1   **THE GENOMIC ARCHITECTURE OF A RAPID ISLAND RADIATION: MAPPING**

2   **CHROMOSOMAL REARRANGEMENTS AND RECOMBINATION RATE**

3   **VARIATION IN *LAUPALA***

4

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

6

7   *[1] Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA*

8   *[2] Department of Horticulture, Virginia Tech, Blacksburg, VA, USA*

9

10

11

12

13 **Running Title:** *Laupala* Genomic Architecture

14

15 **Key words:** co-adaptation, recombination rate, linkage map, genome assembly, crickets

16

17

18 **Corresponding author:**

19 Thomas Blankers, Department of Neurobiology and Behavior, Cornell University, W319 215

20 Tower rd, 14850, Ithaca, NY, USA. thomasblankers@gmail.com. 1- (607) 254 4326

21

22

ABSTRACT

Speciation depends on the (local) reduction of recombination between the genomes of partially

isolated, diverging populations. Chromosomal rearrangements and generic molecular

mechanisms in the genome affect the recombination rate thereby constraining the efficacy of

(linked) selection. This can have profound impacts on trait divergence, in particular on complex

phenotypes associated with speciation by sexual selection. Because of the obligate co-evolution

of traits and preferences in sexual communication signals, it may be expected that these co-

adapted gene complexes reside in regions of low recombination, because of the increased

potential for linked selection. Here we test this hypothesis in *Laupala*, a genus of crickets

distributed across the Hawaiian archipelago that underwent recent and rapid speciation. We

generate three dense linkage maps from interspecies crosses and use the linkage information to

anchor a substantial portion of a *de novo* genome assembly to chromosomes. Local

recombination rates were then estimated as a function of the genetic and physical distance of

anchored markers. These data provide important genomic resources for Orthopteran model

systems in molecular biology. 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. Contrary to our expectations, a genomic

region where a male song QTL peak co-localizes with a female preference QTL peak was not

associated with particularly low recombination rates. This study shows that trait-preference co-

evolution in sexual selection is not necessarily constrained by local recombination rates.

INTRODUCTION

Speciation is a complex process contingent on the accumulation of genomic divergence over

time. Genomes diverge under the influence of selection and drift, while gene flow and

46   recombination counteract this divergence by homogenizing the genome (Kirkpatrick and

47   Ravigne 2002; Gavrilets 2003). The efficacy of these processes as well as the signatures they

48   leave on the genome are expected to be constrained not only by the genetic architecture of traits

49   under selection but also by structural properties of the chromosomes.

50   Chromosomal rearrangements, transpositions, and the location of GC-rich regions,

51   heterochromatin, or centromeres shape the genomic background and generate local variation in

52   recombination rates (Fullerton *et al.* 2001; Nachman 2002; Noor and Bennett 2009; Ortiz-

53   Barrientos *et al.* 2016). The resulting recombination landscape in turn affects local levels of

54   linkage disequilibrium (Felsenstein 1974, 1981) and, as such, the potential for linked selection

55   and genetic hitchhiking (Smith and Haigh 1974; Gillespie 2000; Cutter *et al.* 2014; Wolf and

56   Ellegren 2016). The recombination landscape also affects the genomic signatures of positive

57   selection on advantageous alleles we aim to detect in genomic scans (Roesti *et al.* 2013;

58   Cruickshank and Hahn 2014; Burri *et al.* 2015). However, the role of variation in chromosome

59   structure and recombination across the genome and among species during speciation remains

60   understudied (Wolf and Ellegren 2016), constraining our ability to understand the patterns of

61   genomic divergence associated with natural and sexual selection (Slatkin 2008; Noor and

62   Bennett 2009; Wolf *et al.* 2010; Wolf and Ellegren 2016).

63   Genetic mapping of genome-scale variation can provide powerful insights into the genomic

64   architecture among closely related species. The earliest genetic maps were used to identify

65   inversions (Sturtevant 1921). Considerably more common than originally thought (Kirkpatrick

66   2010), inversions can effectively halt gene flow locally in the genome by suppressing

67   recombination and can thus have dramatic effects on genomic divergence and speciation

68   (Sturtevant 1921; Noor *et al.* 2001; Rieseberg 2001; Kirkpatrick 2010). Inversions have been

69 found to influence the divergence between species (Kulathinal *et al.* 2008; Barb *et al.* 2014) as

70 well as between different sexual morphs within species (Küpper *et al.* 2015; Lamichhaney *et al.*

71 2015). Other chromosomal rearrangements, such as genetic transpositions or translocations, can

72 likewise contribute to 'chromosomal speciation' (Rieseberg 2001) as well as preventing gene

73 flow and furthering genetic divergence among heterospecifics.

74 Marker orders in linkage maps constitute important preliminary evidence for inversions and

75 other chromosomal rearrangements. In addition, deviation in allele frequencies from 1:2:1 ratios,

76 i.e. segregation distortion, can reveal meiotic drive and local genetic incompatibilities. It is

77 common practice to filter out strongly distorted loci in the mapping process as they may be

78 indicative of sequencing errors. However, closely linked groups of markers with intermediate

79 levels of segregation distortion can result from meiotic drive associated with selfish genetic

80 elements and other active segregation distorters as well as from the incompatibility of genomic

81 material from one parent into the background of another parent (Burt and Trivers 2006;

82 Presgraves 2010). When sufficiently dense, genetic maps can also be used to anchor and order

83 scaffolds of genome assemblies (Fierst 2015). The combination of classical (genetic mapping)

84 and next-gen (high throughput sequencing) approaches offers the opportunity to estimate

85 recombination rate variation along the genome by observing local variation in the relationship

86 between physical distance (usually measured in megabases, Mb) and genetic distance (measured

87 in centimorgans, cM) at various scales and across sexes, species, and populations (Kong *et al.*

88 2010; Smukowski and Noor 2011).

89 There is substantial variation in recombination rates among taxa (Wilfert *et al.* 2007; Smukowski

90 and Noor 2011), with social hymenopterans showing the highest rates in insects (16.1 cM/Mb for

91 *Apis melifera*) and dipterans among the lowest (0.1 cM/Mb for the mosquito *Armigeres*

92   *subalbatus*); however, just within the genus *Drosophila* recombination rates can vary from 0.3

93   cM/Mb to 3.1 cM/Mb (Wilfert *et al.* 2007) and 50 fold differences have been observed within

94   single chromosomes of humans and birds (Myers *et al.* 2005; Singhal *et al.* 2015). Variation in

95   recombination is intimately associated with adaptation (Hill and Robertson 1966; Felsenstein

96   1974), sex chromosome evolution (Wilson and Makova 2009), and genomic heterogeneity in

97   genetic variation (Smith and Haigh 1974; Begun and Aquadro 1992; Nachman 2002; Cutter *et*

98   *al.* 2014). In addition, the role of recombination in generating or eroding co-adapted gene

99   complexes, and thereby promoting or discouraging speciation, are well known in the context of

100   sympatric speciation (Felsenstein 1981) and reinforcement (Stevison *et al.* 2011). Therefore,

101   characterizing the recombination landscape of closely related species is fundamental to

102   understanding how genome architecture and evolution contributes to the mechanics of speciation

103   Here, we generate and compare three marker-dense interspecific linkage maps and infer genome-

104   wide and local recombination rates using four closely related species of Hawaiian sword-tail

105   crickets of the genus *Laupala*. The genus *Laupala* is one of the fastest speciating taxa known to

106   date (Mendelson and Shaw 2005) and the product of a very recent evolutionary radiation, with

107   each of the 38 species endemic to a single island of the Hawaiian archipelago (Otte 1994; Shaw

108   2000a); moreover, the four focal species in this study are each endemic to Hawaii Island, the

109   youngest island in the chain. Evidence suggests that speciation by sexual selection on the

110   acoustic communication system has driven this rapid diversification, as both male mating song

111   and female acoustic preferences have diverged extensively among *Laupala* species (Otte 1994;

112   Shaw 2000b; Mendelson and Shaw 2002). Sexual trait evolution strongly contributes to the onset

113   and maintenance of reproductive isolation (Mendelson and Shaw 2002; Grace and Shaw 2011).

