Genetics: Early Online, published on June 6, 2018 as 10.1534/genetics.118.300894

THE GENOMIC ARCHITECTURE OF A RAPID ISLAND RADIATION: RECOMBINATION RATE VARIATION, CHROMOSOME STRUCTURE, AND GENOME ASSEMBLY OF THE HAWAIIAN CRICKET *LAUPALA*

Thomas Blankers[1], Kevin P. Oh[1], Aureliano Bombarely[2], Kerry L. Shaw[1]

[1] *Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA*

[2] *Department of Horticulture, Virginia Tech, Blacksburg, VA, USA*

Running Title: *Laupala* Genomic Architecture

Key words: speciation, sexual selection, recombination, chromosomal rearrangements, genome, crickets

Corresponding author:

Thomas Blankers, Department of Neurobiology and Behavior, Cornell University, 215 Tower Rd, 14850, Ithaca, NY, USA. thomasblankers@gmail.com. 1- (607) 254 4326

ABSTRACT

Phenotypic evolution and speciation depend on recombination in many ways. Within populations, recombination can promote adaptation by bringing together favorable mutations and decoupling beneficial and deleterious alleles. As populations diverge, cross-over can give rise to maladapted recombinants and impede or reverse diversification. Suppressed recombination due to genomic rearrangements, modifier alleles, and intrinsic chromosomal properties may offer a shield against maladaptive gene flow eroding co-adapted gene complexes. Both theoretical and empirical results support this relationship. However, little is known about this relationship in the context of behavioral isolation, where co-evolving signals and preferences are the major hybridization barrier. Here we examine the genomic architecture of recently diverged, sexually isolated Hawaiian swordtail crickets (*Laupala*). We assemble a *de novo* genome and generate three dense linkage maps from interspecies crosses. In line with expectations based on the species' recent divergence and successful interbreeding in the lab, the linkage maps are highly collinear and show no evidence for large-scale chromosomal rearrangements. Next, the maps were used to anchor the assembly to pseudomolecules and estimate recombination rates across the genome to test the hypothesis that loci involved in behavioral isolation (song and preference divergence) are in regions of low interspecific recombination. Contrary to our expectations, the genomic region where a male song and female preference QTL co-localize is not associated with particularly low recombination rates. This study provides important novel genomic resources for an emerging evolutionary genetics model system and suggests that trait-preference co-evolution is not necessarily facilitated by locally suppressed recombination.

INTRODUCTION

Speciation is spurred by the accumulation of genomic variation resulting in the formation of barriers that prevent gene flow between populations. Genomes diverge under the influence of selection and drift, while gene flow counteracts this divergence by homogenizing the genome (Felsenstein 1981; Kirkpatrick and Ravigne 2002; Gavrilets 2003). To appreciate the speciation process and the origin of the fascinating diversity of life on earth, we need to understand the interaction between the mechanisms that change allele frequencies and the mechanisms that govern the association of beneficial and deleterious alleles with other alleles. A key process in this interaction is recombination, which creates new allelic combinations during meiosis in sexually reproducing organisms.

Any association between loci that underlies environmental adaptation or between loci underlying co-evolving (sexual) signals and signal responses (i.e. co-adapted gene complexes) will be affected by recombination (Felsenstein 1981). Within populations, recombination can mitigate Hill-Robertson interference by combining locally adaptive alleles from different genomic backgrounds and by decoupling beneficial and deleterious alleles (Hill and Robertson 1966); recombination can also influence the covariance between sexual traits and preference across sexes (Smith and Haigh 1974; Smith 1978; Gillespie 2000; Otto 2009). As such, recombination might increase the efficiency of background selection (purging deleterious alleles), sexual selection (through signal-preference co-evolution), and local adaptation (by linking locally adapted alleles) in the earliest stages of speciation (Kirkpatrick and Ravigne 2002; Butlin 2005; Kirkpatrick and Barton 2006; Yeaman and Whitlock 2011).

Between divergent populations with some (but incomplete) reproductive isolation, recombination can also counteract population divergence and prevent the closure of a reproductive boundary by

creating combinations of alleles that are favorable in different contexts (Noor *et al.* 2001; Rieseberg 2001; Coyne and Orr 2004; Ortiz-Barrientos *et al.* 2016). Interspecific recombination is constrained both by intrinsic properties of the species' genomes that also constrain intraspecific recombination, as well as by genetic and chromosomal changes between populations, such as incompatibilities, locally adaptive alleles, and adaptive modifications of between-population recombination (Yeaman and Whitlock 2011; Feder *et al.* 2012). The most intensely studied modifier of recombination between divergent populations are inversions. Inversions can suppress recombination locally in the genome and, thus, promote reproductive isolation, by trapping genetic incompatibilities in linkage blocks (Noor *et al.* 2001), acting synergistically with other genes causing isolation (Rieseberg 2001), or by linking locally adaptive alleles (Kirkpatrick and Barton 2006). Other chromosomal rearrangements, such as translocations and transposable elements, can likewise contribute to 'chromosomal speciation' (Rieseberg 2001) as well as to preventing gene flow and furthering genetic divergence among heterospecifics.

Interestingly, there is ubiquitous among-species variation in recombination rates (Wilfert *et al.* 2007; Smukowski and Noor 2011). In insects, for example, rates vary from 16.1 cM/Mb (centi-Morgans per megabase) in *Apis melifera* to 0.1 cM/Mb in the mosquito *Armigeres subalbatus* (Wilfert *et al.* 2007). There is also variation across the genome within individuals. For example, 50-fold differences have been observed within single chromosomes of humans and birds (Myers *et al.* 2005; Singhal *et al.* 2015). These patterns of variation underline that the efficacy of selection acting within species may differ across taxa and across genomes of the same species.

A major prediction following from theoretical work is that favorable allele combinations that promote ecological adaptation are more likely to reside in regions of low recombination.

Recombination counteracts adaptation by natural selection by breaking up associations between segregating alleles that are locally adaptive within the resident population and counteracts divergent selection if there is gene flow between recently diverged populations (Bürger and Akerman 2011; Yeaman and Whitlock 2011; Yeaman 2013). So far, empirical evidence for the prediction that locally adaptive alleles reside in regions of low recombination is not conclusive (Roesti *et al.* 2013; Burri *et al.* 2015; Marques *et al.* 2016). However, a recent study indicated that the interaction between gene flow and divergent selection is a strong predictor for the association between adaptive alleles and regions of low recombination in multiple species of stickleback fish (Samuk *et al.* 2017).

It is less clear how these predictions apply to the evolution of behavioral isolation. Theoretical models of speciation by sexual selection depend on linkage disequilibrium between sexual signaling traits and corresponding preference genes (Fisher 1930; Lande 1981; Kirkpatrick 1982). Linkage disequilibrium between trait and preference genes can come about by assortative mating (Lande 1981; Andersson and Simmons 2006) or by physical linkage (Kirkpatrick and Hall 2004), either through closely linked loci or through pleiotropy (a single gene affecting both signal and preference phenotypes). On the one hand, recombination can help consolidate loci brought together by nonrandom mating and as such facilitate linkage disequilibrium between trait and preference (Kirkpatrick and Ravigne 2002). On the other hand, recombination can also tear apart co-adapted trait and preference alleles if genes are exchanged between populations that differ in mating phenotypes. Therefore, recombination between sexually divergent populations in sympatry and parapatry often compromises differentiation in mating phenotypes and hinders speciation (Arnegard and Kondrashov 2004; Servedio 2009, 2015; Servedio and Burger 2014). However, there has been limited empirical insight into the relationship between trait-preference

co-evolution and genome-wide variation in recombination rates (see Davey *et al.* 2017 for a recent exception).

Here, we examine the genomic architecture, specifically structural variation and heterogeneity in interspecific recombination, of four closely related, sexually isolated species of Hawaiian swordtail crickets from the genus *Laupala*. *Laupala* is one of the fastest speciating taxa known to date (Mendelson and Shaw 2005). The 38 morphologically cryptic species, each endemic to a single island of the Hawaiian archipelago (Otte 1994; Shaw 2000a) are the product of a recent evolutionary radiation. Evidence suggests that speciation by sexual selection on the acoustic communication system has driven this rapid diversification, as both male mating song and female acoustic preferences have diverged extensively among *Laupala* species (Otte 1994; Shaw 2000b; Mendelson and Shaw 2002). Sexual trait evolution strongly contributes to the onset and maintenance of reproductive isolation (Mendelson and Shaw 2002; Grace and Shaw 2011). Quantitative variation in one key temporal property of male song (pulse rate) and corresponding female preference strongly covaries across species and across populations within species (Shaw 2000b; Grace and Shaw 2011). Although the mechanisms of trait-preference co-evolution require further study, there is evidence that both are associated with a polygenic basis and that genetic loci controlling quantitative variation in traits and preferences are physically linked in the genome (Shaw and Lesnick 2009; Wiley *et al.* 2012). Notably, one of the major song quantitative trait loci (QTL; haploid effect size ~ 9%) co-localizes with the first mapped preference QTL (haploid effect size ~ 14%). Directional effects of song QTL provide additional evidence that (sexual) selection is driving divergence between species (Shaw *et al.* 2007).

The species pairs involved in this study, *L. kohalensis* and *L. pruna*, and *L. paranigra* and *L. kona*, are endemic to the Big Island, the youngest island of the Hawaiian archipelago (Fig 1A,

B). Although these species pairs have apparently diverged in allopatry within the Big Island, past or future migration is likely, given their geographical proximity. Indeed, although allopatric and more closely related to *L. kohalensis*, *L. pruna* currently overlaps in distribution with *L. paranigra* (Fig 1B). The discordance between nuclear and mitochondrial phylogenies (Shaw 2002) and the limited degree of postzygotic isolation between some species pairs further emphasize the possibility of gene flow across natural populations. Together, the biogeography and the genetics of song and preference variation in this system provide a unique opportunity to explore the interaction between interspecific recombination rate variation, co-evolution of mating traits, and speciation.

