* ($n = 2$) were included as outgroups and to complete the exon capture data set.

Exon Capture Approach

We applied a custom in-solution exon capture approach using target sequences from a yellow-footed rock-wallaby (*Petrogale xanthopus*) transcriptome (Bragg et al. 2016; see Potter et al. [2017]). Briefly, we identified orthologs of exon sequences in the transcriptome based on reciprocal best BLAST (blastall v 2.2.25, program BlastN; Altschul et al. 1990) hits to the Tasmanian devil (*Sarcophilus harrisii*) genome (derived from GTF file, Ensembl release 74, Flicek et al. 2014). A total of 3,960 target exons were filtered as being >200 bp in length and based on BLAST hits to *Sarcophilus harrisii* and the tammar wallaby genome (*N. eugenii*) (O'Neill RJ, unpublished data) expected to be single copy. Probes were then synthesized for these targets (1.83 Mb) in a SeqCap EZ Developer Library (Roche NimbleGen; scripts to identify targets are available in Dryad Repository <https://doi.org/10.3389/fgene.2017.00010>).

Genomic libraries for each individual were prepared following the protocol of Meyer and Kircher (2010), including modifications made by Bi et al. (2013). In addition, we included a double-sided size selection bead clean up prior to the blunt-end repair step to remove fragments <200 and >500 bp. Each individual sample was given a unique barcode (supplementary table S1, Supplementary Material online) and quantified using a LabChip DS Droplet Spectrophotometer (PerkinElmer). For the in-solution hybridization, we pooled samples in equimolar ratios (1.2 μ g total). Samples were prepared across multiple experiments, limited by the number of individuals pooled together for a single hybridization reaction ($n = 56$). In addition to the genomic libraries, 5 μ g of mouse Cot-1 DNA (Life Technologies Corporation) and 56 barcode specific blocking oligos (1,000 pmol) designed to block the unique barcodes and adapters used in the Meyer and Kircher (2010) protocol were combined. Following the manufacturer's protocol (SeqCap EZ Developer Library), samples were hybridized for ~ 72 h and the hybridization reaction then cleaned up. Two independent enrichment PCRs (17 cycles) were run on the cleaned "postcapture" hybridization reaction and checked against the precapture pooled libraries as a quality control check, to assess global enrichment efficiency using the DyNAmo Flash SYBR green qPCR kit (Thermo Fisher Scientific Inc.; see Bi et al. [2012]). The qPCR amplified target loci from the capture using primers specifically designed to hit targets of the hybridization probes to test for enrichment, as well as primers to amplify non-targets that are known to amplify but not in the capture to test de-enrichment. The enriched hybridization samples were

run on a BioAnalyzer (2100; Agilent Technologies, Inc.) to check the quality and quantity of the sample and then sequenced on a single lane of an Illumina HiSeq 2500 (100 bp paired-end run) at the ACRF Biomolecular Resource Facility.

SNP Screening

Genomic DNA samples (20 μ l of 50–100 ng/ μ l concentration) were sent to Diversity Arrays Technologies Pty Ltd (DART). SNP genotyping was performed using the DART complexity reduction approach and proprietary DART analytical pipelines (Kilian et al. 2012). This approach uses restriction enzymes, double digest complexity reduction, fragment size selection, and sequencing. Technical replicates were included by DART to assess allele calls and accuracy. We requested raw sequencing reads and mapped these to a *P. penicillata* genome to call SNPs (see below), rather than using genotypes produced by DART pipelines.

Bioinformatics—Exon Capture Data

Illumina sequencing reads were cleaned following a workflow developed by Singhal (2013). Briefly, this workflow removes duplicate, contaminant (human: GRCh37, ensemble version 67; *Escherichia coli*: str. K-12 substr. MG1655, GenBank accession number U00096.2), and low complexity reads. Then, TRIMMOMATIC (v 0.22; Bolger et al. 2014) is used to remove adaptors and low-quality bases, and FLASH (v 1.2.2; Magoc and Salzberg 2011) is used to merge overlapping read pairs. All scripts used in this analysis are archived in Dryad Repository (<https://doi.org/10.5061/dryad.mm856>).