114   As a consequence, substantial co-evolution of male song and female preference exists among

115   species and among populations within species (Grace and Shaw 2011). Although the

116   mechanisms of trait-preference co-evolution require further study, there is evidence that genetic

117   loci controlling quantitative variation in traits and preferences are physically linked in the

118   genome (Shaw and Lesnick 2009; Wiley *et al.* 2012).

119   A handful of other speciation model system have suggested a critical role for genome

120   architecture in promoting speciation (Ortiz-Barrientos *et al.* 2016; Wolf and Ellegren 2016) but

121   to date these have not included systems characterized by sexual selection. We have a limited

122   basis to expect that genetic co-evolution of the sexual traits may be accompanied by

123   chromosomal rearrangements (e.g. Küpper *et al.* 2015; Lamichhaney *et al.* 2015). In addition,

124   theoretical models of speciation by sexual selection imply strong genetic linkage (linkage

125   disequilibrium) of trait and preference genes (Fisher 1930; Lande 1981; Kirkpatrick 1982). The

126   colocalization of a moderate effect quantitative trait locus (QTL) of the male song rhythm and a

127   corresponding female preference QTL (Shaw and Lesnick 2009) thus provides a unique

128   opportunity to test whether genetic linkage of these coevolving traits has been facilitated by

129   chromosomal rearrangements or locally reduced recombination rates.

130   To characterize the genomic architecture and estimate rates of recombination in *Laupala* we first

131   produce a *de novo L. kohalensis* draft genome and obtain thousands of SNP markers for hybrid

132   offspring from interspecific crosses. We then generate three medium to high density genetic

133   (linkage) maps and compare these maps to examine collinearity, chromosomal rearrangements,

134   and genetic incompatibilities across the genome. There is some variation in the level of overall

135   differentiation in the species pairs studied here, but all lineages are young (approximately 0.5

136   million years or less, Fig 1). We thus expect limited chromosomal differentiation. We also

137   expect inversions and levels of genetic incompatibility to be more pronounced in the crosses of

138   the more distantly related species. We then anchor our assembled scaffolds to chromosomes, and

139   infer recombination rates across the chromosomes. Using an additional map that integrates the

140   AFLP markers from previous QTL studies in *L. kohalensis* and *L. paranigra*, we identify the

141   location of known male song QTL, including one co-localizing with a female acoustic preference

142   QTL, to the chromosomes. We examine local variation in recombination rates across the genome

143   and in relation to the location of the genetic architecture of song and preference divergence. This

144   study provides important insight into the role of the genomic architecture in the divergence of

145   closely related, sexually divergent species as well as pivotal resources for future studies.

146   **MATERIAL & METHODS**

147   *De novo* **genome assembly**

148   The *Laupala kohalensis* draft genome was sequenced using the Illumina HiSeq2500 platform.

149   DNA was isolated with the DNeasy Blood & Tissue Kits (Qiagen Inc., Valencia, CA, USA)

150   from six immature female crickets (c. five months of age) chosen randomly from a laboratory

151   stock population (approximately lab generation=14). Females were chosen to balance DNA

152   content of sex chromosomes to autosomes (female crickets are XX; male crickets are XO). DNA

153   was subsequently pooled for sequencing. Four different libraries were created: a paired-end

154   library with an estimated insert size of 200 bp (sequenced by Cornell Biotechnology Resource

155   Center), a paired-end library with an estimated insert size of 500 bp, and two mate-pair libraries

156   with insert sizes of 2 and 5 Kb (sequenced by Cornell Weill College Genomics Resources Core

157   Facility).

158   Reads were processed using Fastq-mcf from the Ea-Utils package (Aronesty 2011) with

159   the parameters -q 30 (trim nucleotides from the extremes of the read with qscore below 30) and -

160   l 50 (discard reads with lengths below 50 bp). Read duplications were removed using PrinSeq

161   (Schmieder and Edwards 2011) and reads were corrected using Musket with the default

162   parameters (Liu *et al.* 2013).

163   Reads were assembled using SoapDeNovo2 (Luo *et al.* 2012). The reads were assembled

164   using different Kmers (k = 31, 39, 47, 55, 63, 71, 79 and 87). The 87-mer assembly produced the

165   best assembly (based on N50/L50, assembly size, and number of scaffolds). Scaffolds and

166   contigs were renamed using an in-house Perl script. Gaps were filled using GapCloser from the

167   SoapDeNovo2 package.

168   The gene space covered by the assembly were evaluated using three different approaches.

169   (1) *Laupala kohalensis* unigenes produced by the Gene Index initiative (Cricket release 2.0:

170   http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=cricket) were mapped using Blat

171   (Kent 2002). Only unigenes mapping with 90% or more of their length were considered; (2)

172   RNASeq from a congeneric species, *L. cerasina* was mapped using Tophat2 (Kim *et al.* 2013).

173   Reads were processed using the same methodology that was described above, but using a

174   minimum length of 30 bp; (3) using BUSCO (Simão *et al.* 2015) search results for conserved

175   eukaryotic and arthropod genes.

Samples

177   We generated three F_2 interspecies hybrid families to estimate genetic maps. F_1 males and

178   females from one parental pair per hybrid cross type were intercrossed to generate an F_2 mapping

179   population, including the following: (1) a *L. kohalensis* female and *L. paranigra* male

180   ("ParKoh", 178 genotyped F_2 hybrid offspring; previously reported in Shaw et al. 2007); (2) a *L.*

181   *kohlanesis* female and a *L. pruna* male ("PruKoh", 193 genotyped F_2 hybrid offspring); a *L.*

182   *paranigra* female and a *L. kona* male ("ParKon", 263 genotyped F_2 hybrid offspring). Crickets

183   used in crosses were a combination of lab stock and wild-caught juveniles from their respective

184   species and reared to adulthood in a temperature-controlled room (20°C) with access to Purina

185   cricket chow and water *ad libitum.* Population and geographical locations from which

186   individuals were sampled are shown in Fig 1 and in Table S1.

187   We lost the parents for PruKoh, so we used sequence data from a single, non-parental *L.*

188   *pruna* individual, and three available *L. kohalensis* individuals, each from the same populations

189   as the parents, to select segregation patterns and establish informativeness of the markers

190   (resulting in a smaller number of available markers compared to the other crosses, see Results).

191   These 'surrogate parental samples' were included in the genotyping pipeline described below

192   instead of the actual parents to this cross.

193   **Genotyping**

194   DNA was extracted from whole adults using the DNeasy Blood & Tissue Kits (Qiagen,

195   Valencia, CA, USA). Genotype-by-Sequencing library preparation and sequencing were done in

196   2014 at the Genomic Diversity Facility at Cornell University following (Elshire *et al.* 2011). The

197   ApeKI methylation sensitive restriction enzyme was used for sequence digestion and DNA was

198   sequenced on the Illumina HiSeq 2500 platform (Illumina Inc., USA).

199   Reads were trimmed and demultiplexed using Flexbar (Dodt *et al.* 2012) and then mapped to the

200   *L. kohalensis de novo* draft genome using Bowtie2 (Langmead and Salzberg 2012) with default

201   parameters. We then called SNPs using two different pipelines: The Genome Analysis Toolkit

202   (GATK; DePristo *et al.* 2011; Van der Auwera *et al.* 2013) and FreeBayes (Garrison and Marth

203   2012). For GATK we used individual BAM files to generate gVCF files using 'HaplotypeCaller'

204   followed by the joint genotyping step 'GenotypeGVCF'. We then evaluated variation in SNP

205   quality across all genotypes using custom R scripts to determine appropriate settings for hard

206   filtering based on the following metrics (based on the recommendations for hard filtering section

207   "Understanding and adapting the generic hard-filtering recommendations" at

208   https://software.broadinstitute.org/gatk/ accessed on 28 February 2017): quality-by-depth, Phred-

209   scaled P-value using Fisher's Exact Test to detect strand bias, root mean square of the mapping

210   quality of the reads, u-based z-approximation from the Mann-Whitney Rank Sum Test for

211   mapping qualities, u-based z-approximation from the Mann-Whitney Rank Sum Test for the

212   distance from the end of the read for reads with the alternate allele. For FreeBayes we called

213   variants from a merged BAM file using standard filters. After variant calling we filtered the

214   SNPs using 'vcffilter', a Perl library part of the VCFtools package (Danecek *et al.* 2011) based

215   on the following metrics: quality (> 30), depth of coverage (> 10), and strand bias for the

216   alternative and reference alleles (SAP and SRP, both > 0.0001). Finally, the variant files from the

217   GATK pipeline and the FreeBayes pipeline were filtered to only contain biallelic SNPs with less

218   than 10% missing genotypes using VCFtools.