We first assemble a *de novo L. kohalensis* draft genome and then obtain thousands of SNP markers for heterogeneously hybrid offspring from three laboratory-generated interspecific crosses. We then generate three dense linkage maps and compare these maps to test the hypothesis that the genomic architectures of young, sexually differentiated species are largely collinear (similar marker order) and have few genetic incompatibilities and conserved interspecific recombination frequencies (similar marker distances). There is some variation in the level of overall differentiation in the species pairs studied here, but all lineages are young (approximately 0.5 million years or less, Fig 1). It is commonly expected that strong prezygotic isolation can evolve rapidly and largely in the absence of intrinsic postzygotic isolating mechanisms (Coyne and Orr 2004), but explicit comparisons of chromosomal architectures across behaviorally isolated species are rare. We compare the maps visually and use variation in maker order and length (measured in genetic distance, or centi-Morgans [cM]) as indicators of possible chromosomal rearrangements affecting the recombination rates differently in different crosses (Fig 1C). Then, from the large amount of information on linkage across many genomic

markers from three hybrid crosses, we anchor the draft genome assembly to pseudomolecules and estimate the landscape of recombination across the genome. Finally, using an additional map that integrates the amplified fragment length polymorphism (AFLP) markers from previous QTL studies of the *L. kohalensis* and *L. paranigra* species pair, we approximate the location of known male song QTL, including one co-localizing with a female acoustic preference QTL, on the pseudomolecules. We examine local variation in recombination rates across the genome and in relation to the location of the song and preference QTL to test the hypothesis that song-preference co-evolution is facilitated by suppressed interspecific recombination. This study provides important insight into the role of the genomic architecture during divergence of closely related species separated by premating barriers.

MATERIAL & METHODS

De novo genome assembly

The *Laupala kohalensis* draft genome (estimated genome size ~ 1.9 Gb; Petrov *et al.* 2000) was sequenced using the Illumina HiSeq 2500 platform. DNA was isolated with the DNeasy Blood & Tissue Kits (Qiagen Inc., Valencia, CA, USA) from six immature female crickets (c. five months of age) chosen randomly from a laboratory stock population (approximate lab generation=14). Females were chosen to balance DNA content of sex chromosomes to autosomes (female crickets are XX; male crickets are XO). DNA was subsequently pooled for sequencing. Four different libraries were created: a paired-end library with an estimated insert size of 200 bp (sequenced by Cornell Biotechnology Resource Center), a paired-end library with an estimated insert size of 500 bp, and two mate-pair libraries with insert sizes of 2 and 5 Kb (sequenced by Cornell Weill College Genomics Resources Core Facility).

Reads were processed using Fastq-mcf from the Ea-Utils package (Aronesty 2011) with the parameters -q 30 (trim nucleotides from the extremes of the read with qscore below 30) and -l 50 (discard reads with lengths below 50 bp). Read duplications were removed using PrinSeq v0.20.4 (Schmieder and Edwards 2011) and reads were corrected using Musket v1.0.8 with the default parameters (Liu *et al.* 2013).

Reads were assembled using SoapDeNovo2 (Luo *et al.* 2012). The reads were assembled using different Kmer sizes (k = 31, 39, 47, 55, 63, 71, 79 and 87). The 87-mer assembly produced the best assembly (based on N50/L50, assembly size, and number of scaffolds). Scaffolds and contigs were renamed using an in-house Perl script. Gaps were filled using GapCloser from the SoapDeNovo2 package.

The gene space covered by the assembly was evaluated using three different approaches: (1) *Laupala kohalensis* unigenes produced by the Gene Index initiative (Cricket release 2.0: http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=cricket) were mapped using Blat v35x1 (Kent 2002). Only unigenes mapping with 90% or more of their length were considered; (2) 50 bp paired-end RNA-seq reads from a congeneric species, *L. cerasina* were mapped using Tophat2 (Kim *et al.* 2013). Reads were processed using the same methodology described above, but using a minimum length of 30 bp; (3) using BUSCO v.2.0.1 (Simão *et al.* 2015) to search for conserved eukaryotic and arthropod genes.

Samples

We generated three F_2 interspecies hybrid families to estimate genetic maps. Multiple F_1 male and sibling females were intercrossed to generate F_2 mapping populations for the following species crosses: (1) a *L. kohalensis* female and *L. paranigra* male (“ParKoh”, 178 genotyped F_2

hybrid offspring; previously reported in Shaw et al. 2007); (2) a *L. kohlanesis* female and a *L. pruna* male ("PruKoh", 193 genotyped F_2 hybrid offspring); (3) a *L. paranigra* female and a *L. kona* male ("KonPar", 263 genotyped F_2 hybrid offspring). These four species are part of a recently radiated clade showing conspicuous mating song divergence (Mendelson and Shaw 2005). Approximate geographic distributions of the species, phylogenetic relationships and parent collection localities are shown in Fig 1 and in Table S1. Crickets used in crosses were a combination of lab stock and outbred individuals (*L. kohalensis* [for ParKoh] and *L. paranigra* [for ParKoh and KonPar] were both lab reared for 3-15 generations; *L. kohalensis* [for PruKoh], *L. pruna* and *L. kona* were wild-caught). All parental and hybrid generations were reared in a temperature-controlled room (20°C) on Purina cricket chow and provided water *ad libitum*.

Genotyping

DNA was extracted from whole adults using the DNeasy Blood & Tissue Kits (Qiagen, Valencia, CA, USA). Genotype-by-Sequencing library preparation and sequencing were done at the Genomic Diversity Facility at Cornell University in 2014 following Elshire *et al.* (2011). The *Pst* I restriction enzyme was used and DNA was sequenced (1x100bp) on the Illumina HiSeq 2000 platform (Illumina Inc., USA).

Reads were trimmed and demultiplexed using Flexbar v2.5 (Dodt *et al.* 2012) and then mapped to the *L. kohalensis de novo* draft genome using Bowtie2 v2.2.6 (Langmead and Salzberg 2012) with default parameters. We then called SNPs using two different pipelines: The Genome Analysis Toolkit v3.6.0 (GATK; DePristo *et al.* 2011; Van der Auwera *et al.* 2013) and FreeBayes v0.9.13 (Garrison and Marth 2012). For GATK we used individual BAM files to generate gVCF files using 'HaplotypeCaller' followed by the joint genotyping step 'GenotypeGVCF'. We then evaluated variation in SNP quality across all genotypes using custom

R scripts (R Development Core Team 2016) to determine appropriate settings for hard filtering using the following metrics (based on the recommendations for hard filtering section "Understanding and adapting the generic hard-filtering recommendations" at https://software.broadinstitute.org/gatk/ accessed on 28 February 2017): quality-by-depth, Phred-scaled *P*-value using Fisher's Exact Test to detect strand bias, root mean square of the mapping quality of the reads, u-based z-approximation from the Mann-Whitney Rank Sum Test for mapping qualities, u-based z-approximation from the Mann-Whitney Rank Sum Test for the distance from the end of the read for reads with the alternate allele. For FreeBayes we called variants from a merged BAM file using standard filters. After variant calling we filtered the SNPs using 'vcffilter', a Perl library part of the VCFtools package (Danecek *et al.* 2011) based on the following metrics: quality (> 30), depth of coverage (> 10), and strand bias for the alternative and reference alleles (SAP and SRP, both > 0.0001). Finally, the variant files from the GATK pipeline and the FreeBayes pipeline were filtered to only contain biallelic SNPs with less than 10% missing genotypes using VCFtools v0.1.15.

We retained two final variant sets: a high-confidence set including only SNPs with identical genotype calls between the two variant discovery pipelines and the full set of SNPs which included all variants called using FreeBayes but limited to positions that were shared among the GATK and FreeBayes pipelines.

Linkage mapping

The genotype information from the parental lines was used to assign ancestry to the SNP loci. The parents of the crosses were heterogeneously heterozygous and only ancestry informative loci were retained, i.e. all loci for which one or both parents were heterozygous were discarded. We were unable to obtain sequence data from the parents for PruKoh, but used sequence data from a

252 single, non-parental *L. pruna* female, and three available *L. kohalensis* females, all from the
253 same populations as the parents. Ancestry was inferred if all three *L. kohalensis* individuals were
254 homozygous for one allele and the *L. pruna* individual was homozygous for the alternative allele.
255 All other loci were discarded. The loci were then further filtered based on genotype similarity
256 and segregation distortion (see below for details).

257 The linkage maps deriving from the three species crosses were generated independently
258 and by taking a three-step approach, employing both the regression mapping and the maximum
259 likelihood (ML) mapping functions in JoinMap 4.0 (van Ooijen 2006) as well as the three-point
260 error-corrected ML mapping function in MapMaker 3.0 (Lander *et al.* 1987; Lincoln *et al.* 1993).
261 We used an intercross (F2) scheme in both programs, because we limited our SNP markers to
262 those that are fixed between the parents (i.e. the only genotypes are A, B, and H).

263 In the first step, we estimated "initial" maps that are relatively low resolution (5 cM) but
264 with high marker order certainty. For initial maps, we first grouped ($3.0 \leq \text{LOD} \leq 5.0$) and then
265 ordered the high-confidence markers that showed no segregation distortion (markers with χ^2-
266 square associated *P*-value for deviation from Mendelian inheritance < 0.05 were discarded) and
267 for which no marker had more than 95% similarity in genotypes across individuals compared to
268 other markers (otherwise, one of each pair was excluded). When excluding similar loci, we
269 favored those marker loci shared among the three mapping populations over markers unique to
270 one or two crosses. We then checked for concordance among the three mapping algorithms. In
271 most cases, the maps had highly concordant marker orders; any (rare) marker order discrepancies
272 between maps for the same cross were resolved by choosing the order that resulted in the highest
273 map likelihood, shortest total map length, and the best fit of markers based on pairwise
274 recombination frequencies with neighboring markers (see also JoinMap 4.0 manual).

These initial maps were then filled out using MapMaker with marker loci passing slightly more lenient criteria: markers drawn from the full set of SNPs, with false discovery rate (Benjamini and Hochberg 1995) corrected *P* value for χ^2- square test of deviation from Mendelian inheritance ≤ 0.05 and fewer than 99% of their genotypes in common with other markers loci. First, more informative markers (no missing genotypes, > 2.0 cM distance from other markers) were added satisfying a log-likelihood threshold of 4.0 for the positioning of the marker (i.e., assigned marker position is 10,000 times more likely than any other position in the map). Remaining markers were added at the same threshold, followed by a second round for all markers at a log-likelihood threshold of 3.0. We then used the ripple algorithm on 5-marker windows and explored alternative orders.