Each locus was assembled de novo from the cleaned sequencing reads, with heterozygous sites called and phased using a workflow described in Bragg et al. (2016) (see Dryad Repository <https://doi.org/10.5061/dryad.mm856> for code that was used to assemble exon sequences). Using BLAST (blastall v 2.2.25, program BlastX; Altschul et al. 1990), sequencing reads with homology to each target exon were identified. Then, using VELVET (K values of 31, 41, 51, 61, and 71; v 1.2.08; Zerbino and Birney 2008), sequencing reads homologous to the target were assembled. CAP3 (Huang and Madan 1999) was used to assemble contigs from different K values. These were then aligned to the target *Sarcophilus* protein using EXONERATE (v. 2.2.0; Slater and Birney 2005), and the contigs were trimmed to the target exon boundaries. Where multiple contigs were present, we took the one with the maximum bit score for the BlastX hit and checked it did not hit any other *Sarcophilus* proteins. Using BOWTIE 2 (v 2.2.2; Langmead and Salzberg 2012) and GATK (v. 3.3-0-g37228af; McKenna et al. 2010), heterozygous sites were identified by mapping clean sequencing reads to the best assembled contigs and then phased. We inserted unknown (N) bases at sites that did not exceed a minimum genotyping quality (GQ) score of 20. This pipeline has alternative fully automated containerized version available from the Docker hub: <https://hub.docker.com/repository/docker/trust1/ubuntu>.

Finally, haplotype sequences obtained from the steps above were aligned and filtered using the EAPhy (v1.2; Blom 2015) pipeline. This uses MUSCLE (v 3.8.31; Edgar 2004) to align haplotype sequences, perform checks of coding

frame to ensure amino acid coding (starting in frame 1), removes missing data from the ends of alignments and creates data files ready for analysis (fasta and phylip format) allowing for different amounts of missing data (we set 0–10%). Samples were visually inspected using Geneious Prime 2020.1.1 (<https://www.geneious.com>). We used snp-sites (Page et al. 2016) to extract SNP data from FASTA alignments of exon capture loci.

We located our target sequences on the scaffolds of the *N. eugenii* genome assembly using BLAST (blastall v 2.2.25, program BlastN; Altschul et al. 1990). Based on knowledge of scaffold physical maps, and relationship of tammar wallaby chromosome to homologous *Petrogale* chromosomes (Eldridge and Close 1997; O'Neill et al. 1999), we were able to group loci into chromosomes. Average pairwise divergences between species across the X, rearranged autosomal (chromosomes 5, 6, 9, 10) and nonrearranged autosomal loci (chromosomes 1, 2, 7, 8) were estimated using PopGenome (Pfeifer et al. 2014). Averages were then estimated across all groups and divided by the total length of the alignment of the X, autosomal loci, as well as rearranged autosomal and nonrearranged autosomal loci to get comparisons for average net divergence.

Mapped Loci

We estimated average net divergence (d_{XY}) between species of loci mapped to chromosomes on *N. eugenii*. Based on our knowledge of chromosome synteny (Deakin and O'Neill 2020), we were able to estimate the location of these loci on the *penicillata* group species' chromosomes. Loci were separated into three groups: X chromosome, rearranged chromosomes (including chromosomes 5, 6, 9, and 10), and nonrearranged chromosomes (including chromosomes 1, 2, 7, and 8).

Bioinformatics—DART Data

The DART sequencing reads were mapped to a new draft *Petrogale penicillata* genome sequence (<https://doi.org/10.5061/dryad.6m905qg0d>) for variant calling.

To generate the draft genome sequence, we extracted high molecular weight DNA using a Qiagen genomic tips and DNA buffer kit (Qiagen, Hilden, Germany). Then a 10 \times Genomics linked read sequencing library (10 \times Genomics, Pleasanton, CA) was prepared and sequenced using an Illumina HiSeq X Ten instrument (The Ramaciotti Centre for Gene Analysis, UNSW). A draft genome was assembled using the Supernova assembler (“pseudohap” format; v 1.2.0, Weisenfeld et al. 2017). We assessed the completeness of the draft genome using BUSCO (v 2.0.1, Simão et al. 2015, tetrapoda_odb9 data set), which measures the fraction of expected single-copy ortholog genes that were present in the assembly. BUSCO was configured with HMMer (v 3.1b2, Eddy 2011), AUGUSTUS (v 3.2.2, Keller et al. 2011), EMBOSS (v 6.5.7, Rice et al. 2000), and BLAST+ (v 2.2.31, Altschul et al. 1990). The draft genome assembly for *Petrogale penicillata* had a scaffold N50 of 5.5 Mb, L50 of 179, and contained 89.7% of expected single-copy ortholog genes ($n = 3,950$) rated as Complete with a low level of duplication (50 of the

3,543 Complete BUSCOs). A further 5.4% of BUSCO genes were fragmented, with only 4.3% missing.