219   We retained two final variant sets: a 'high-confidence' set including only SNPs with identical

220   genotype calls between the two variant discovery pipelines and a "full" set of SNPs which

221   included all variants called using FreeBayes but limited to positions that were shared among the

222   GATK and FreeBayes pipelines.

223   **Linkage mapping**

224   The linkage maps deriving from the three species crosses were generated independently

225   and by taking a three-step approach, employing both the regression mapping and the maximum

226   likelihood (ML) mapping functions in JoinMap 4.0 (van Ooijen 2006) as well as the three-point

227   error-corrected ML mapping function in MapMaker 3.0 (Lander *et al.* 1987; Lincoln *et al.* 1993).

228   In the first step, we estimated "initial" maps that are relatively low (5 centimorgan, cM)

229   resolution but with high marker order certainty. For initial maps, we first grouped ($3.0 \leq$ LOD $\leq$

230   5.0) and then ordered a subset of evenly, well-spaced, highly informative (see above) markers

231   with no segregation distortion (χ^2- square associated *P*-value $\geq$ 0.05) and fewer than 95% of their

232   genotypes in common. We then checked for concordance among the three mapping algorithms.

233   In most cases, the maps were highly concordant (in ordering of the markers; with respect to

234   centimorgan (cM) among markers, distances differed depending on the algorithm, especially

235   between the regression and ML methods in JoinMap). Discrepancies among the maps produced

236   by the different algorithms for the same cross were resolved by optimizing the likelihood and

237   total length of a given map as well as by using the information in JoinMap's "Genotype

238   Probabilities" and "Plausible Positions".

239   These initial maps were then filled out using MapMaker with marker loci passing more

240   lenient criteria (drawn from the full set of SNPs, with segregation distortion χ^2- square associated

241   q-value $\leq$ 0.05, i.e. a 5% false discovery rate [Benjamini & Hochberg, 1995] and fewer than 99%

242   of their genotypes in common with other markers loci (i.e. excluding marker loci with identical

243   genotypes for all individuals), favoring those marker loci shared among the three mapping

244   populations. First, more informative markers (no missing genotypes, > 2.0 cM distance from

245   other markers) were added satisfying a log-likelihood threshold of 4.0 for the positioning of the

246   marker (i.e., assigned marker position is 10,000 times more likely than any other position in the

247   map). Remaining markers were added at the same threshold, followed by a second round for all

248   markers at a log-likelihood threshold of 3.0.

249   In the second step, "comprehensive" maps were obtained in MapMaker by sequentially

250   adding markers (meeting the lenient criteria described above) to the initial map, satisfying a log-

251   likelihood threshold of 2.0 for the marker positions, followed by a second round with log-

252   likelihood threshold equaling 1.0. We then used the ripple algorithm on 5-marker windows and

253   explored alternative orders. Typically, MapMaker successfully juxtaposes SNP markers from the

254   same scaffold. However, in marker dense regions with low recombination rates, the likelihoods

255   of alternative marker orders coalesce. In such regions, when multiple markers from the same

256   genomic scaffold were interspersed by markers from a different scaffold, we repositioned the

257   former markers by forcing them in the map together. If the log-likelihood of the map decreased

258   by more than 3.0 (factor 1000), only one of the markers from that scaffold was used in the map.

259   The comprehensive maps provide a balance between marker density and confidence in marker

260   ordering and spacing.

261   The third step was to create "dense" maps. We added all remaining markers that were not

262   yet incorporated in step two, first at a log-likelihood threshold of 0.5, followed by another round

263   at a log-likelihood threshold of 0.1. We then used the ripple command as described above. The

264   dense maps are useful for anchoring of scaffolds and for obtaining the highest possible resolution

265   of variation in recombination rates, but with the caveat that there is some uncertainty in marker

266   order. Uncertainty is expected to be higher towards the centers of the linkage groups where

267   crossing over events between adjacent markers become substantially less frequent (see Results).

268   **Comparative analyses**

269   Based on the recent divergence times and high interbreeding successes, we anticipated few

270   chromosomal rearrangements and, accordingly, a large degree of collinearity of the linkage

271   maps. We first searched for evidence for inversions or other chromosomal rearrangements by

272 comparing marker ordering among the initial and comprehensive linkage maps visually using

273 map graphs from MapChart (Voorrips 2002). Inverted or transposed markers present in two or

274 all maps can be detected by connecting "homologs" in MapChart. Then, we tested whether

275 marker order and marker spacing (the genetic distance between pairs of markers) were conserved

276 across the maps using linear models. If the marker order and genetic distance among markers is

277 similar in a pair of linkage maps, a linear relationship with regression slope equal to unity is

278 expected between marker positions in the two maps. We regressed the marker positions in the

279 ParKon and PruKoh maps, respectively, against the marker positions in the ParKoh map.

280 Deviations from linearity (reflected by $R^2 < 1.00$ of the model) indicate that relative marker

281 positions within the maps differs. The slope of the fitted regression curve indicates whether

282 marker distances increase at higher rates in the PruKoh or ParKon map relative to the shared

283 markers in the ParKoh map. If the slope of the regression curve equals 1, the markers are spaced

284 out equally in the two maps and deviations from 1 indicate that markers are closer together or

285 further apart in one map relative to the other. The null hypothesis of collinear maps thus predicts

286 $R^2 \approx 1.00$ and unit slope in both comparisons.

287   Genetic incompatibilities are expected to increase with increasing differentiation of the

288 genome and the chromosomal architecture. In a linkage map, local genetic incompatibility is

289 indicated by an excess of one of the parental alleles or a deficiency in heterozygotes in groups of

290 adjacent markers. Because the *L. kohalensis* and *L. paranigra* are more distantly related to each

291 other than they are to *L. pruna* and *L. kona*, respectively (Mendelson and Shaw 2005; see Fig 1.),

292 we expected more regions with significant segregation distortion in the ParKoh map relative to

293 the ParKon and PruKoh maps. We checked for regions of elevated segregation distortion across

294 the linkage groups in R (R Development Core Team 2016) using the R/qtl package (Broman *et*

295   *al.* 2003). Although we filtered out markers with very high levels of segregation distortion (using

296   a 5% FDR cutoff), markers with medium to high levels of segregation distortion were still

297   present. Grouped markers in a given region of a linkage group with medium to high levels of

298   segregation distortion represent genomic regions with biased parental allele contributions,

299   suggesting genetic incompatibilities (or, less common, selfish alleles and other active segregation

300   distorters).

301   The chromosomal rearrangements and genetic incompatibilities as well as several other

302   properties of genomes result in variation in recombination rates across and among chromosomes,

303   constraining genetic divergence. To examine the patterns of variation in crossing over in the

304   *Laupala* genome we first merged the three maps using ALLMAPS (Tang *et al.* 2015). Then, we

305   calculated species-specific average recombination rates for the linkage groups by dividing the

306   total length of the linkage group (in cM) by the physical length of the chromosome obtained

307   using ALLMAPS (in million bases, Mb). Lastly, to evaluate recombination rate variation along

308   the linkage groups, we fitted smoothing splines (with 10 degrees of freedom) in R to obtain

309   functions that describe the relationship between the consensus physical distance (as per the

310   anchored scaffolds) and the genetic distance specific to each linkage map. Variation in the

311   recombination rate was then assessed by taking the first derivative of the fitted spline function.

312   The estimated recombination rates are likely to be an overestimate of the true recombination rate,

313   because unplaced/unordered parts of the assembly do not contribute to the physical length of the

314   chromosomes but are reflected in the genetic distances obtained from crossing-over events in the

315   recombining hybrids.

316   To test the hypothesis that linked trait and preference genes reside in low recombination

317   regions to facilitate linkage, we integrated the AFLP map and song and preference QTL peaks

318   identified in previous work (Shaw and Lesnick 2009) with the current ParKoh SNP map and

319   projected the QTL peaks onto the anchored genome. The SNPs used in the present study were

320   obtained from the same mapping population (same individuals) as in the 2009 AFLP study.