In the second step, "comprehensive" maps were obtained in MapMaker by sequentially adding markers from the full set of SNPs that met the more lenient criteria described above to the initial map. Markers were added if they satisfied a log-likelihood threshold of 2.0 for the marker positions, followed by a second round with a log-likelihood threshold of 1.0. We then used the ripple algorithm again on 5-marker windows and explored alternative orders. Typically, MapMaker successfully juxtaposes SNP markers from the same scaffold. However, in marker dense regions with low recombination rates, the likelihoods of alternative marker orders coalesce. In such regions, when multiple markers from the same genomic scaffold were interspersed by markers from a different scaffold, we repositioned the former markers by forcing them in the map together. If the log-likelihood of the map decreased by more than 3.0 (factor 1000), only one of the markers from that scaffold was used in the map. The comprehensive maps provide a balance between marker density and confidence in marker ordering and spacing.

The third step was to create "dense" maps. We added all remaining markers that were not yet incorporated in step two, first at a log-likelihood threshold of 0.5, followed by another round at a log-likelihood threshold of 0.1. We then used the ripple command as described above. The dense maps are useful for anchoring of scaffolds and for obtaining the highest possible resolution of variation in recombination rates, but with the caveat that there is some uncertainty in marker order. Uncertainty is expected to be higher towards the centers of the linkage groups where crossing over events between adjacent markers become substantially less frequent (see Results).

Comparative analyses

Based on the recent divergence times and high interbreeding successes, we predict a large degree of collinearity of the linkage maps. We note that interpretations must take into account the non-independence of the ParKoh and PruKoh/KonPar maps, as only comparing PruKoh and KonPar comprises a fully independent contrast. We first examined whether inversions or other chromosomal rearrangements were common (affecting linkage map lengths and marker orders) or whether maps were generally collinear by comparing among the initial and comprehensive linkage maps visually using map graphs from MapChart v2.3 (Voorrips 2002). Inverted or transposed markers present in two or all maps can be detected by connecting "homologs" in MapChart (a homolog in this case means a scaffold that is represented in two or more maps). Then, we tested whether linkage maps are generally collinear across the species pairs quantitatively. We calculated the Spearman's rank order correlation test statistic (ρ) and corresponding *P*-value (the probability of observing the measured or stronger correlation given no true correlation exists) to examine the correlation between the order of shared markers in homologous linkage groups using the cor.test() function in R v3.3.1.

We then tested for genetic incompatibilities among the genomes of the four species, by measuring segregation distortion in sliding, 10 cM windows. Although we filtered out markers with very high levels of segregation distortion (using a 5% FDR cutoff) to purge markers with potential sequencing errors, groups of distorted markers in a single region of a linkage group represent genomic regions with biased parental allele contributions, suggesting genetic incompatibilities (or, less common, selfish alleles and other active segregation distorters). Because *L. kohalensis* and *L. paranigra* are more distantly related to each other (and, thus, allowing more time for genetic incompatibilities to accumulate) than they are to *L. pruna* and *L. kona*, respectively (Mendelson and Shaw 2005; see Fig 1.), we expected more regions with significant segregation distortion in the ParKoh map relative to the KonPar and PruKoh maps. We calculated genotype frequency and the negative 10-base logarithm of the *P* value for the χ^2-square test of deviation from Mendelian inheritance across the linkage groups in R using the R/qtl package v1.42 (Broman *et al.* 2003). Windows with $P < 0.01$ were considered to have significant segregation distortion, and thus potentially reflecting genetic incompatibilities.

After establishing that the linkage maps were generally collinear (see Results), we merged the maps and examined patterns of variation in crossing over along the *Laupala* genome. Maps were consolidated using ALLMAPS v0.7.1 (Tang *et al.* 2015). Then, we calculated species-specific average recombination rates for the linkage groups by dividing the total length of the linkage group (in cM) by the physical length of the pseudomolecule (in million bases, Mb) obtained by merging homologous linkage groups using ALLMAPS. Lastly, to evaluate recombination rate variation along the linkage groups, we fitted smoothing splines (with 10 degrees of freedom, based on the fit of the spline to the observed data) in R to describe the relationship between the consensus physical distance (as per the anchored scaffolds) and the genetic distance specific to

each linkage map. Variation in the recombination rate was then assessed by taking the first derivative (i.e. the rate) of the fitted spline function. The estimated recombination rates are likely to be an overestimate of the true recombination rate, because unplaced/unordered parts of the assembly do not contribute to the physical length of the pseudomolecules but are reflected in the genetic distances obtained from crossing-over events in the recombining hybrids.

To test the hypothesis that linked trait and preference genes reside in low recombination regions, we integrated the AFLP map and song and preference QTL peaks identified in previous work on *L. kohalensis* and *L. paranigra* (Shaw and Lesnick 2009) with the current ParKoh SNP map and projected the QTL peaks onto the anchored genome. The SNPs used in the present study were obtained from the same mapping population (same individuals) as in the 2009 AFLP study. Therefore, we combined the high confidence SNPs described above (for the "initial" map) with the AFLP markers reported in (Shaw *et al.* 2007) that were of the same individuals as the SNP markers used in this study and created a new linkage map using the same stringent criteria as for the "initial" maps described above. We projected this map onto the anchored draft genome based on common markers (scaffolds). We then approximated the physical location of the QTL peaks by looking for SNP markers on scaffolds present in the draft genome flanking AFPL markers underneath the QTL peaks identified in the 2009 study.

DATA AVAILABILITY

Raw data (genotype data, linkage maps, pseudomolecule agp file) and supplementary files are available on FigShare (doi to be announced). Supplementary data: Table S1. Geographic locations of sampled populations; Table S2. Segregation distortion (count of heterozygotes per genotype) statistics; Table S3. Summary statistics for anchored assembly; Table S4. Integrated AFLP and SNP map for the *L. kohalensis* x *L. paranigra* cross; Figure S1. Comprehensive

linkage maps; Figure S2. ALLMAPS output; Figure S3. Coverage per cross per linkage group. Custom R-scripts are also available on FigShare (doi to be announced). The *L. kohalensis* genome assembly and sequencing reads are available on NCBI's GenBank under BioProject number PRJNA392944. The genotype-by-sequencing reads are available at NCBI's short read archive under BioProject number PRJNA429815.

RESULTS

De novo genome assembly

The sequencing of the four libraries yielded 162.5 Gb of raw sequences (Table 1). After read processing, 145.5 Gb was used for the sequence assembly. We compared among assemblies resulting from different Kmer sizes (k = 31, 39, 47, 55, 63, 71, 79 and 87). Based on the N50/L50 and the total assembly size, the assembly produced with k = 87 was retained for the final draft genome. Despite a large number of scaffolds in the final assembly (149,424), the median length of the scaffolds was high and the total length of the assembly covers about 83% of the expected complete genome in *Laupala* (Table 1).

Gene space coverage in the assembly was evaluated using the *L. kohalensis* cricket gene index (Danley *et al.* 2007) (release 2.0), RNASeq from *Laupala cerasina* (Blankers *et al.* 2018), and by performing a BUSCO search using eukaryotic and arthropod specific conserved genes. Respectively 95% and 92% of the *Laupala* gene index and RNAseq sequences mapped to the assembled genome. In addition, the BUSCO search using either database found most genes in the *Laupala* assembly (Table 1).

Collinear linkage maps

We obtained 815,109,126; 522,378,849; and 311,558,401 reads after demultiplexing for ParKoh, KonPar, and PruKoh, respectively. Average sequencing depth ± standard deviation across all individuals in the F2 mapping population after filtering, (before and) after marker selection based on segregation distortion and ancestry information for linkage mapping was (62.4 x ± 162.5) 52.2 x ± 31.4, (44.3 x ± 58.5) 38.1 x ± 23.8, and (56.1 x ± 105.7) 41.8 x ± 29.3, respectively.

In the initial maps, 158 (ParKoh), 170 (KonPar), and 138 (PruKoh) markers were grouped into eight linkage groups at a LOD score of 5.0, corresponding to the seven autosomes and one X-chromosome in *Laupala.* Markers group in 8 linkage groups for a range of LOD scores: 5.0 < LOD < 15.0 for ParKoh (no markers unlinked), 5.0 < LOD < 42.0 for KonPar (no markers unlinked), and 4.0 < LOD < 14.0 for PruKoh (between 4 and 7 markers are unlinked depending on the LOD threshold). The corresponding marker spacing was 5.14, 4.85, and 7.33 cM. The comprehensive maps contained 526, 650, and 325 markers with an average marker spacing of 1.91, 1.37, and 3.25 cM and on the dense maps we placed 608, 823, and 383 markers with on average 1.69, 1.37, and 3.25 cM between markers

The recent divergence times and the limited levels of post-zygotic isolation observed in this system led us to hypothesize that the linkage maps would show a high degree of collinearity. The visual comparison of marker positioning showed that the relative locations of shared scaffolds were similar across the linkage maps in both the initial and the comprehensive maps (Fig 2, Fig S1). However, we also observe substantial variation in the total genetic length of homologous linkage groups, indicating recombination rate variation (Fig 2, Fig S1). This variation may in part result from chromosomal rearrangements. However, we can only reliably detect rearrangements in our maps if they are alternatively fixed between *L. pruna* and *L. kohalensis* on the one side and *L. paranigra* and *L. kona* on the other side. In that specific scenario the inverted

409 marker order is visible when contrasting the PruKoh and KonPar maps, while the ParKoh map
410 would show reduced recombination in that area (Fig 1C). Despite the apparent variation in
411 recombination rates among homologous linkage groups, Spearman's rank correlation of pairwise
412 linkage group comparisons was high (ρ varied between 0.91 and 1.00) and similar to values seen
413 in comparisons of intraspecific linkage maps (e.g. Poursarebani *et al.* 2013); the quantitative
414 measure of collinearity was largely consistent across linkage groups and across cross types
415 (Table 2). Finally, merging the maps into a consensus pseudomolecule assembly allowed us to
416 measure the error between individual maps and the merged assembly. Correlations between
417 linkage maps and the pseudomolecule assembly were generally high (> 0.95), indicating
418 substantial synteny (Fig S2).