The raw DArT Illumina single-end short reads were mapped to the draft genome sequence using “bwa mem” (v. 0.7.17, [Li and Durbin 2009](#)). Variant calling followed by GATK best practice protocol with GATK (v. 4.1.4.1, [Van der Auwera et al. 2014](#)). Briefly, this protocol first ran single sample variant calling which generates a panel of polymorphic sites detected across all samples. Then, a second stage processing step where sequence base qualities in the alignment files were recalibrated with information taken from the first stage variant calling. Recalibrated alignment files proceeded with second pass calling via making combined g.vcf file across all samples including technical replicates. This protocol allows us to accurately detect low frequency allelic variants. Default parameters and thresholds suggested by GATK best practice recommendations were applied. GATK pipeline script and all used tools were containerized with Docker image and available from the `trust1/gatk` repository on the official Docker website alone with additional documentation: <https://hub.docker.com/r/trust1/gatk>.

The DArT SNPs produced by the GATK pipeline were filtered prior to analysis. We used `vcftools` to apply an initial set of SNP filters including minimum coverage (8), maximum coverage (150), and genotyping quality (20). After applying these proposed filters, we compared the genotypes that were estimated for each pair of technical replicate samples. We found that, on average, technical replicates had the same estimated genotype at 99.73% of called loci. We were satisfied that this level of genotyping precision was likely to produce robust downstream analyses. We therefore removed technical replicates from the data set (keeping the replicate sample with the most complete data) and applied the same set of filters. We also applied several additional filters. This included the removal of SNPs that had missing data for more than 30 samples.

Genetic Structure

We first evaluated genetic clustering of individuals using a PCoA in the package `dartR` ([Gruber et al. 2018](#)) in R (v. 3.6.3; R Core Team). This clustering approach uses Euclidean genetic distances which we applied to our SNP data generated from DArT. We applied an hierarchical approach, whereby we removed the most divergent species and reran the analysis to determine clustering of more closely related species. We also ran a DAPC using the package `adegenet` ([Jombart 2008; Jombart et al. 2010](#)) and `poppr` ([Kamvar et al. 2014](#)) packages in R on the most closely related SAM group. We used the species identity to inform this analysis. This multivariate approach identifies clusters of genetically related individuals and finds a model which maximizes the between-group variation whilst minimizes the within-group variation, thus can make use of the a priori species identity.

In addition, we explored genetic structure of the DArT SNP data using a Bayesian clustering algorithm implemented in `STRUCTURE` ([Pritchard et al. 2000](#)), which uses Hardy–Weinberg equilibrium to estimate individual ancestry coefficients and determine distinct genetic groupings. Again, we used a

hierarchical approach, where we pulled out distinct populations/species and reran the program with fewer individuals to detect more fine-scale structure. This approach is suggested to counteract the $K = 2$ conundrum ([Janes et al. 2017](#)). These analyses were performed using the admixture model, independent allele frequencies and lambda set to 1.0. Genetic clusters (K) were estimated across 10 replicates, using a burn-in of 100,000 and sampling every 1 million iterations. We then ran `CLUMPAK` ([Kopelman et al. 2015](#)) to infer the best K value based on evaluation of the maximum posterior probability ($L[K]$; [Pritchard et al. 2000](#)) and the maximum delta log likelihood (ΔK ; [Evanno et al. 2005](#)).

Phylogenetic Analyses

Given our population-scale sampling with a large number of loci, use of full Bayesian species-tree or network methods is infeasible. We therefore estimated phylogenetic relationships using `SVDquartets` in `PAUP*` ([Swofford 2003; Chifman and Kubatko 2014](#)), which analyses a decomposition matrix of site pattern frequencies to compute quartet scores and infer a species phylogeny. We analyzed the concatenated phased exons (h0 haplotype only) from the exon capture data set to estimate the phylogeny using two approaches. One assigning tips to species and running the multispecies coalescent and the other not assigning tips to species. Both analyses evaluated 100,000 random quartets and performed 1,000 bootstrap replicates.

Tests for Introgression

We implemented a phylogenetic network analysis using the Neighbor-Net approach in `SplitsTree4` ([Bryant and Moulton 2002; Huson and Bryant 2006](#)) to examine the patterns of reticulation. The Neighbor-Net approach creates a distance matrix based on uncorrected P distances and agglomerates clusters of individuals and is a useful exploratory statistical approach to examine complex evolutionary processes of reticulation. Again, we used the concatenated nuclear alignment (h0 haplotypes), effectively ignoring coalescent variance in gene trees to visualize potential reticulation.