321   Therefore, we simply combined high confidence SNPs described above with the AFLP markers

322   and created a new linkage map using the stringent criteria for the "initial" maps. We projected

323   this map onto the anchored draft genome based on common contigs. We then approximated the

324   physical location of the QTL peaks by looking for SNP markers on scaffolds present in the draft

325   genome flanking AFPL markers underneath the QTL peaks identified in the 2009 study.

326   **DATA ACCESIBILITY**

327   Raw data and R-scripts will be deposited in Dryad upon final acceptance and are available upon

328   request. The genome sequences and chromosome assembly are available on NCBI's GenBank

329   under BioProject number PRJNA392944 (as of the date of manuscript submission no accession

330   numbers are available, but these will be included in a next version of this manuscript).

331   **RESULTS**

332   *De novo* **genome assembly**

333   The sequencing of the four libraries yielded 162.5 Gb of raw sequences (Table 1). After read

334   processing, 145.5 Gb was used for the sequence assembly. We compared among assemblies

335   resulting from different Kmer sizes (k = 31, 39, 47, 55, 63, 71, 79 and 87). Based on the

336   N50/L50 and the total assembly size, the assembly produced with k = 87 was retained for the

337   final draft genome. Despite a large number of scaffolds in the final assembly, the median length

338   of the scaffolds was high and the total length of the assembly covers about 83% of the expected

339   complete genome in *Laupala*.

340   Gene space coverage in the assembly was evaluated using the *L. kohalensis* cricket gene

341   index (Danley *et al.* 2007) (release 2.0), RNASeq from *Laupala cerasina* (unpublished) and by

342   performing a BUSCO search using eukaryotic and arthropod specific conserved genes.

343   Respectively 95% and 92% of the L*aupala* gene index and RNAseq sequences mapped to the

344   current genome. In addition, the BUSCO search indicated very few missing genes in either

345   database (Table 1).

346

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

348

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		149,424	
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		67.4	
N50 index	9,917		756	
N50 length (Kb)	43.6		581.3	
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

349

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

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

355

356 **Linkage maps**

357 We obtained 815,109,126; 522,378,849; and 311,558,401 reads after demultiplexing for ParKoh,

358 ParKon, and PruKoh, respectively. Average sequencing depth ± standard deviation was 105.6 x

359 ± 52.0, 35.5 x ± 16.5, and 44.4 x ± 13.1, respectively.

360 In the initial maps, 158 (ParKoh), 170 (ParKon), and 138 (PruKoh) markers were grouped into

361 eight linkage groups at a LOD score of 5.0, as expected based on the existence of seven

362 autosomes and one X-chromosome in *Laupala.* The corresponding marker densities were 5.14,

363 4.85, and 7.33 markers per cM, containing 526, 650, and 325 markers at 1.91, 1.37, and 3.25

364 markers per cM and on the dense maps we placed 608, 823, and 383 markers at 1.69, 1.15, and

365 2.80 markers per cM.

366 **Collinear linkage maps with few rearrangements**

367 We expected few rearrangements and a high degree of collinearity among the linkage maps. The

368 visual comparison of marker positioning showed that relative locations of shared scaffolds were

369 similar across the linkage maps in both the initial and the comprehensive maps with only few

370 putative inversions or genetic transpositions (Fig 2, Fig S1). Using linear regression, we then

371 tested whether the marker order (represented by the R^2) and the relative distance between

372 markers (represented by the slope, β) shared across the maps were similar. Our results generally

373 supported the null hypothesis of high collinearity. Although there was some variation in marker

374 order and map length among the three maps, marker positions on most linkage groups differed

375 only marginally from linearity and increased in genetic distance at similar rates (Table 2).

376 Table 2. Linkage map comparison.

	ParKoh ~ ParKon		ParKoh ~ PruKoh	
	$R^{2\,a}$	slope (95% CI) [b]	R^2	slope (95% CI)
1	0.93	0.92 (0.86,0.98)	0.94	1.08 (0.99, 1.17)
2	0.95	1.61 (1.52,1.69)	0.95	1.31 (1.21,1.42)
3	0.98	1.29 (1.25,1.33)	0.97	0.76 (0.71,0.81)
4	0.94	0.90 (0.83,0.97)	0.95	0.95 (0.84,1.06)
5	0.92	1.36 (1.27,1.45)	0.87	1.08 (0.93,1.23)
6	0.99	0.95 (0.92,0.99)	0.94	0.51 (0.45,0.58)
7	0.95	0.71 (0.65,0.77)	0.67	0.86 (0.34,1.37)
X	0.68	1.04 (0.81,1.26)	0.91	1.04 (0.93,1.14)

377 [a] R^2 indicates the extend of linearity in the relative marker positions.

378 [b] slope measures whether the marker positions tend to be significantly larger in the ParKoh map (values

379 larger than 1 +/- 95% confidence interval) or larger in the ParKon or PruKoh map (values smaller than 1).

380 Further support came from Pearson correlation coefficients between the linkage groups from the

381 cross-specific maps and the assembled chromosomes. These were generally high (> 0.95),

382 indicating substantial synteny between the different genetic maps and the (consensus) anchored

383 reference genome (Fig. S2). We did, however, observe some evidence for inverted regions on all

384 linkage groups except LG 4, mostly in the peripheral regions of the chromosomes (Fig. 2, Fig.

385   S1). Additionally, there was some variation in the length of the linkage groups among the crosses

386   most notably for LG 3, 6, and 2 (Table 3).

387   Limited heterogeneity in segregation distortion

388   We expected genetic incompatibility to be more likely in the ParKoh cross than in the ParKon

389   and PruKoh cross, because *L. kohalensis* and *L. paranigra* are more distantly related than any of

390   the other species pairs (Fig 1). We tested this hypothesis by examining the degree of segregation

391   distortion in groups of adjacent markers among the linkage maps. There were few patterns of

392   regionally high segregation distortion in all three crosses (Fig 3). However, LG 1 showed a bias

393   against *L. kohalensis* homozygotes in the ParKoh cross but not in any of the other crosses,

394   lending some support to the hypothesis of limited genetic incompatibility. Additionally, there

395   was significant variation in the frequency of heterozygotes across the linkage groups (linear

396   model Freq[heterozygotes]~LG x cross: $R^2 = 0.21$, $F_{20,1547} = 20.7$, $P < 0.0001$). Post-hoc Tukey

397   Honest Significant Differences revealed that linkage group 6 had the lowest abundance of

398   heterozygotes overall and within each of the intercrosses and that levels of heterozygosity on LG

399   6 were similar across the maps (Table S2).

400   Variable recombination rates across the genome

401   We anchored a total of 1054 scaffolds covering 720 million base pairs, a little below half the

402   current genome assembly (see Table S3 for anchoring statistics). Average recombination rates

403   (cM/Mb) varied from between 0.75 (ParKon) and 0.93 (ParKoh) on the X chromosome to

404   between 3.12 (ParKon) and 4.24 (PruKoh) on LG 7 (Table 3). Both including and excluding the

405   sex chromosome, there is a significant negative relationship (with X: $\beta = -0.016$, $R^2 = 0.72$, $F_{1,5} =$

406   15.500, $P = 0.0076$; without X: $\beta = -0.015$, $R^2 = 0.66$, $F_{1,5} = 9.754$, $P = 0.0261$) between

407   chromosome size and broad-scale recombination rate.

408   Table 3. Linkage map summary statistics.

LG	Length (Mb)	ParKoh		ParKon		PruKoh	
		Map length (cM)	Rec. Rate (cM/Mb)	Map length (cM)	Rec. Rate (cM/Mb)	Map length (cM)	Rec. Rate (cM/Mb)
1	137	169	1.23	167	1.22	173	1.26
2	117	207	1.77	156	1.33	156	1.33
3	102	167	1.64	128	1.25	205	2.01
4	90	100	1.11	117	1.30	99	1.10
5	62	91	1.47	84	1.35	103	1.66
6	53	78	1.47	84	1.58	139	2.62
7	25	85	3.40	106	4.24	78	3.12
X	134	124	0.93	101	0.75	114	0.85
Total	**720**	**1021**	**1.42**	**943**	**1.31**	**1067**	**1.48**

409

410   Most linkage groups showed wide regions of strongly reduced recombination rates in the center

411   of the linkage groups, revealing the metacentric location of the centromeres and the coincidence

412   of the centromere with large recombination 'deserts' (Fig. 4). The general pattern of

413   recombination rate variation across the genome was similar among the three intercrosses, but

414   some cross-specific peaks in recombination rates were observed on almost all linkage groups

415   (Fig 4).