419 *Limited heterogeneity in segregation distortion*

420 We expected genetic incompatibilities to be more likely to occur in the ParKoh cross than in the
421 KonPar and PruKoh cross, because *L. kohalensis* and *L. paranigra* are more distantly related
422 than any of the other species pairs (Fig 1). We tested this hypothesis by examining the degree of
423 segregation distortion in markers within 10 cM sliding windows across the linkage maps.
424 Overall, segregation distortion was limited and average genotype frequencies were close to
425 Mendelian expectations (Fig 3). However, LG3 showed a bias against *L. kohalensis*
426 homozygotes in the ParKoh cross but not in any of the other crosses. Additionally, there was
427 significant variation in the frequency of heterozygotes across the linkage groups (linear model
428 Freq[heterozygotes]~LG x cross: $R^2 = 0.21$, $F_{20,1547} = 20.7$, $P < 0.0001$). *Post-hoc* Tukey Honest
429 Significant Differences revealed that linkage group 7 had the lowest abundance of heterozygotes
430 overall and within each of the intercrosses and that levels of heterozygosity on LG 7 were similar
431 across the maps (Table S2). Together, these results show that from some LGs and in some

crosses, certain genotype combinations were less common than expected, potentially as a result for genetic incompatibilities or meiotic drive.

Variable recombination rates across the genome

We anchored a total of 1,054 scaffolds covering 720 million base pairs, a little below half the current genome assembly. The pseudomolecule length ranges from 25,387,579 bp (LG6) to 137,279,277 bp (LG3), with N50 values between 569,059 bp (LG6) and 1,249,941 bp (LGX; see Table S3 for scaffold number, N50, and assembly size per LG and Fig S3 for coverage variation across the linkage groups). This gives us enough power to make inferences about broad-scale recombination rate variation, but not about the existence of small-scale recombination hotspots. Average recombination rates (cM/Mb) varied from between 0.75 (KonPar) and 0.93 (ParKoh) on the X chromosome to between 3.12 (KonPar) and 4.24 (PruKoh) on LG6 (Table 3). We note that the recombination rate for LG6 might be artificially inflated because of lower assembly quality (expressed as N50) of this LG relative to the other LGs in all linkage maps and in the pseudochromosomes (Table S3). Both including and excluding the sex chromosome, there is a significant linear relationship between chromosome size and genetic length (linear mixed effect model with cross as random variable; with X: $\beta = 0.62$, $F_{1,23} = 14.95$, $P = 0.0008$; without X: $\beta = 0.69$, $F_{1,20} = 29.7$, $P < 0.0001$) and between chromosome size and broad-scale recombination rate (with X: $\beta = -34.1$, $F_{1,23} = 29.1$, $P < 0.0001$); without X: $\beta = -24.0$, $F_{1,23} = 63.7$, $P < 0.0001$).

Most linkage groups showed wide regions of strongly reduced recombination rates in the center of the linkage groups (Fig. 4). The general pattern of peripheral peaks in recombination rates juxtaposing large recombination “desserts” was observed in all three intercrosses, but some additional cross-specific peaks in recombination rates were observed on almost all linkage groups (Fig 4).

Trait-preference co-evolution despite high recombination

Contrary to our expectation, the approximate location of the colocalizing song and preference QTL peak from (Shaw and Lesnick 2009) was associated with average recombination rates in the ParKoh and KonPar map and low recombination rates in the PruKoh map (Fig 4; Table S4). However, most other QTL peaks are located in regions of low recombination (Fig 4; Table S4).

DISCUSSION

The evolutionary trajectory of diverging populations and the likelihood of speciation can be heavily influenced by recombination. Within species, recombination can create favorable combinations of alleles or decouple deleterious from beneficial alleles. Among species, regions with low recombination can provide a genetic shield against introgression of maladaptive loci (Noor *et al.* 2001; Rieseberg 2001; Butlin 2005; Slatkin 2008; Noor and Bennett 2009; Cutter and Payseur 2013; Ortiz-Barrientos *et al.* 2016). Understanding recombination is thus critical to understanding adaptation and speciation. Recombination also has important implications for the analysis of genotype-phenotype relationships (Mackay 2001), demographic inference (Li and Durbin 2011), and analyses of genomic variation (Cutter and Payseur 2013; Wolf and Ellegren 2016). However, we still have limited insight into the patterns of recombination rate variation among species and across genomes, in particular for radiations powered largely by behavioral isolation.

Here, we study four species of sexually divergent Hawaiian swordtail crickets and generate the first pseudomolecule-level assembly for Orthoptera and the first published genome assembly for crickets, an important model system in neurobiology, behavioral ecology, and evolutionary genetics (Horch *et al.* 2017). Below, we discuss how our results provide insight into the potential

for structural variation (linkage map collinearity) and genetic incompatibilities to drive reproductive isolation among closely related *Laupala* species. We also elaborate on the patterns of variation in recombination rates across the genome. We then discuss the surprising finding that colocalizing male song and female preference QTL did not fall in a region with particularly low recombination. This is important because it challenges the hypothesis that co-evolution of traits and preferences is facilitated by locally reduced recombination between recently diverged populations.

Collinearity of Genetic Maps

Based on the recent divergence (Mendelson and Shaw 2005) and strong premating isolation of *Laupala* species in the absence of conspicuous morphological and ecological differences (Otte 1994; Shaw 1996; Mendelson and Shaw 2005), we expected limited variation in chromosome structure and few signatures of genetic incompatibility between the species. In line with these expectations, we found that linkage groups are collinear across interspecies crosses (Fig 2). This was true both for comparisons of non-independent species pairs (between ParKoh and the other two crosses), as well as for the independent contrast of the PruKoh map versus the KonPar map. We saw some instances where markers occupied regions that may have been translocated or inverted. However, these instances were rare (Fig 2, Fig S1) and recombination rates were similar among homologous linkage groups across the hybrid families (Fig 4). Moreover, the quantitative measures of collinearity, Spearman's rank correlation among maps and between maps and the pseudomolecule assembly, as well as limited segregation distortion (but see discussion of one exception below) both supported the collinearity hypothesis; if there were larger or many small inversions between the species pairs, the order of markers in the different maps would be more variable in which case significantly lower values for Spearman's rank

correlation would be expected than the uniformly high correlations (most values are between 0.95 and 0.99) we observe here.

Variation in the organization and structure of chromosomes can contribute to postzygotic reproductive isolation after speciation as well as to the speciation process directly (Noor *et al.* 2001; Rieseberg 2001). We conclude that at least for the *Laupala* group that radiated on the Big Island of Hawaii in the last 500,000 years, structural rearrangements have not played a major role in the evolution of reproductive isolation. This is because, similar to two hybridizing *Heliconius* species (Davey *et al.* 2017), we observed that chromosome-wide recombination rates are relatively conserved and large chromosomal rearrangements are absent. We hypothesize that for *Laupala* on the Big Island premating isolation combined with (partial) geographic separation (i.e. low migration rates) provides a sufficiently strong barrier to gene flow between sister species. Indeed, as has been shown in recent models of the role of inversions in speciation, genomic rearrangements can only invade and spread in diverging populations if levels of gene flow and the contribution of structural variation to isolation (by linking adaptive alleles or incompatibilities) is high relative to the strength of assortative mating (Feder *et al.* 2014; Dagilis and Kirkpatrick 2016).

We acknowledge that the power to detect rearrangements and recombination rate variation between the genomes of the species studied here is limited by the resolution of the maps, the sample size and the error rate of the sequencing platform. The average spacing of markers is between 1.37 and 3.25 cM; the resolution at which we can detect rearrangements is on the order of 10^5 and 10^6 bp. Therefore, it is difficult to attribute subtle variation in marker order and genetic distance (< 1-2 cM) between the maps to genomic rearrangements, as opposed to mapping errors and sampling variance, because variation in recombination frequencies between

closely related species might be detected only at finer scales (Stevison *et al.* 2017). However, the overall collinearity of the maps supports our conclusion that chromosomal architectures are broadly conserved between the species studied here.

Genetic incompatibilities

We expected genetic incompatibilities to be more likely between genomes of more distantly related species. Accordingly, we discovered a single region covering approximately half of linkage group 3 with high segregation distortion in ParKoh; we found no such deviations in the other two crosses (Fig 3). Inspection of genotype frequencies indicated that fewer individuals than expected were homozygous for *L. kohalensis* alleles at loci in this region. In a controlled cross, segregation distortion can be caused by prezygotic effects such as meiotic drive of selfish genetic elements and distorter genes (e.g. like *sd* in *Drosophila melanogaster* (Larracuente and Presgraves 2012)), and by postzygotic genetic incompatibilities (Dobzhansky 1937; Muller 1942; Burt and Trivers 2006; Presgraves 2010). Genotypic errors may produce superficially similar patterns but are unlikely to distort segregation over large genomic regions and with consistent bias towards the same genotypes. Although meiotic drive is a possible alternative to genetic incompatibilities, we do not see the same effect in the other cross involving *L. paranigra*, where selfish genetic elements or segregation distorters ought to have a similar effect. Overall, the large region on linkage group 3 reveals a potential local post-mating barrier to gene flow that could contribute to strengthening existing prezygotic barriers in secondary contact zones or following episodes of migration.

Recombination landscape

Chromosomal rearrangements influence genomic divergence by locally altering recombination rates within and among species. Felsenstein (1974, 1981) illuminated the role of *intraspecific* recombination in purging deleterious alleles and the role of *interspecific* recombination in decoupling co-adapted alleles. In recent years, the role of recombination and its interaction with divergent selection and adaptation on genomic scales have received considerable attention (e.g. Yeaman and Whitlock 2011; Feder *et al.* 2012; Samuk *et al.* 2017) and technological advances are shifting focus towards characterizing recombination rates across genomes (Butlin 2005; Slatkin 2008; Noor and Bennett 2009; Barb *et al.* 2014; Burri *et al.* 2015). Comparisons of genomic recombination landscapes illuminate the role of recombination in speciation, but focus has been on systems where environmental and ecological pressures are driving divergence between locally adapted populations.