We then used a population genomic approach to evaluate the phylogenetic relationships of species and explore admixture between species whilst accounting for incomplete lineage sorting. Using the DArT SNP data set, we ran `TreeMix` ([Pickrell and Pritchard 2012](#)), which uses allele frequency data to evaluate historical relationships, estimating patterns of population splitting, drift, and mixing. We calculated allele frequencies for each species and estimated a maximum likelihood tree using *P. godmani* as the outgroup and different inferences of admixture events were compared (zero to four events). We analyzed the data in windows of size $k = 1$, $k = 100$, and $k = 500$ to account for any nonindependence in SNPs. The residual plots of observed covariance, the percent of variance explained by the models (`get_f()` script in R from documentation accompanying `TreeMix`) and the package `OptM` in R ([Fitak 2021](#)), were examined to determine the best model fit to the data and infer the number of admixture events. `OptM` calculates the second-order rate of change in the log likelihood between the different models of

introgression and infers delta M to estimate the most likely model. Based on the residual plots of covariance, positive residuals indicate the model underestimates the observed covariance between pairs of populations, whilst negative residuals imply overestimation of the model. Therefore, residuals close to zero infer the best fit model to the data. Once a model infers 99.8% of the variance in the data, it has been customary to stop adding migration events. Given the varied approaches to determining the best fit model to the data, we evaluated across them to estimate the number of migration events using *TreeMix*. In addition, to formally test for admixture, we used the three-population test (f_3 statistic; *threepop* analysis) within *TreeMix*, where a negative statistic infers admixture.

In addition to *TreeMix*, we calculated Patterson's D (Green et al. 2010; Durand et al. 2011) also known as the ABBA-BABA statistic as well as the f_4 -ratio (f_{4G} ; Patterson et al. 2012) for all possible trios of species in *Dsuite* (Malinsky et al. 2021). Based on biallelic sites and the phylogeny (((P1, P2),P3),O), where the outgroup (O) defines the ancestral allele, and the pattern of the derived allele determines if there is introgression between P3 and P1 or P2. This site pattern approach assumes under a null hypothesis that there are equal patterns of shared derived alleles between P3 and P1 and between P3 and P2, and that deviation from this (greater than sampling error) implies introgression. We used this approach because it is computed for allele frequency estimates and thus, we could include multiple individuals from each species. We tested for introgression among parapatric species using the DArT data set. To test for introgression among SAM, we used *P. inornata* as the outgroup, and for evaluation of introgression between *P. assimilis* and *P. inornata*, we used *P. godmani* as the outgroup. Statistics were calculated for all possible trios of species using the *Dtrio* command and jackknife blocks of 20 and 100 ($k = 20$ or $k = 100$) SNPs. We also ran the analysis using user-defined input trees (-t option) to ensure all pairwise comparisons were conducted. To assess if D is significantly different from zero, the standard block-jackknife procedure was used to obtain an approximately normally distributed standard error (Green et al. 2010; Durand et al. 2011; Malinsky et al. 2021). The Z score, the value of the D -statistic divided by its standard error represents strong significance if Z score >3 (Zheng and Janke 2018). A combined result with P values (<0.05) and Z scores >3 was taken as significant evidence of introgression. We evaluated the D_{min} , the lowest D -statistic reported for each trio, as a conservative approach. P1 and P2 are ordered so that $n_{ABBA} \geq n_{BABA}$ and thus is not negative, so the program does make some assumptions of relatedness of species based on the shared allele patterns. If P1 is more closely related to P3 (more shared alleles) then it swaps to P2 and P2 becomes P1. Therefore, the results range from 0 = no introgression between P2 and P3, to 1 = no incomplete lineage sorting but introgression.

Lastly, to exploit having extensive haplotype sequence data from exon-capture, we used a Bayesian coalescent-based approach to estimate effective population sizes (Θ) and past migration rates (M) in MIGRATE 3.6 (Beerli and Palczewski 2010). Using the exon capture data (*P. godmani* excluded due