416   **The colocalizing trait and preference QTL are not associated with low recombination rates**

417   Contrary to our expectation, the approximate location of the colocalizing song and preference

418   QTL peak from (Shaw and Lesnick 2009) was associated with average recombination rates in the

419   ParKoh and ParKon map and low recombination rates in the PruKoh map (Fig 4; Table S4). The

420   majority of the other QTL peaks are located in regions of low recombination (Fig 4; Table S4).

421   **DISCUSSION**

422 The likelihood of speciation and the evolutionary trajectory of diverging populations is

423 contingent on the chromosomal architecture, both by directly invoking speciation events through

424 chromosomal rearrangements, as well as by locally affecting the efficacy of selection on

425 complex traits through recombination rate variation (Noor *et al.* 2001; Rieseberg 2001; Butlin

426 2005; Slatkin 2008; Noor and Bennett 2009; Cutter *et al.* 2014; Ortiz-Barrientos *et al.* 2016). We

427 find large, centric regions with low recombination rates but a major male song rhythm QTL that

428 co-localizes with a female preference QTL did not map to a region with particularly low

429 recombination. This is important because it challenges the hypothesis that co-evolution of traits

430 and preferences is constrained by local recombination rates. Additionally, we projected three

431 high density linkage maps onto a *de novo* assembly of the *Laupala* draft genome – the first

432 published draft genome for crickets – to generate the first assembly of pseudomolecules for

433 Orthoptera. This is a major step forward in creating genomic resources for an important model

434 system in neurobiological, behavioral ecological, and evolutionary genetics studies (Horch *et al.*

435 2017).

436 *Chromosomal rearrangements*

437 Variation in the organization and structure of chromosomes contributes to postzygotic

438 reproductive isolation after speciation and has been suggested to contribute to the speciation

439 process directly (Noor *et al.* 2001; Rieseberg 2001). The *Laupala* radiation comprises many

440 morphologically and ecologically cryptic, but sexually divergent species that can be successfully

441 interbred and have diverged only recently (Otte 1994; Shaw 1996; Mendelson and Shaw 2005).

442 We therefore expected limited variation in chromosome structure and number and few signatures

443 of genetic incompatibility between the species. In line with these expectations, we observed

444 limited large scale structural variation and found that linkage maps were collinear across

445   interspecies crosses (Fig 2). In addition, apart from linkage group 1 in the cross of the most

446   distantly related species pair, we found no sign of broadly distorted allele frequencies in the F_2

447   generations (Fig 3). Structural genomic variation thus appears to be an unlikely contributor to

448   isolation in the recent evolutionary history of *Laupala*.

449   However, we did find signatures of several inversions when comparing the interspecific linkage

450   maps. Inversions can locally suppress recombination and put heterozygotes at a selective

451   disadvantage, thereby contributing to reproductive barriers between species as well as initiating

452   the speciation process (Noor *et al.* 2001; Rieseberg 2001; Stevison *et al.* 2011). Regions showing

453   inversions between the genetic maps were mostly located in the peripheral regions of the linkage

454   groups. Particularly linkage group 6 showed substantial variation in marker order and genetic

455   map length across the intercrosses. The observed deficit in heterozygotes on LG 6 compared to

456   the other linkage groups might thus be related to the substantial variation in marker order in the

457   periphery as well as to the variability in the total length of LG6. *Laupala* species hybrids have

458   not yet been discovered in nature, however, these findings indicate a potential for underdominant

459   selection, i.e. selection against heterozygotes for rearranged chromosomal regions (Ortiz-

460   Barrientos *et al.* 2016), to act against interspecific introgression in the case of secondary contact.

461   *Segregation distortion*

462   We found a large region on linkage group 1 showing substantial segregation distortion due to an

463   excess of *L. paranigra* alleles when crossed with *L. kohalensis*, but not in the other two crosses

464   (Fig 3). This observation is in line with expectations of stronger genetic incompatibility between

465   more distantly related species pairs. Segregation distortion by itself does not complicate the

466   ordering of genetic markers (Hackett and Broadfoot 2002), except if it is the result of genotyping

467   errors; accordingly, we filtered out loci with high levels of segregation distortion prior to

468 creating the maps. Any remaining variability in levels of segregation distortion is expected to be

469 due to random variation and genetic incompatibility (Burt and Trivers 2006; Presgraves 2010),

470 which plays an important role in the reinforcement of species boundaries. An important cause of

471 genetic incompatibilities are inversions and other rearrangements. However, our results do not

472 indicate that linkage group 1 is particularly variable among the crosses. Other genetic elements,

473 such as selfish alleles and distorting gene complexes such as the *sd* complex in *Drosophila*

474 (Larracuente and Presgraves 2012) could also lead to the observed pattern on LG1. However,

475 these are likely less common and would be expected to affect the different crosses in similar

476 ways.

477 *Recombination landscape*

478 Chromosomal rearrangements specifically, and the structural variability within the genome in

479 general, influence genomic divergence by locally altering recombination rates. Felsenstein drew

480 important attention to the role of recombination rates in speciation (Felsenstein 1974, 1981). In

481 recent years, characterization of the recombination landscape has received increasing attention

482 (Butlin 2005; Slatkin 2008; Noor and Bennett 2009; Barb *et al.* 2014; Burri *et al.* 2015), in

483 particular in relation to its role in linked and background selection – linkage of neutral loci to

484 positively selected alleles and linkage of neutral loci to deleterious alleles, respectively (Cutter *et*

485 *al.* 2014). Here, we show that there is limited variation in recombination rates among the species

486 pairs (Fig 4). The genome-wide average varied between 1.3 and 1.5 cM/Mb (Table 3), similar to

487 dipterans and substantially lower than social hymenopterans and lepidopterans (Wilfert *et al.*

488 2007).

489 However, for all three species pairs we document high variability in recombination across

490 genomic regions. We found large regions of low recombination in all three maps, with

491   recombination rates well below 1 cM/Mb and occasionally approaching zero, flanked by steep

492   inclines reaching rates up to 6 cM/Mb (Fig 4). This pattern is consistent with earlier findings in

493   plants (Anderson *et al.* 2003), invertebrates (Rockman and Kruglyak 2009; Niehuis *et al.* 2010),

494   and vertebrates (Backström *et al.* 2010; Roesti *et al.* 2013; Singhal *et al.* 2015), but differs from

495   observations in, for example, *Drosophila* (Kulathinal *et al.* 2008) and humans (Myers *et al.*

496   2005).

497   We found that the wide regions of low recombination were characteristic of the larger linkage

498   groups (size in terms of the sum of the lengths of the anchored scaffolds), which also had lower

499   average recombination rates. The smaller linkage groups also show substantial recombination

500   rate heterogeneity with decreasing recombination rates closer to the centromere, however low

501   levels of recombination are limited to only a narrow, centric region. Although recombination rate

502   heterogeneity can arise from selection against recombination due to negative epistasis or the

503   maintenance of linkage disequilibrium between mutually beneficial alleles (Smukowski and

504   Noor 2011; Stevison *et al.* 2011; Smukowski Heil *et al.* 2015; Ortiz-Barrientos *et al.* 2016), the

505   generality of this pattern across the genome and across taxa implies there are generic molecular

506   mechanisms underlying the peripheral bias in crossing over.

507   In addition to the relation between recombination rates and linked and background selection,

508   linkage disequilibrium as a result of reduced crossing over along extensive genomic regions

509   could have important evolutionary consequences on sexual trait evolution in this system. There is

510   known sexual trait-preference matching in this system, with co-localizing QTL for male song

511   and female preference (Shaw and Lesnick 2009; Wiley *et al.* 2012). The highly polygenic

512   genetic architectures associated with sexual (acoustic) communication in crickets would likely

513   require genetic linkage (physical or otherwise) of trait and preference genes to co-evolve at the

514   elevated rates observed in Hawaiian crickets (Mendelson and Shaw 2005; Shaw *et al.* 2011).