Here, we show for species separated by conspicuous sexual barriers that there is limited variation in recombination rates across the maps of three interspecific crosses (Fig 4), but strong heterogeneity in recombination rates across the genome. Genome-wide average interspecific recombination rate varied between 1.3 and 1.5 cM/Mb (Table 3), similar to intraspecific rates observed in dipterans and substantially lower than social hymenopterans and lepidopterans (Wilfert *et al.* 2007). We note that our estimates are derived from interspecific maps, which may lead to somewhat lower estimates compared to intraspecific maps (e.g. Beukeboom *et al.* 2010), because genetic incompatibilities and rearrangements may reduce rates of crossing over; however, differences between intra and interspecific recombination might be negligible if rearrangements are rare (e.g., Davey *et al.* 2017). Moreover, we anchored about 50% of the nucleotides in the draft assembly to linkage groups, and there remain many scaffolds not mapped to a genomic position. These ‘missing’ scaffolds are expected to add to the physical length of the

chromosomes more so than to the genetic length of the chromosomes, thus lowering the overall recombination rate and increasing heterogeneity in recombination rates among the crosses. However, our study emphasized relative patterns of recombination, which should not be affected by our sampling. And while we can only approximate intraspecific recombination rates at this point, we note that recent divergence of the study species and collinearity of the linkage maps support conservation of recombination landscapes across intraspecific and interspecific comparisons.

Interestingly, for all three species pairs we document high variability in interspecific recombination across genomic regions. We found a stereotypical U-shaped pattern: large regions of low recombination in the center, with recombination rates well below 1 cM/Mb and occasionally approaching zero, flanked by steep inclines reaching rates up to 6 cM/Mb (Fig 4). This pattern is consistent with earlier findings in plants (Anderson *et al.* 2003), invertebrates (Rockman and Kruglyak 2009; Niehuis *et al.* 2010), and vertebrates (Backström *et al.* 2010; Roesti *et al.* 2013; Singhal *et al.* 2015) including humans (Kong *et al.* 2002), but differs from observations in, for example, *Drosophila* (Kulathinal *et al.* 2008), which show heterogeneity in recombination rates, but not necessarily much higher rates on the periphery of the chromosomes. Commonly invoked drivers of local recombination suppression, such as selection against recombination due to negative epistasis or the maintenance of linkage disequilibrium between mutually beneficial alleles (Smukowski and Noor 2011; Stevison *et al.* 2011; Smukowski Heil *et al.* 2015; Ortiz-Barrientos *et al.* 2016), are not likely to leave chromosome wide signatures. Rather, the observed U-shaped pattern is more likely attributable to structural properties of chromosomes, such as the location of the centromere and heterochromatin-rich regions (Copenhaver *et al.* 1999; Haupt *et al.* 2001). Roesti *et al*. (2013) observe similar recombination

landscapes in stickleback and suggest it might be due to peripheral clustering during meiosis prophase I to facilitate homolog pairing (Harper *et al.* 2004; Brown *et al.* 2005). Regardless of the mechanism, the observed genomic architecture will drive substantial heterogeneity in the propensity of favorable and/or maladaptive alleles to come together, break apart, and introgress in heterospecific backgrounds.

Trait-Preference Co-evolution

Recombination heterogeneity may be important in this study system as a facilitator of trait-preference co-evolution. If trait and preference genes are coupled through physical linkage (Kirkpatrick and Hall 2004), linkage can be stronger and span wider physical distances in regions with reduced recombination. We hypothesized that recombination facilitates linkage between trait and preference genes in *Laupala* because a previous study showed that a major song QTL (~ 9% of the parental difference in male song) co-localizes with a preference QTL (~14% of parental difference for female preference) in a cross between *L. kohalensis* and *L. paranigra* (Shaw and Lesnick 2009). Contrary to our expectation, we show that the co-localizing QTL fall in a region with intermediate to high recombination rates (> 2.0 cM/Mb) compared to chromosomal averages (typically 1 - 2 cM/Mb). This suggests that reduced recombination over larger physical distances is unlikely to facilitate trait-preference co-evolution in this system. Importantly, a high speciation rate and wide-spread divergence in sexual signaling phenotypes suggest a primary role for trait-preference co-evolution in *Laupala* speciation (Mendelson and Shaw 2005; Shaw *et al.* 2011). Additionally, although these species likely diverged in allopatry (Mendelson and Shaw 2005), some level of interspecific gene exchange is likely given historical biogeography, widespread secondary contact and evidence derived from discordant nuclear and mitochondrial gene trees (Shaw 2002).

How then is linkage disequilibrium between traits and preferences maintained? First, QTL may co-localize due to very tight physical linkage or pleiotropy instead of looser linkage. Under pleiotropy, a lack of physical space for cross-overs to occur rather than locally low recombination rates maintains linkage disequilibrium. Linkage disequilibrium might also persist in the face of recombination if strong assortative mating results from female mate preference. In this case, genetic correlations in signal and preference variation will evolve, independent of their genomic locations (Fisher 1930; Lande 1981). Concordantly, recent simulation studies showed that the probability with which recombination rate modifiers that link co-adaptive alleles spread in a population is lower when assortative mating is strong, recombination between loci is low, and selection on the loci themselves is strong (Feder *et al.* 2014; Dagilis and Kirkpatrick 2016). Third, the current test involves only a single locus and additional tests are required to more robustly examine the relationship between recombination and trait-preference co-evolution. We observed that several known male song QTL on other linkage groups fall in regions of low recombination. Additional female preference QTL covary with these song QTL as well (Wiley *et al.* 2012) although precise map locations are not yet known.

In summary, we find limited variation in chromosome structure among species, but strong heterogeneity in the recombination landscape across the genome. We present a *de novo* genome assembly for *L. kohalensis* and anchor a substantial part of the genome to pseudomolecules. Crickets are an important model system for evolutionary and neurobiological research (Horch *et al.* 2017), but limited genomic resources are available. The first Orthopteran pseudomolecule-level draft genome and recombination rate map are thus important new contributions to future speciation genomics research. This study further provides important insight into the extent to which structural variation and genetic incompatibilities contribute to reproductive barriers

between closely related, sexually divergent species. We also shine light on the role of recombination in trait-preference co-evolution and argue, at least in *Laupala*, that the evolution of behavioral isolation is not contingent on structural genomic variation and locally reduced recombination.

ACKNOWLEDGEMENTS

We thank Stephen Chenoweth and two anonymous reviewers for helpful comments that strongly improved the quality of this manuscript. We further thank the Shaw lab, in particular Mingzi Xu, as well as Michael Sheehan and other members from Cornell's Neurobiology and Behavior department for input that contributed to the interpretation of the findings. This work was supported by the National Science Foundation (DEB 1241060, IOS 1257682 and IOS 0843528).

REFERENCES

Anderson L. K., Doyle G. G., Brigham B., Carter J., Hooker K. D., *et al.*, 2003 High-resolution crossover maps for each bivalent of Zea mays using recombination nodules. Genetics 165: 849–865.

Andersson M., Simmons L. W., 2006 Sexual selection and mate choice. Trends Ecol. Evol. 21: 296–302.

Arnegard M. E., Kondrashov A. S., 2004 Sympatric speciation by sexual selection alone is unlikely. Evolution. 58: 222–237.

Aronesty E., 2011 ea-utils: Command-line tools for processing biological sequencing data. Expr. Anal. Durham, NC.

Auwera G. A. Van der, Carneiro M. O., Hartl C., Poplin R., Angel G. del, *et al.*, 2013 From

FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinforma. 43: 1–11.

Backström N., Forstmeier W., Schielzeth H., Mellenius H., Nam K., *et al.*, 2010 The recombination landscape of the zebra finch *Taeniopygia guttata* genome. Genome Res. 20: 485–95.

Barb J. G., Bowers J. E., Renaut S., Rey J. I., Knapp S. J., *et al.*, 2014 Chromosomal Evolution and Patterns of Introgression in *Helianthus*. Genetics 197: 969–979.

Benjamini Y., Hochberg Y., 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57: 289–300.

Beukeboom L. W., Niehuis O., Pannebakker B. A., Koevoets T., Gibson J. D., *et al.*, 2010 A comparison of recombination frequencies in intraspecific versus interspecific mapping populations of *Nasonia*. Heredity. 104: 302–309.

Blankers T., Oh K. P., Shaw K. L., 2018 The genetics of behavioral isolation in an island system. bioRxiv. https://doi.org/10.1101/250852

Broman K. W., Wu H., Sen Ś., Churchill G. A., 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Brown P. W., Judis L., Chan E. R., Schwartz S., Seftel A., *et al.*, 2005 Meiotic Synapsis Proceeds from a Limited Number of Subtelomeric Sites in the Human Male. Am. J. Hum. Genet. 77: 556–566.

Bürger R., Akerman A., 2011 The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. Theor. Popul. Biol. 80: 272–288.

678 Burri R., Nater A., Kawakami T., Mugal C. F., Olason P. I., *et al.*, 2015 Linked selection and
679 recombination rate variation drive the evolution of the genomic landscape of differentiation
680 across the speciation continuum of Ficedula flycatchers. Genome Res. 25: 1656–1665.

681 Burt A., Trivers R., 2006 *Genes in Conflict: The Biology of Selfish Genetic Elements.* Belknap,
682 Cambridge, MA.

683 Butlin R. K., 2005 Recombination and speciation. Mol. Ecol. 14: 2621–2635.

684 Copenhaver G. P., Nickel K., Kuromori T., Benito M.-I., Kaul S., *et al.*, 1999 Genetic definition
685 and sequence analysis of *Arabidopsis* centromeres. Science. 286: 2468–2474.

686 Coyne J. A., Orr H. A., 2004 *Speciation*. Sinauer, Sunderland, MA.

687 Cutter A. D., Payseur B. A., 2013 Genomic signatures of selection at linked sites: unifying the
688 disparity among species. Nat. Rev. Genet. 14: 262–274.