to small sample size), we estimated Θ for all species and M for an asymmetric migration matrix. Past migration was estimated for all pairs of species except for the geographically nonadjacent. These estimates are mutation-scaled estimates, such that $\Theta = 4 N_e \mu$ and $M = m/\mu$, where μ is the mutation rate per generation and x is a multiplier that depends on the ploidy and inheritance of the data (e.g., $x = 4$ for nuclear data). The analysis was run using 1,617 nuclear loci, the Jukes–Cantor sequence model with base frequency 0.25, a uniform prior distribution for Θ (0,0.1,0.01; minimum, maximum, Delta respectively) and M (0, 50,000, 10,000). Random starting parameters from the prior distribution were used for estimation of Θ and M , and analyses were run starting from a random number seed, a constant mutation rate estimated from the data and two independent analyses were run, each with one long chain and four heated chains. A static heating scheme was used with temperatures set to 1, 1.5, 3, and 10^6 ordered from cold to hot, with sampling every 100 steps and run for 1 million steps after a burnin of 50,000.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by an Australian Research Council Discovery Grant (DP160100187), Strategic Research Funds from the University of Canberra as well as Australian Museum and National Institutes of Health Grant No. R01 GM116835 to M.K. We would like to thank the Australian Museum and South Australian Museum for samples. In addition, we would like to thank all the individuals that contributed tissues that were used in this study. We would like to thank Rachel O'Neill for access to the tammar wallaby genome and Peter Beerli for advice on MIGRATE analyses. We thank Maxine Piggott for assistance with laboratory work.

Data Availability

The scripts and data underlying this article are available in Dryad Digital Repository at <https://doi.org/10.5061/dryad.mm856>, <https://doi.org/10.3389/fgene.2017.00010>, as well as <https://doi.org/10.5061/dryad.6m905qg0d>. In addition, sequence data underlying this article are available in the Short Read Archive on the NCBI GenBank Nucleotide Sequence Database in SRA BioProjects PRJNA360868 and PRJNA741610 (refer to [supplementary table S1](#), [Supplementary Material](#) online, for Biosample IDs). The draft pseudohaplotype genome of *Petrogale penicillata* is available from Dryad Digital Repository <https://doi.org/10.5061/dryad.6m905qg0d>.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Ayala JF, Coluzzi M. 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci U S A.* 102(Suppl 1):6535–6542.
- Baker RJ, Bickham JW. 1986. Speciation by monobrachial centric fusions. *Proc Natl Acad Sci U S A.* 83(21):8245–8248.