515   Low recombination rates might enhance or facilitate linkage disequilibrium, thus easing co-

516   evolution of signals and preference.

517   However, our data lend no support for this hypothesis, as the colocalized song and preference

518   QTL peaks in the *L. kohalensis* x *L. paranigra* cross map to a region of relatively high

519   recombination in the genome (Fig. 4). Interestingly, many of the remaining song QTL do map to

520   the center of their respective chromosomes, which are associated with low recombination. Our

521   results thus suggest that on the one hand, low recombination is not a likely driving force in the

522   establishment of linkage between the trait and preference genes in this system. On the other

523   hand, limited recombination around most other song QTL, suggest that interspecific gene flow in

524   these regions is likely limited.

525   There is some uncertainty in the estimated recombination rates in this study. Using interspecific

526   maps may lead to somewhat lower estimates compared to intraspecific maps. For example, in

527   *Nasonia vitripennis* recombination rates estimated from interspecific maps were 1.8% lower than

528   estimates from intraspecific maps (Beukeboom *et al.* 2010). In addition, although we were able

529   to anchor about 50% of the assembly in the first version of the *Laupala* reference genome, there

530   are still many scaffolds unaccounted for. These 'missing' scaffolds are expected to add to the

531   length of the physical map (from which they are absent) more than to the length of the genetic

532   map (where they are largely accounted for by genetic distance of the surrounding markers), thus

533   lowering the recombination rate. However, our study focuses on patterns of recombination,

534   which should not be affected, rather than aiming to infer the true recombination rate, which we

535   can only approximate at this point.

536 In summary, comparing three interspecific linkage maps from a rapid insect radiation, we find

537 limited variation in chromosome structure among species, but similarly strong heterogeneity in

538 the recombination landscape across their genomes. We present a *de novo* genome assembly and

539 anchor a substantial part of the *L. kohalensis* genome to putative chromosomes. Crickets are an

540 important model system for evolutionary and neurobiological research. However, there are

541 limited genomic resources and this first Orthopteran chromosome-level draft genome and

542 recombination rate map offer important background information for future research on the role

543 of selection in genetic divergence and the genomic basis of speciation. In this study, we further

544 provide insight into the extent to which chromosomal rearrangements and structural variation

545 contribute to genetic incompatibilities in closely related but sexually divergent populations.

546 Importantly, we show that trait-preference co-evolution is not necessarily coupled to regions of

547 low recombination, emphasizing that, at least in *Laupala*, other processes drive interspecific

548 variation in sexual communication phenotypes during divergence.

549 **ACKNOWLEDGEMENTS**

552 **REFERENCES**

553 Anderson L. K., Doyle G. G., Brigham B., Carter J., Hooker K. D., *et al.*, 2003 High-resolution

554 crossover maps for each bivalent of Zea mays using recombination nodules. Genetics 165:

555 849–865.

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

557 Anal. Durham, NC.

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

559   FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices

560   Pipeline. Curr. Protoc. Bioinforma. 43: 1–11.

561   Backström N., Forstmeier W., Schielzeth H., Mellenius H., Nam K., *et al.*, 2010 The

562   recombination landscape of the zebra finch *Taeniopygia guttata* genome. Genome Res. 20:

563   485–95.

564   Barb J. G., Bowers J. E., Renaut S., Rey J. I., Knapp S. J., *et al.*, 2014 Chromosomal Evolution

565   and Patterns of Introgression in Helianthus. Genetics 197: 969–979.

566   Begun D. J., Aquadro C. F., 1992 Levels of naturally occurring DNA polymorphism correlate

567   with recombination rates in D. melanogaster.

568   Benjamini Y., Hochberg Y., 1995 Controlling the false discovery rate: a practical and powerful

569   approach to multiple testing. J. R. Stat. Soc. Ser. B: 289–300.

570   Beukeboom L. W., Niehuis O., Pannebakker B. A., Koevoets T., Gibson J. D., *et al.*, 2010 A

571   comparison of recombination frequencies in intraspecific versus interspecific mapping

572   populations of Nasonia. Heredity (Edinb). 104: 302–309.

573   Broman K. W., Wu H., Sen Ś., Churchill G. A., 2003 R/qtl: QTL mapping in experimental

574   crosses. Bioinformatics 19: 889–890.

575   Burri R., Nater A., Kawakami T., Mugal C. F., Olason P. I., *et al.*, 2015 Linked selection and

576   recombination rate variation drive the evolution of the genomic landscape of differentiation

577   across the speciation continuum of Ficedula flycatchers. Genome Res. 25: 1656–1665.

578   Burt A., Trivers R., 2006 Genes in Conflict: The Biology of Selfish Genetic Elements (Belknap,

579   Cambridge, MA).

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

581   Cruickshank T. E., Hahn M. W., 2014 Reanalysis suggests that genomic islands of speciation are
582   due to reduced diversity, not reduced gene flow. Mol. Ecol. 23: 3133–3157.

583   Cutter A. D., And, Payseur B. A., 2014 Genomic signatures of selection at linked sites: unifying
584   the disparity among species. Nat. Rev. Genet. 14: 262–274.

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

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

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

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

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

597   Felsenstein J., 1974 The evolutionary advantage of recombination. Genetics 78: 737–756.

598   Felsenstein J., 1981 Skepticism towards Santa Rosalia, or why are there so few kinds of animals?
599   Evolution (N. Y).: 124–138.

600   Fierst J. L., 2015 Using linkage maps to correct and scaffold de novo genome assemblies:

601   methods, challenges, and computational tools. Front. Genet. 6: 220.

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

603   Fullerton S. M., Carvalho A. B., Clark A. G., 2001 Local rates of recombination are positively

604   correlated with GC content in the human genome. Mol. Biol. Evol. 18: 1139–1142.

605   Garrison E., Marth G., 2012 Haplotype-based variant detection from short-read sequencing.

606   arXiv Prepr. arXiv1207.3907.

607   Gavrilets S., 2003 Perspective: models of speciation: what have we learned in 40 years?

608   Evolution 57: 2197–2215.

609   Gillespie J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model.

610   Genetics 155: 909–919.

611   Grace J. L., Shaw K. L., 2011 Coevolution of male mating signal and female preference during

612   early lineage divergence of the Hawaiian cricket, Laupala Cerasina. Evolution (N. Y). 65:

613   2184–2196.

614   Hackett C. A., Broadfoot L. B., 2002 Effects of genotyping errors, missing values and

615   segregation distortion in molecular marker data on the construction of linkage maps.

616   Heredity (Edinb). 90: 33–38.

617   Hill W. G., Robertson A., 1966 The effect of linkage on limits to artificial selection. Genet. Res.

618   8: 269–294.

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

620   *organism*. Springer Japan, Tokyo, Japan.

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

622 Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., *et al.*, 2013 TopHat2: accurate alignment

623   of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol.

624   14: R36.

625 Kirkpatrick P., 1982 Sexual selection and the evolution of female mate choice. Evolution (N. Y).

626   36: 1–12.

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

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

629 Kirkpatrick M., 2010 How and Why Chromosome Inversions Evolve. PLOS Biol. 8: 1–5.

630 Kong A., Thorleifsson G., Gudbjartsson D. F., Masson G., Sigurdsson A., *et al.*, 2010 Fine-scale

631   recombination rate differences between sexes, populations and individuals. Nature 467:

632   1099–1103.

633 Kulathinal R. J., Bennett S. M., Fitzpatrick C. L., Noor M. A. F., 2008 Fine-scale mapping of

634   recombination rate in Drosophila refines its correlation to diversity and divergence. Proc.

635   Natl. Acad. Sci. 105: 10051–10056.

636 Küpper C., Stocks M., Risse J. E., Remedios N., Farrell L. L., *et al.*, 2015 A supergene

637   determines highly divergent male reproductive morphs in the ruff. Nat. Publ. Gr. 48: 79–83.

638 Lamichhaney S., Fan G., Widemo F., Gunnarsson U., Thalmann D. S., *et al.*, 2015 Structural

639   genomic changes underlie alternative reproductive strategies in the ruff (Philomachus

640   pugnax). Nat. Genet. 48: 84–88.