689 Dagilis A. J., Kirkpatrick M., 2016 Prezygotic isolation, mating preferences, and the evolution of
690 chromosomal inversions. Evolution 70: 1465–1472.

691 Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., *et al.*, 2011 The variant call format
692 and VCFtools. Bioinformatics 27: 2156–2158.

693 Danley P. D., Mullen S. P., Liu F., Nene V., Quackenbush J., *et al.*, 2007 A cricket Gene Index:
694 a genomic resource for studying neurobiology, speciation, and molecular evolution. BMC
695 Genomics 8: 109.

696 Davey J. W., Barker S. L., Rastas P. M., Pinharanda A., Martin S. H., *et al.*, 2017 No evidence
697 for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions.
698 Evol. Lett. 1: 138–154.

DePristo M. A., Banks E., Poplin R., Garimella K. V, Maguire J. R., *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–498.

Dobzhansky T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York, NY.

Dodt M., Roehr J. T., Ahmed R., Dieterich C., 2012 FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. Biology. 1: 895–905.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 6.

Feder J. L., Egan S. P., Nosil P., 2012 The genomics of speciation-with-gene-flow. Trends Genet. 28: 342–350.

Feder J. L., Nosil P., Flaxman S. M., 2014 Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. Front. Genet. 5: 295.

Felsenstein J., 1981 Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution. 35: 124–138.

Fisher R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, New York.

Garrison E., Marth G., 2012 Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv1207.3907.

Gavrilets S., 2003 Perspective: models of speciation: what have we learned in 40 years? Evolution 57: 2197–2215.

Gillespie J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model. Genetics 155: 909–919.

Grace J. L., Shaw K. L., 2011 Coevolution of male mating signal and female preference during early lineage divergence of the Hawaiian cricket, *Laupala Cerasina*. Evolution. 65: 2184–2196.

Harper L., Golubovskaya I., Cande W. Z., 2004 A bouquet of chromosomes. J. Cell Sci. 117: 4025 LP-4032.

Haupt W., Fischer T. C., Winderl S., Fransz P., Torres-Ruiz R. A., 2001 The centromere1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. Plant J. 27: 285–296.

Hill W. G., Robertson A., 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Horch H. W., Mito T., Popadic A., Ohuchi H., Noji S. (Eds.), 2017 *The cricket as a model organism*. Springer Japan, Tokyo, Japan.

Kent W. J., 2002 BLAT—the BLAST-like alignment tool. Genome Res. 12: 656–664.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., *et al.*, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14: R36.

Kirkpatrick M., 1982 Sexual selection and the evolution of female mate choice. Evolution. 36: 1–12.

Kirkpatrick M., Ravigne V., 2002 Speciation by natural and sexual selection: models and

741 experiments. Am. Nat. 159 supple: S22–S35.

742 Kirkpatrick M., Hall D. W., 2004 Sexual selection and sex linkage. Evolution. 58: 683–691.

743 Kirkpatrick M., Barton N., 2006 Chromosome inversions, local adaptation and speciation.
744 Genetics 173: 419–434.

745 Kong A., Gudbjartsson D. F., Sainz J., Jonsdottir G. M., Gudjonsson S. A., *et al.*, 2002 A high-
746 resolution recombination map of the human genome. Nat. Genet. 31: 241.

747 Kulathinal R. J., Bennett S. M., Fitzpatrick C. L., Noor M. A. F., 2008 Fine-scale mapping of
748 recombination rate in Drosophila refines its correlation to diversity and divergence. Proc.
749 Natl. Acad. Sci. 105: 10051–10056.

750 Lande R., 1981 Models of speciation by sexual selection on polygenic traits. Proc Natl Acad Sci
751 78: 3721–3725.

752 Lander E., Green P., Abrahamson J., Barlow A., Daly M., *et al.*, 1987 MAPMAKER: An
753 interactive computer package for constructing primary genetic linkage maps of
754 experimental and natural populations. Genomics 1: 174–181.

755 Langmead B., Salzberg S. L., 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:
756 357–359.

757 Larracuente A. M., Presgraves D. C., 2012 The Selfish Segregation Distorter Gene Complex of
758 *Drosophila melanogaster*. Genetics 192: 33–53.

759 Li H., Durbin R., 2011 Inference of human population history from individual whole-genome
760 sequences. Nature 475: 493–496.

761 Lincoln S. E., Daly M. J., Lander E. S., 1993 Constructing genetic linkage maps with

762 MAPMAKER/EXP Version 3.0: a tutorial and reference manual. A Whitehead Inst.
763 Biomed. Res. Tech. Rep.: 78–79.

764 Liu Y., Schröder J., Schmidt B., 2013 Musket: a multistage k-mer spectrum-based error corrector
765 for Illumina sequence data. Bioinformatics 29: 308–315.

766 Luo R., Liu B., Xie Y., Li Z., Huang W., *et al.*, 2012 SOAPdenovo2: an empirically improved
767 memory-efficient short-read de novo assembler. Gigascience 1: 18.

768 Mackay T. F. C., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:
769 303–339.

770 Marques D. A., Lucek K., Meier J. I., Mwaiko S., Wagner C. E., *et al.*, 2016 Genomics of rapid
771 incipient speciation in sympatric threespine stickleback. PLoS Genet. 12: 1–34.

772 Mendelson T. C., Shaw K. L., 2002 Genetic and behavioral components of the cryptic species
773 boundary between Laupala cerasina and L. kohalensis (Orthoptera: Gryllidae). Genetica
774 116: 301–310.

775 Mendelson T. C., Shaw K. L., 2005 Rapid speciation in an arthropod. Nature 433: 375–376.

776 Muller H., 1942 Isolating mechanisms, evolution, and temperature. Biol. Symp. 6: 71–125.

777 Myers S., Bottolo L., Freeman C., McVean G., Donnelly P., 2005 A fine-scale map of
778 recombination rates and hotspots across the human genome. Science. 310: 321–324.

779 Niehuis O., Gibson J. D., Rosenberg M. S., Pannebakker B. A., Koevoets T., *et al.*, 2010
780 Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*.
781 PLoS One 5: e8597.

782 Noor M. A., Grams K. L., Bertucci L. A., Reiland J., 2001 Chromosomal inversions and the

783 reproductive isolation of species. Proc. Natl. Acad. Sci. 98: 12084–8.

784 Noor M. A. F., Bennett S. M., 2009 Islands of speciation or mirages in the desert? Examining the
785 role of restricted recombination in maintaining species. Heredity. 103: 439–44.

786 Ooijen J. W. van, 2006 JoinMap 4, Software for the calculation of genetic linkage maps in
787 experimental populations.

788 Ortiz-Barrientos D., Engelstädter J., Rieseberg L. H., 2016 Recombination rate evolution and the
789 origin of species. Trends Ecol. Evol. 31: 226–236.

790 Otte D., 1994 *The Crickets of Hawaii: Origin, Systematics, and Evolution*. Orthoptera
791 Society/Academy of Natural Sciences of Philadelphia, Philadelphia, PA.

792 Otto S. P., 2009 The evolutionary enigma of sex. Am. Nat. 174: S1--S14.

793 Petrov D. A., Sangster T. A., Johnston J. S., Hartl D. L., Shaw K. L., 2000 Evidence for DNA
794 loss as a determinant of genome size. Science. 287: 1060–1062.

795 Poursarebani N., Ariyadasa R., Zhou R., Schulte D., Steuernagel B., *et al.*, 2013 Conserved
796 synteny-based anchoring of the barley genome physical map. Funct. Integr. Genomics 13:
797 339–350.

798 Presgraves D. C., 2010 Darwin and the origin of interspecific genetic incompatibilities. Am. Nat.
799 176: S45–S60.

800 R Development Core Team R., 2016 R: A language and environment for statistical computing
801 (RDC Team, Ed.). R Found. Stat. Comput. 1: 409.

802 Rieseberg L. H., 2001 Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16: 351–
803 357.

Rockman M. V., Kruglyak L., 2009 Recombinational landscape and population genomics of *Caenorhabditis elegans*. PLoS Genet 5: e1000419.

Roesti M., Moser D., Berner D., 2013 Recombination in the threespine stickleback genome - Patterns and consequences. Mol. Ecol. 22: 3014–3027.

Samuk K., Owens G. L., Delmore K. E., Miller S. E., Rennison D. J., *et al.*, 2017 Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Mol. Ecol. 26: 4378–4390.

Schmieder R., Edwards R., 2011 Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.

Servedio M. R., 2009 The role of linkage disequilibrium in the evolution of premating isolation. Heredity. 102: 51–56.

Servedio M. R., Burger R., 2014 The counterintuitive role of sexual selection in species maintenance and speciation. Proc. Natl. Acad. Sci. 111: 8113–8118.

Servedio M. R., 2015 Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. Evol. Appl. 9: 91–102.

Shaw K. L., 1996 Polygenic inheritance of a behavioral phenotype: interspecific genetics of song in the Hawaiian cricket genus *Laupala*. Evolution. 50: 256–266.

Shaw K. L., 2000a Further acoustic diversity in Hawaiian forests: two new species of Hawaiian cricket (Orthoptera: Gryllidae: Trigonidiinae: *Laupala*). Zool. J. Linn. Soc. 129: 73–91.

Shaw K. L., 2000b Interspecific genetics of mate recognition: inheritance of female acoustic preference in Hawaiian crickets. Evolution 54: 1303–1312.

825 Shaw K. L., 2002 Conflict between nuclear and mitochondrial DNA phylogenies of a recent
826 species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian
827 crickets. Proc. Natl. Acad. Sci. 99: 16122–16127.

828 Shaw K. L., Parsons Y. M., Lesnick S. C., 2007 QTL analysis of a rapidly evolving speciation
829 phenotype in the Hawaiian cricket *Laupala*. Mol. Ecol. 16: 2879–2892.

830 Shaw K. L., Lesnick S. C., 2009 Genomic linkage of male song and female acoustic preference
831 QTL underlying a rapid species radiation. Proc. Natl. Acad. Sci. 106: 9737–9742.