- Balloux F, Brunner H, Lugo-Moulin N, Hausser J, Goudet J. 2000. Microsatellites can be misleading: an empirical and simulation study. *Evolution* 54(4):1414–1422.
- Basset P, Yannic G, Hausser J. 2019. Is it really the chromosomes? In: Searle JB, Polly PD, Zima J, editors. *Shrews, chromosomes and speciation*. Cambridge: Cambridge University Press. p. 365–383.
- Bateson W. 1909. Heredity and variation in modern lights. In: Seward AC, editor. *Darwin and modern science*. Cambridge: Cambridge University Press. p. 85–101.
- Baverstock PR, Gelder M, Jahnke A. 1983. Chromosome evolution in Australian *Rattus*—G-banding and hybrid meiosis. *Genetica* 60(2):93–103.
- Bee CA, Close RL. 1993. Mitochondrial DNA analysis of introgression between adjacent taxa of rock-wallabies, *Petrogale* species (Marsupialia: Macropodidae). *Genet Res.* 61(1):21–37.
- Beerli P, Mashayekhi S, Sadeghi M, Khodaei M, Shaw K. 2019. Population genetic inference with MIGRATE. *Curr Protoc Bioinformatics.* 68(1):e87.
- Beerli P, Palczewski M. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185(1):313–326.
- Belton J, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58(3):268–276.
- Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Mol Ecol.* 22(24):6018–6032.
- Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* 13:403.
- Blom M. 2015. EAPhy: a flexible tool for high-throughput quality filtering of exon-alignments and data processing for phylogenetic methods. *PLoS Curr.* 7. doi: 10.1371/currents.tol.75134257bd389c04bc1d26d42aa9089f.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Borodin PM, Fedyk S, Chêtnicki W, Torgasheva AA, Pavlova SV, Searle JB. 2019. Meiosis and fertility associated with chromosomal heterozygosity. In: Searle JB, Polly PD, Zima J, editors. *Shrews, chromosomes and speciation*. Cambridge: Cambridge University Press. p. 217–270.
- Bragg JG, Potter S, Bi K, Moritz C. 2016. Exon capture phylogenomics: efficacy across scales of divergence. *Mol Ecol Resour.* 16(5):1059–1068.
- Briscoe DA, Calaby JM, Close RL, Maynes GM, Murtagh CE, Sharman GB. 1982. Isolation, introgression and genetic variation in rock wallabies. In: Groves RH, Ride WDL, editors. *Species at risk: research in Australia*. Canberra (ACT): Australian Academy of Science. p. 73–87.
- Britton-Davidian J, Caminade P, Davidian E, Pagès M. 2017. Does chromosomal change restrict gene flow between house mouse populations (*Mus musculus domesticus*)? Evidence from microsatellite polymorphisms. *Biol J Linn Soc.* 122(1):224–240.
- Brown JD, O'Neill RJ. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet.* 11:291–316.
- Bryant D, Moulton V. 2002. NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In: Guigó R, Gusfield D, editors. *International workshop on algorithms in bioinformatics*. Heidelberg (Germany): Springer. p. 375–391.
- Campbell CR, Poelstra JW, Yoder AD. 2018. What is speciation genomics? The role of ecology, gene flow, and genomic architecture in the formation of species. *Biol J Linn Soc.* 124(4):561–583.
- Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513(7517):195–201.
- Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23):3317–3324.
- Chmátal L, Gabriel SI, Mitsainas GP, Martínez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. 2014. Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr Biol.* 24(19):2295–2300.
- Close RL, Bell JN, Dollin AE, Harding HR. 1996. Spermatogenesis and synaptonemal complexes of hybrid *Petrogale* (Marsupialia). *J Hered.* 87(2):96–107.
- Close RL, Bell JN. 1997. Fertile hybrids in two genera of wallabies: *Petrogale* and *Thylogale*. *J Hered.* 88(5):393–397.
- Coyne JA, Orr HA. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43(2):362–381.
- Coyne JA, Orr HA. 2004. *Speciation*. Sunderland (United Kingdom): Sinauer Associates.
- Dagilis AJ, Kirkpatrick M. 2016. Prezygotic isolation, mating preferences, and the evolution of chromosomal inversions. *Evolution* 70(7):1465–1472.
- Damas J, Corbo M, Lewin HA. 2021. Vertebrate chromosome evolution. *Annu Rev Anim Biosci.* 9:1–27.
- Deakin JE, O'Neill RJ. 2020. Evolution of marsupial genomes. *Annu Rev Anim Biosci.* 8:25–45.
- de Vos JM, Augustijnen H, Livio B, Kay L. 2020. Speciation through chromosomal fusion and fission in Lepidoptera. *Philos Trans R Soc Lond B Biol Sci.* 375(1806):20190539.
- Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21(2):113–135.
- Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol.* 28(8):2239–2252.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* 7(10):e1002195.
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–599.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.
- Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63(1):1–19.
- Eldridge MDB, Close RL. 1992. Taxonomy of rock wallabies, *Petrogale* (Marsupialia, Macropodidae). I. A revision of the eastern *Petrogale* with the description of 3 new species. *Aust J Zool.* 40(6):605–625.
- Eldridge MDB, Close RL. 1993. Radiation of chromosome shuffles. *Curr Opin Genet Dev.* 3(6):915–922.
- Eldridge MDB, Close RL. 1997. Chromosomes and evolution in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). *Aust Mammal.* 19:123–136.
- Eldridge MDB, Close RL, Johnston PG. 1990. Chromosomal rearrangements in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). III. G-banding analysis of the *Petrogale inornata* and *P. penicillata*. *Genome* 33(6):798–802.
- Eldridge MDB, Close RL, Johnston PG. 1991. Chromosomal rearrangements in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). IV. G-banding analysis of the *Petrogale lateralis* complex. *Aust J Zool.* 39(6):621–627.
- Eldridge MDB, Johnston PG, Close RL. 1992. Chromosomal rearrangements in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). VI. Determination of the plesiomorphic karyotype: G-Banding comparison of *Thylogale* with *Petrogale persephone*, *P. xanthopus*, and *P. l. lateralis*. *Cytogenet Cell Genet.* 61(1):29–33.
- Eldridge MDB, Johnston PG, Close RL, Lowry PS. 1989. Chromosomal rearrangements in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). II. G-banding analysis of *Petrogale godmani*. *Genome* 32(6):935–940.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14(8):2611–2620.
- Fan H, Wu Q, Wei F, Yang F, Ng BL, Hu Y. 2019. Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biol.* 20(1):267–278.