641 Lande R., 1981 Models of speciation by sexual selection on polygenic traits. Proc Natl Acad Sci

642 U S A 78: 3721–3725.

643 Lander E., Green P., Abrahamson J., Barlow A., Daly M., *et al.*, 1987 MAPMAKER: An

644 interactive computer package for constructing primary genetic linkage maps of

645 experimental and natural populations. Genomics 1: 174–181.

646 Langmead B., Salzberg S. L., 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:

647 357–359.

648 Larracuente A. M., Presgraves D. C., 2012 The Selfish Segregation Distorter Gene Complex of

649 Drosophila melanogaster. Genetics 192: 33–53.

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

651 MAPMAKER/EXP Version 3.0: a tutorial and reference manual. A Whitehead Inst.

652 Biomed. Res. Tech. Rep.: 78–79.

653 Liu Y., Schröder J., Schmidt B., 2013 Musket: a multistage k-mer spectrum-based error corrector

654 for Illumina sequence data. Bioinformatics 29: 308–315.

655 Luo R., Liu B., Xie Y., Li Z., Huang W., *et al.*, 2012 SOAPdenovo2: an empirically improved

656 memory-efficient short-read de novo assembler. Gigascience 1: 18.

657 Mendelson T. C., Shaw K. L., 2002 Genetic and behavioral components of the cryptic species

658 boundary between Laupala cerasina and L. kohalensis (Orthoptera: Gryllidae). Genetica

659 116: 301–310.

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

661 Myers S., Bottolo L., Freeman C., McVean G., Donnelly P., 2005 A Fine-Scale Map of

662 Recombination Rates and Hotspots Across the Human Genome. Science (80-.). 310: 321–

663   324.

664   Nachman M. W., 2002 Variation in recombination rate across the genome: evidence and

665   implications. Curr. Opin. Genet. Dev. 12: 657–663.

666   Niehuis O., Gibson J. D., Rosenberg M. S., Pannebakker B. A., Koevoets T., *et al.*, 2010

667   Recombination and its impact on the genome of the haplodiploid parasitoid wasp Nasonia.

668   PLoS One 5: e8597.

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

670   reproductive isolation of species. Proc. Natl. Acad. Sci. U. S. A. 98: 12084–8.

671   Noor M. A. F., Bennett S. M., 2009 Islands of speciation or mirages in the desert? Examining the

672   role of restricted recombination in maintaining species. Heredity (Edinb). 103: 439–44.

673   Ooijen J. W. van, 2006 JoinMap 4, Software for the calculation of genetic linkage maps in

674   experimental populations.

675   Ortiz-Barrientos D., Engelstädter J., Rieseberg L. H., 2016 Recombination rate evolution and the

676   origin of species. Trends Ecol. Evol. 31: 226–236.

677   Otte D., 1994 *The Crickets of Hawaii: Origin, Systematics, and Evolution*. Orthoptera

678   Society/Academy of Natural Sciences of Philadelphia, Philadelphia, PA.

679   Petrov D. A., Sangster T. A., Johnston J. S., Hartl D. L., Shaw K. L., 2000 Evidence for DNA

680   loss as a determinant of genome size. Science (80-.). 287: 1060–1062.

681   Presgraves D. C., 2010 Darwin and the Origin of Interspecific Genetic Incompatibilities. Am.

682   Nat. 176: S45–S60.

683   R Development Core Team R., 2016 R: A Language and Environment for Statistical Computing

684 (RDC Team, Ed.). R Found. Stat. Comput. 1: 409.

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

686 357.

687 Rockman M. V, Kruglyak L., 2009 Recombinational landscape and population genomics of

688 Caenorhabditis elegans. PLoS Genet 5: e1000419.

689 Roesti M., Moser D., Berner D., 2013 Recombination in the threespine stickleback genome -

690 Patterns and consequences. Mol. Ecol. 22: 3014–3027.

691 Schmieder R., Edwards R., 2011 Quality control and preprocessing of metagenomic datasets.

692 Bioinformatics 27: 863–864.

693 Shaw K. L., 1996 Polygenic Inheritance of a Behavioral Phenotype: Interspecific Genetics of

694 Song in the Hawaiian Cricket Genus Laupala. Evolution (N. Y). 50: 256–266.

695 Shaw K. L., 2000a Further acoustic diversity in Hawaiian forests: two new species of Hawaiian

696 cricket (Orthoptera: Gryllidae: Trigonidiinae: Laupala). Zool. J. Linn. Soc. 129: 73–91.

697 Shaw K. L., 2000b Interspecific genetics of mate recognition: inheritance of female acoustic

698 preference in Hawaiian crickets. Evolution 54: 1303–1312.

699 Shaw K. L., Lesnick S. C., 2009 Genomic linkage of male song and female acoustic preference

700 QTL underlying a rapid species radiation. Proc. Natl. Acad. Sci. U. S. A. 106: 9737–9742.

701 Shaw K. L., Ellison C. K., Oh K. P., Wiley C., 2011 Pleiotropy, "sexy" traits, and speciation.

702 Behav. Ecol. 22: 1154–1155.

703 Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V, Zdobnov E. M., 2015 BUSCO:

704 assessing genome assembly and annotation completeness with single-copy orthologs.

705   Bioinformatics 31: 3210.

706   Singhal S., Leffler E. M., Sannareddy K., Turner I., Venn O., *et al.*, 2015 Stable recombination

707   hotspots in birds. Science (80-.). 350: 928–932.

708   Slatkin M., 2008 Linkage disequilibrium—understanding the evolutionary past and mapping the

709   medical future. Nat. Rev. Genet. 9: 477–485.

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

711   Smukowski C. S., Noor M. A. F., 2011 Recombination rate variation in closely related species.

712   Heredity (Edinb). 107: 496–508.

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

714   Hotspots: A Comprehensive Evolutionary Portrait of Recombination in Two Closely

715   Related Species of Drosophila. Genome Biol. Evol. 7: 2829–42.

716   Stevison L. S., Hoehn K. B., Noor M. A. F., 2011 Effects of Inversions on Within- and Between-

717   Species Recombination and Divergence. Genome Biol. Evol. 3: 830.

718   Sturtevant A. H., 1921 A case of rearrangement of genes in Drosophila. Proc. Natl. Acad. Sci. 7:

719   235–237.

720   Tang H., Zhang X., Miao C., Zhang J., Ming R., *et al.*, 2015 ALLMAPS: robust scaffold

721   ordering based on multiple maps. Genome Biol. 16: 3.

722   Voorrips R. E., 2002 MapChart: software for the graphical presentation of linkage maps and

723   QTLs. J. Hered. 93: 77–78.

724   Wiley C., Ellison C. K., Shaw K. L., 2012 Widespread genetic linkage of mating signals and

725   preferences in the Hawaiian cricket Laupala. Proc. R. Soc. B Biol. Sci. 279: 1203–1209.

726   Wilfert L., Gadau J., Schmid-Hempel P., 2007 Variation in genomic recombination rates among

727   animal taxa and the case of social insects. Heredity (Edinb). 98: 189–197.

728   Wilson M. A., Makova K. D., 2009 Genomic analyses of sex chromosome evolution. Annu. Rev.

729   Genomics Hum. Genet. 10: 333–354.

730   Wolf J., Lindell J., Backström N., 2010 Speciation genetics: current status and evolving

731   approaches. Philos. Trans. R. Soc. B Biol. Sci. 365: 1717–1733.

732   Wolf J. B. W., Ellegren H., 2016 Making sense of genomic islands of differentiation in light of

733   speciation. Nat. Publ. Gr. 18: 87–100.

734

735   **FIGURE LEGENDS**

736   Figure 1. Study design. (A) shows the phylogenetic relationships of studied *Laupala* species

737   based on a neighbor joining tree generated from genetic distances among the parental lines

738   used in this study. Dashed grey lines connect species pairs that were crossed. (B) Approximate

739   sampling location of the studied species on the Big Island of Hawaii.

740

741   Figure 2. Initial linkage maps. Bars represent linkage groups (LG) for ParKoh, ParKon, and

742   PruKoh. Lines within the bars indicate marker positions. The scale on the left measures marker

743   spacing in centimorgans (cM). Blue lines connect markers on the same scaffold between the

744   different maps. The map for ParKoh is shown twice to facilitate comparison across all three

745   maps. See Fig S1 for comprehensive maps.