832 Shaw K. L., Ellison C. K., Oh K. P., Wiley C., 2011 Pleiotropy, "sexy" traits, and speciation.
833 Behav. Ecol. 22: 1154–1155.

834 Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V, Zdobnov E. M., 2015 BUSCO:
835 assessing genome assembly and annotation completeness with single-copy orthologs.
836 Bioinformatics 31: 3210.

837 Singhal S., Leffler E. M., Sannareddy K., Turner I., Venn O., *et al.*, 2015 Stable recombination
838 hotspots in birds. Science. 350: 928–932.

839 Slatkin M., 2008 Linkage disequilibrium—understanding the evolutionary past and mapping the
840 medical future. Nat. Rev. Genet. 9: 477–485.

841 Smith J. M., Haigh J., 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

842 Smith J. M., 1978 *The Evolution of Sex*. Cambridge University Press Cambridge.

843 Smukowski C. S., Noor M. A. F., 2011 Recombination rate variation in closely related species.
844 Heredity. 107: 496–508.

845 Smukowski Heil C. S., Ellison C., Dubin M., Noor M. A. F., 2015 Recombining without

Hotspots: A Comprehensive Evolutionary Portrait of Recombination in Two Closely Related Species of *Drosophila*. Genome Biol. Evol. 7: 2829–42.

Stevison L. S., Hoehn K. B., Noor M. A. F., 2011 Effects of Inversions on Within- and Between-Species Recombination and Divergence. Genome Biol. Evol. 3: 830.

Stevison L. S., Sefick S., Rushton C., Graze R. M., 2017 Recombination rate plasticity: revealing mechanisms by design. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372: 20160459.

Tang H., Zhang X., Miao C., Zhang J., Ming R., *et al.*, 2015 ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 16: 3.

Voorrips R. E., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93: 77–78.

Wiley C., Ellison C. K., Shaw K. L., 2012 Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. Proc. R. Soc. B Biol. Sci. 279: 1203–1209.

Wilfert L., Gadau J., Schmid-Hempel P., 2007 Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity. 98: 189–197.

Wolf J. B. W., Ellegren H., 2016 Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. 18: 87–100.

Yeaman S., Whitlock M. C., 2011 The genetic architecture of adaptation under migration-selection balance. Evolution. 65: 1897–1911.

Yeaman S., 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc. Natl. Acad. Sci. 110: E1743--E1751.

FIGURE LEGENDS

Figure 1. Study design. (A) The phylogenetic relationships of studied *Laupala* species based on a neighbor joining tree generated from genetic distances among the parental lines used in this study. Dashed grey lines connect species pairs that were crossed. (B) Approximate distributions of the studied species on the Big Island of Hawaii (C) Hypothetical segregation and linkage map construction for five genetic loci A, B, C, D, and E in three crosses of fours species. The genetic distance between the loci is 5 centi-Morgan (cM) in each of the four species. Loci [B,C,D] are inverted in the green and black species. When two species that have alternative karyotypes for the inversion are crossed (pair 2), loci in the inversion will not recombine in the first-generation hybrid, resulting in reduced genetic (map) length in the second-generation hybrid. Other chromosomal rearrangements will have similar effects. Only if two crosses involve homokaryotypic species pairs that have alternative karyotypes can an inversion be detected in a comparison of intercross linkage maps.

Figure 2. Initial linkage maps. Bars represent linkage groups (LG) for ParKoh, KonPar, and PruKoh. Lines within the bars indicate marker positions. The scale on the left measures marker spacing in cM. Blue lines connect markers on the same scaffold between the different maps. The map for ParKoh is shown twice to facilitate comparison across all three maps. See Fig S1 for comprehensive maps.

Figure 3. Segregation distortion. For each of the seven autosomal linkage groups within the three comprehensive linkage maps (from top to bottom: ParKoh, KonPar, PruKoh), a sliding window of the negative log-transformed P-values for the $\chi 2$-square test for deviation from a 1:2:1 segregation ratio is shown across markers with black lines in the top panels. In the panel below, the trace of the frequency of heterozygote genotypes (blue lines) and homozygote genotypes for

890 both parental alleles (black and red lines, respectively) is shown. For each intercross, dashed
891 grey lines indicate P = 0.01 (top panels) or expected allele frequencies based on 1:2:1 inheritance
892 (bottom panels).
893 Figure 4. Recombination and Marey maps. Gray-scale symbols and lines indicate the relationship
894 between the physical distance (scaffold midposition) in million base pairs on the x-axis and the
895 genetic distance in cM for each of the 8 linkage groups on the left y-axis. Open dots represent the
896 dense ParKoh linkage map, triangles and diamonds that of the KonPar and PruKoh cross,
897 respectively. The corresponding lines (ParKoh: solid, KonPar: dashed, PruKoh: dotted) indicate
898 the fitted smoothing spline (10 degrees of freedom). The red lines (same stroke style) show the
899 first order derivative of the fitted splines and represent the variation in recombination rate as a
900 function of physical distance (cM per Mb, on the right y-axis). Grey bars indicate the
901 approximate location of male song rhythm QTL peaks. The yellow star in the LG1 panel
902 highlights the QTL peak that co-localizes with a female preference QTL peak (Shaw & Lesnick
903 2009).
904

Table 1. *Laupala kohalensis* sequencing, assembly, and gene space evaluation statistics.

Sequencing Statistics	**Raw data**		**Processed data**	
Library	**Size (Gb)**	**Coverage[a]**	**Size (Gb)**	**Coverage[a]**
Paired End 0.2 Kb inserts	28.9	15	26.1	14
Paired End 0.5 Kb inserts	63.1	33	59.8	31
Mate Pair 2 Kb inserts	36.2	19	31.8	17
Mate Pair 5 Kb inserts	34.3	18	27.8	14
Total	162.5	85	145.5	76

Assembly Statistics	**Contigs**	**Scaffolds**
Total assembly size (Gb)	1.6	1.6
Total assembled sequences	219,073	148,874
Longest sequence length (Kb)	465	4,541
Average sequence length (Kb)	7.2	10.7
N90 index[b]	40,926	3,505
N90 length (Kb)	7.7	67.7
N50 index	9,917	756
N50 length (Kb)	43.6	583
GC content (%)	34.9%	34.9%

Gene Space Statistics	**Mapping percentage**
Laupala unigenes from the Gene Index	95%
Laupala RNASeq reads	92%

BUSCO database	Complete	Single copy	Duplicated	Fragmented	Missing	Total
Eukaryota_odb9	98.7%	93.7%	5.0%	0.3%	1.0%	303
Arthropoda_odb9	99.3%	96.8%	2.5%	0.1%	0.6%	1066

[a]Coverage is based on an estimated genome size of 1.91 Gb (Petrov *et al.* 2000)

[b]When ordering all contigs (or scaffolds) by size, the N50 or N90 index indicates the number of the longest sequences (contigs or scaffolds) that contain 50% or 90% of the total assembled sequence, respectively. The N50 and N90 length indicate the length of the shortest sequence in the set of the largest contigs (or scaffolds) that contain 50% or 90% of all the sequence in the assembly, respectively.

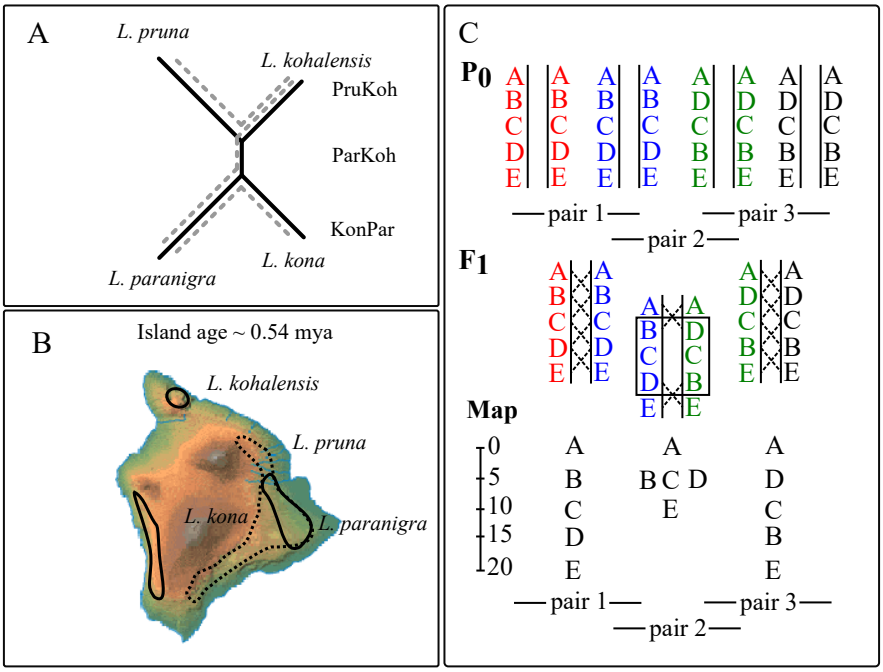
Table 2. Linkage map comparison. Spearman's rank correlation (ρ) is shown for each pairwise comparison of linkage maps across all 8 linkage groups.