- Faria R, Navarro A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol.* 25(11):660–669.
- Fitak RR. 2021. *OptM*: estimating the optimal number of migration edges on population trees using *Treemix*. *Biol Methods Protoc.* 6(1):bpab017.
- Flicke P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2014. Ensembl 2014. *Nucleic Acids Res.* 42(D1):D749–D755.
- Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347(6217):1258524.
- Franchini P, Kautt AF, Nater A, Antonini G, Castiglia R, Meyer A, Solano E. 2020. Reconstructing the evolutionary history of chromosomal races on islands: a genome wide analysis of natural house mouse populations. *Mol Biol Evol.* 37(10):2825–2837.
- Garagna S, Page J, Fernandez-Donoso R, Zuccotti M, Searle JB. 2014. The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma* 123(6):529–544.
- Giménez MD, White TA, Hauffe HC, Panithanarak T, Searle JB. 2013. Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. *Evolution* 67(5):1446–1462.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.
- Gropp A, Winking H, Redi CA. 1982. Consequences of Robertsonian heterozygosity: segregational impairment of fertility versus male-limited sterility. In: Crosignani PG, Rubin BL, Faccaro M, editors. Genetic control of gamete production and function. London (United Kingdom): Academic Press. p. 115–134.
- Gruber B, Unmack PJ, Berry OF, Georges A. 2018. DARTR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour.* 18(3):691–699.
- Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution* 68(10):2747–2756.
- Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. *J Gen.* 12(2):101–109.
- Hoffman AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Syst.* 39:21–42.
- Horn A, Basset P, Yannic G, Banaszek A, Borodin PM, Bulatova NS, Jadwiszczak K, Jones RM, Polyakov AV, Ratkiewicz M, et al. 2012. Chromosomal rearrangements do not seem to affect the gene flow in hybrid zones between karyotypic races of the common shrew (*Sorex araneus*). *Evolution* 66(3):882–889.
- Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9(9):868–877.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23(2):254–267.
- Janes JK, Miller JM, Dupuis JR, Malenfant RN, Gorrell JC, Cullingham CI, Andrew RL. 2017. The K=2 conundrum. *Mol Ecol.* 26(14):3594–3602.
- Jombart T. 2008. Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.
- Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.
- Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27(6):757–763.
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, et al. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. In: Pompanon F, Bonin A, editors. Data production and analysis in population genomics. Methods in molecular biology (methods and protocols). Vol. 888. Totowa (NJ): Humana Press. p. 67–89.
- King M. 1993. Species evolution: the role of chromosome change. Cambridge: Cambridge University Press.
- Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173(1):419–434.
- Kong S, Zhang Y. 2019. Deciphering Hi-C: from 3D genome to function. *Cell Biol Toxicol.* 35(1):15–31.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour.* 15(5):1179–1191.
- Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Larson EL, Keeble S, Vanderpool D, Dean MD, Good JM. 2017. The composite regulatory basis of the large X-effect in mouse speciation. *Mol Biol Evol.* 34:282–295.
- Leaché AD, Banbury BL, Linkem CW, de Oca ANM. 2016. Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in North American spiny lizards. *BMC Evol Biol.* 16:63.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Linck E, Epperly K, Van Els P, Spellman GM, Bryson RW, McCormack JE, Canales-Del-Castillo R, Klicka J. 2019. Dense geographic and genomic sampling reveals paraphyly and a cryptic lineage in a classic sibling species complex. *Syst Biol.* 68:956–966.
- Linkem CW, Minin VN, Leaché AD. 2016. Detecting the anomaly zone in species trees and evidence for misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Syst Biol.* 65(3):465–477.
- Maddison WP. 1997. Gene trees in species trees. *Syst Biol.* 46(3):523–536.
- Magoc T, Salzberg S. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963.
- Malinsky M, Matschiner M, Svardal H. 2021. Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour.* 21(2):584–595.
- Manoukis NC, Powell JR, Toure MB, Sacko A, Edillo FE, Coulibaly MB, Traoré SF, Taylor CE, Besansky NJ. 2008. A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proc Natl Acad Sci U S A.* 105(8):2940–2945.
- Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biol.* 5(9):e243.
- McGee MD, Borstein SR, Meier JJ, Marques DA, Mwaiko S, Taabu A, Kische MA, O'Meara B, Bruggmann R, Excoffier L, et al. 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* 586(7827):75.
- McKenna AH, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.
- McLaughlin JF, Winker K. 2020. An empirical examination of sample size effects on population demographic estimates in birds using single nucleotide polymorphism (SNP) data. *PeerJ* 8:e9939.
- Mérot C, Oomen RA, Tigano A, Wellenreuther M. 2020. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol.* 35(7):561–572.
- Metcalfe CJ, Eldridge MDB, McQuade LR, Johnston PG. 1997. Mapping the distribution of telomeric sequence (T₂AG₃)_n in rock-wallabies, *Petrogale* (Marsupialia: Macropodidae), by fluorescence in situ hybridization. I. The *penicillata* complex. *Cytogenet Cell Genet.* 78(1):74–80.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010(6):pdb.prot5448.doi:10.1101/pdb.prot5448
- Muller H. 1942. Isolating mechanisms, evolution and temperature. *Biol Symp.* 6:71–125.
- Navarro A, Barton NH. 2003. Accumulating postzygotic isolation in parapatry: a new twist on chromosomal speciation. *Evolution* 57(3):447–459.

- Noor MAF, Garfield DA, Schaeffer SW, Machado CA. 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* 177(3):1417–1428.
- Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A*. 98(21):12084–12088.
- O'Neill RJ, Eldridge MD, Toder R, Ferguson-Smith MA, O'Brien PC, Graves JA. 1999. Chromosome evolution in kangaroos (Marsupialia: Macropodidae): cross species chromosome painting between the tamar wallaby and rock wallaby spp. with the 2n=22 ancestral macropodid karyotype. *Genome* 42(3):525–530.
- Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom.* 2(4):e000056.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst Biol.* 64(4):651–662.
- Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 31(7):1929–1936.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Potter S, Bragg JB, Blom MPK, Deakin J, Kirkpatrick M, Eldridge MDB, Moritz C. 2017. Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of rock-wallabies. *Front Genet.* 8:10.
- Potter S, Cooper SJ, Metcalfe CJ, Taggart DA, Eldridge MDB. 2012. Phylogenetic relationships of rock-wallabies *Petrogale* (Marsupialia: Macropodidae) and their biogeographic history within Australia. *Mol Phylogenet Evol.* 62(2):640–652.
- Potter S, Moritz C, Eldridge MDB. 2015. Gene flow despite complex Robertsonian fusions among rock-wallaby (*Petrogale*) species. *Biol Lett.* 11(10):20150731.
- Presgraves DC. 2018. Evaluating genomic signatures of “the large X-effect” during complex speciation. *Mol Ecol.* 27(19):3822–3830.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
- Queney G, Ferrand N, Weiss S, Mougél F, Monnerot M. 2001. Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Mol Biol Evol.* 18(12):2169–2178.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16(6):276–277.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol.* 16(7):351–358.
- Rieseberg LH, Linder CR, Seiler GJ. 1995. Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics* 141(3):1163–1171.
- Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN. 2014. Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol.* 14:254.
- Searle JB. 1993. Chromosomal hybrid zones in eutherian mammals. In: Harrison RG, editor. *Hybrid zones and the evolutionary process*. New York: Oxford University Press. p. 309–353.
- Searle JB, Polly PD, Zima J. 2019. *Shrews, chromosomes and speciation*. Cambridge: Cambridge University Press.
- Sharman GB, Close RL, Maynes GM. 1989. Chromosomal evolution, phylogeny and speciation of rock wallabies (*Petrogale*: Macropodidae). *Aust J Zool.* 37(3):351–363.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Singhal S. 2013. *De novo* transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Mol Ecol Resour.* 13(3):403–416.
- Sites JW, Moritz C. 1987. Chromosomal evolution and speciation revisited. *Syst Zool.* 36(2):153–174.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.
- Sunnucks P, Hales DF. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol.* 13(3):510–524.
- Swofford DL. 2003. PAUP. Phylogenetic analysis using parsimony (and other methods). Version 4. Sunderland (United Kingdom): Sinauer Associates.
- Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, Zhao F, Sun Z, Zhang J. 2017. Population genomics reveals speciation and introgression between Brown Norway rats and their sibling species. *Mol Biol Evol.* 34(9):2214–2228.
- Thawornwattana Y, Dalquen D, Yang Z. 2018. Coalescent analysis of phylogenomic data confidently resolves the species relationships in the *Anopheles gambiae* species complex. *Mol Biol Evol.* 35(10):2512–2527.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J. 2014. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 43:11.10.1–11.10.33.
- Vicoso B, Charlesworth B. 2006. Evolution of the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7(8):645–653.
- Wang Q, Sun M, Zhang Y, Song Z, Zhang S, Zhang Q, Xu J-R, Liu H. 2020. Extensive chromosomal rearrangements and rapid evolution of novel effector superfamilies contribute to host adaptation and speciation in the basal ascomycetous fungi. *Mol Plant Pathol.* 21(3):330–348.
- Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. *Genome Res.* 27(5):757–767.
- Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Syst Biol.* 67(4):735–740.
- White MJD. 1969. Chromosomal rearrangements and speciation in animals. *Annu Rev Genet.* 3(1):75–98.
- Yannic G, Basset P, Horn A, Hausser J. 2019. Gene flow between chromosomal races and species. In: Searle JB, Polly PD, Zima J, editors. *Shrews, chromosomes and speciation*. Cambridge: Cambridge University Press. p. 313–335.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18(5):821–829.
- Zheng Y, Janke A. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics* 19:10.