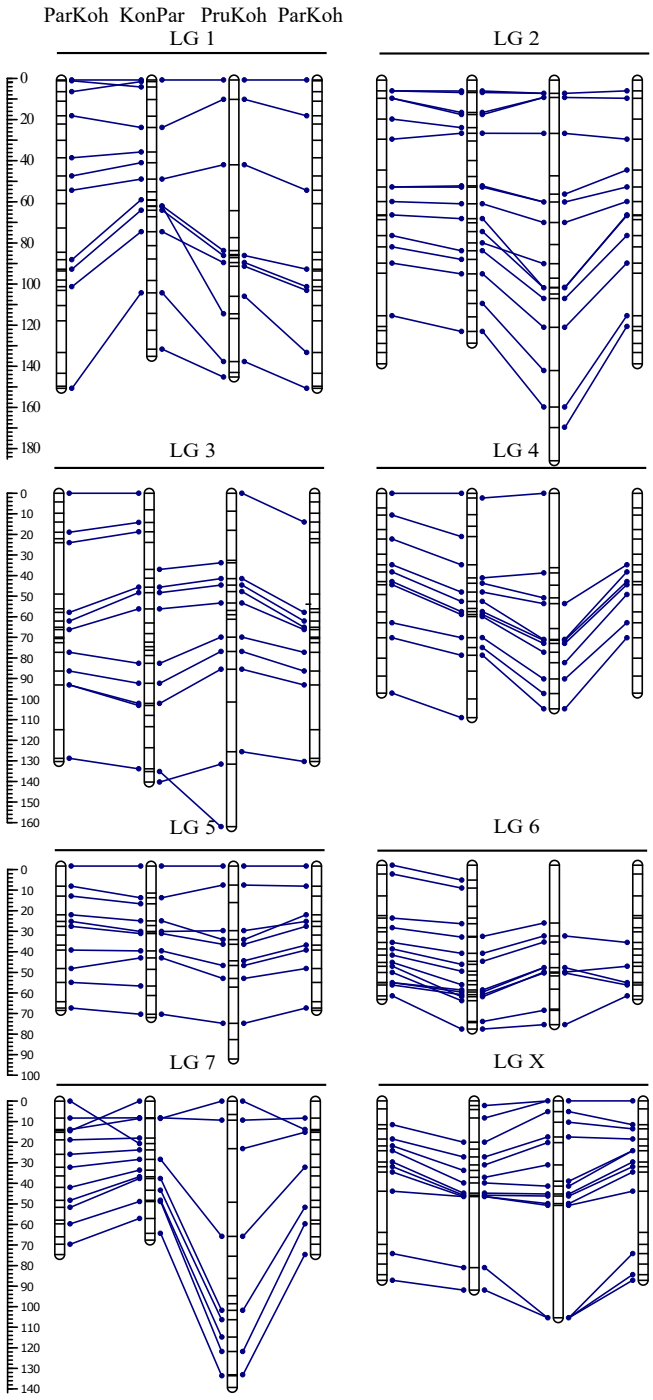
	ParKoh ~ KonPar	**ParKoh ~ PruKoh**	**KonPar ~PruKoh**
1	0.99†	0.90†	0.97†
2	0.99†	0.96†	0.93†
3	1.00†	0.98†	0.95†
4	0.99†	1.00†	0.97†
5	0.97†	0.98†	0.95†
6	0.99†	0.94†	0.94†
7	0.96†	0.93*	0.91†
X	0.92†	0.96†	0.99†

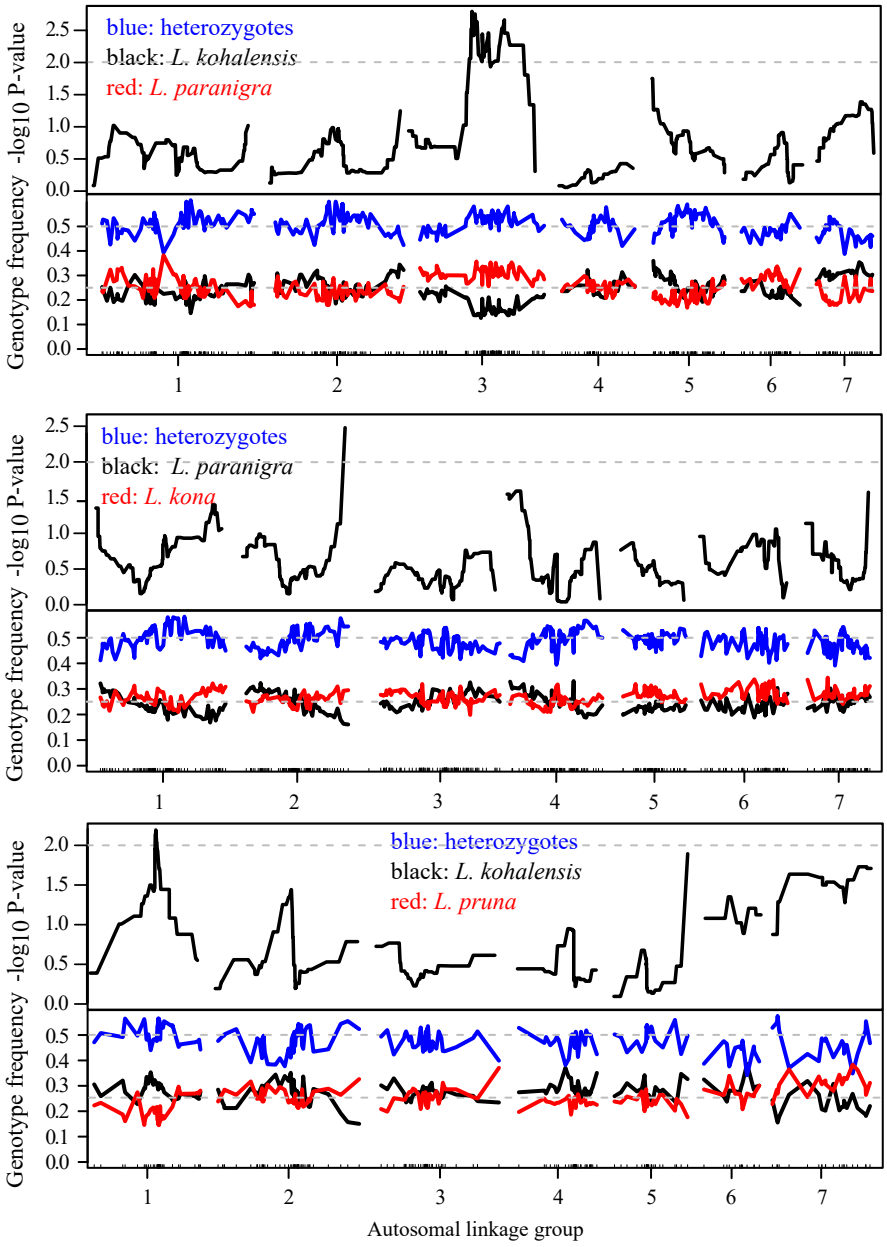
*$P < 0.001$; † $P < 0.0001$

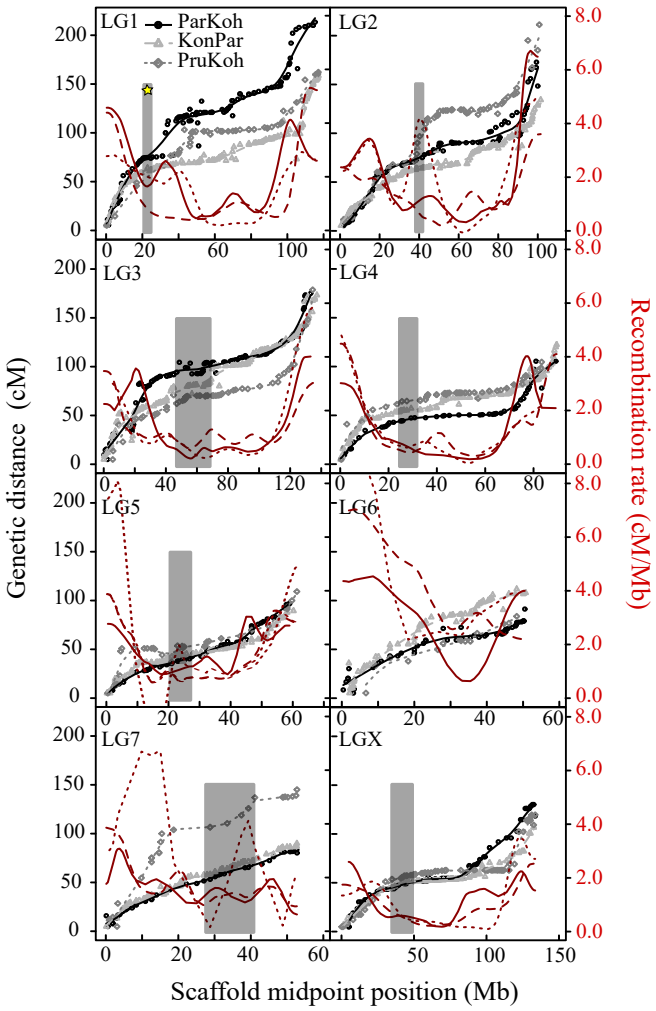
Table 3. Linkage map summary statistics.

LG	Length (Mb)	ParKoh		KonPar		PruKoh	
		Map length (cM)	Rec. Rate (cM/Mb)	Map length (cM)	Rec. Rate (cM/Mb)	Map length (cM)	Rec. Rate (cM/Mb)
1	117	207	1.77	156	1.33	156	1.33
2	102	167	1.64	128	1.25	205	2.01
3	137	169	1.23	167	1.22	173	1.26
4	90	100	1.11	117	1.30	99	1.10
5	62	91	1.47	84	1.35	103	1.66
6	25	85	3.40	106	4.24	78	3.12
7	53	78	1.47	84	1.58	139	2.62
X	134	124	0.93	101	0.75	114	0.85
Total	**720**	**1,021**	**1.42**	**943**	**1.31**	**1,067**	**1.48**


A
L. pruna
L. kohalensis
PruKoh
ParKoh
KonPar
L. kona
L. paranigra
B
Island age ~ 0.54 mya
L. kohalensis
L. pruna
L. kona
L. paranigra
C
P0
pair 1
pair 2
pair 3
F1
Map
0
5
10
15
20
pair 1
pair 2
pair 3


ParKoh KonPar PruKoh ParKoh
LG 1
LG 2
LG 3
LG 4
LG 5
LG 6
LG 7
LG X


blue: heterozygotes
black: L. kohalensis
red: L. paranigra
Genotype frequency -log10 P-value
blue: heterozygotes
black: L. paranigra
red: L. kona
Genotype frequency -log10 P-value
blue: heterozygotes
black: L. kohalensis
red: L. pruna
Genotype frequency -log10 P-value
Autosomal linkage group


LG1
ParKoh
KonPar
PruKoh
LG2
LG3
LG4
LG5
LG6
LG7
LGX
Genetic distance (cM)
Recombination rate (cM/Mb)
Scaffold midpoint position (Mb)