746

747   Figure 3. Segregation distortion. For each of the seven autosomal linkage groups within the three

748   comprehensive linkage maps (from top to bottom: ParKoh, ParKon, PruKoh), a sliding window

749   of the negative log-transformed P-values for the χ^2-square test for deviation from a 1:2:1

750   segregation ratio is shown across markers with black lines in the top panels. In the panel below,

751   the trace of the frequency of heterozygote genotypes (blue lines) and homozygote genotypes for

752   both parental alleles (black and red lines, respectively) is shown. For each intercross, dashed

753   grey lines indicate $P = 0.01$ (top panels) or expected allele frequencies based on 1:2:1 inheritance

754   (bottom panels).

755   Figure 4. Recombination map. Black symbols and lines indicate the relationship between the

756   physical distance (scaffold midposition) in million base pairs on the x-axis and the genetic

757   distance in centimorgans for each of the 8 linkage groups on the left y-axis. Open dots represent

758   the dense ParKoh linkage map, triangles and diamonds that of the ParKon and PruKoh cross,

759   respectively. Black lines (ParKoh: solid, ParKon: dashed, PruKoh: dotted) indicate the fitted

760   smoothing spline (10 degrees of freedom). The red lines (same stroke style annotation) show the

761   first order derivative of the fitted splines and represent the variation in recombination rate (in cM

762   per Mp, on the right y-axis) as a function of physical distance. Grey bars indicate the

763   approximate location of male song rhythm QTL peaks. The yellow start in linkage group 2

764   highlights the QTL peak that co-localizes with a female preference QTL peak (Shaw & Lesnick

765   2009). The numbering of the linkage group in the 2009 study is given in grey text.
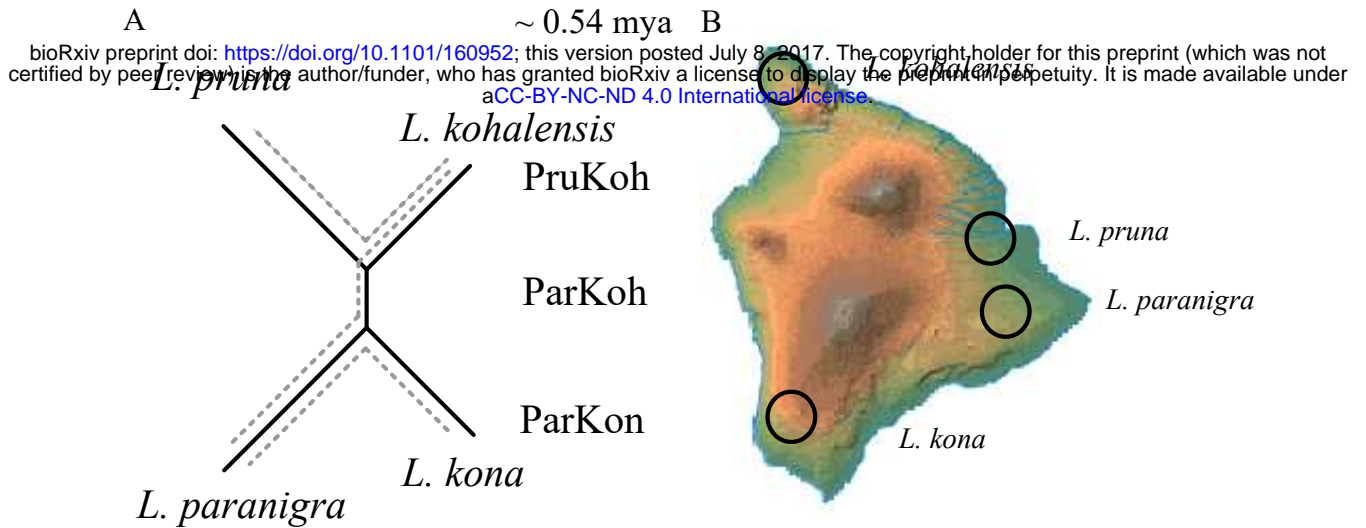
766

767

768

Figure 1. Study design. (A) shows the phylogenetic relationships of studied *Laupala* species based on a neighbourjoining tree generated from genetic distances among the parental lines used in this study. Dashed grey lines connect species pairs that were crossed. (B) Approximate sampling location of the studied species on the Big Island of Hawaii.
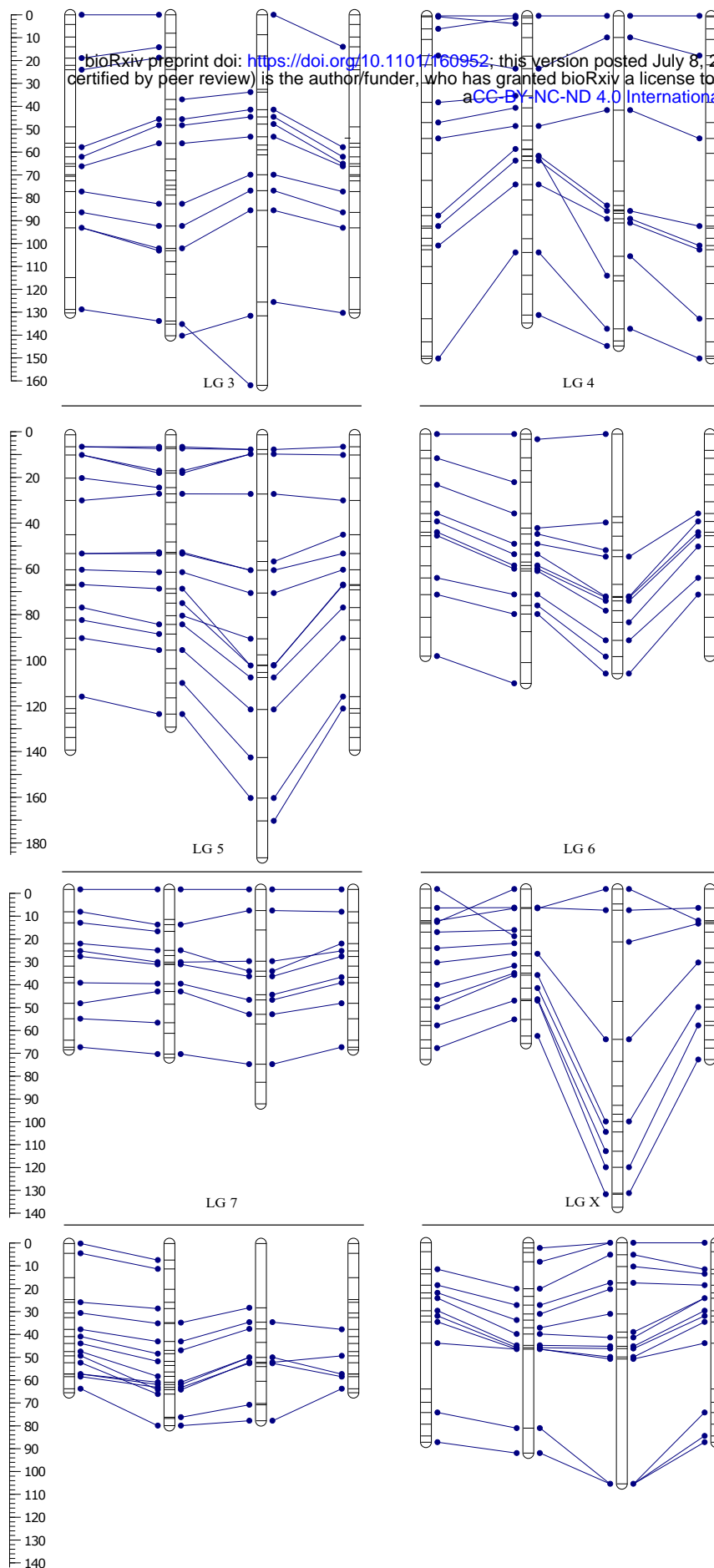
Figure 2. Initial linkage maps. Bars represent linkage groups (LG) for ParKoh, ParKon, and PruKoh. Lines within the bars indicate marker positions. The scale on the left measures marker spacing in centimorgans (cM). Blue lines connect markers on the same scaffold between the different maps. The map for ParKoh is shown twice to facilitate comparison across all three maps. See Fig S1 for comprehensive maps.
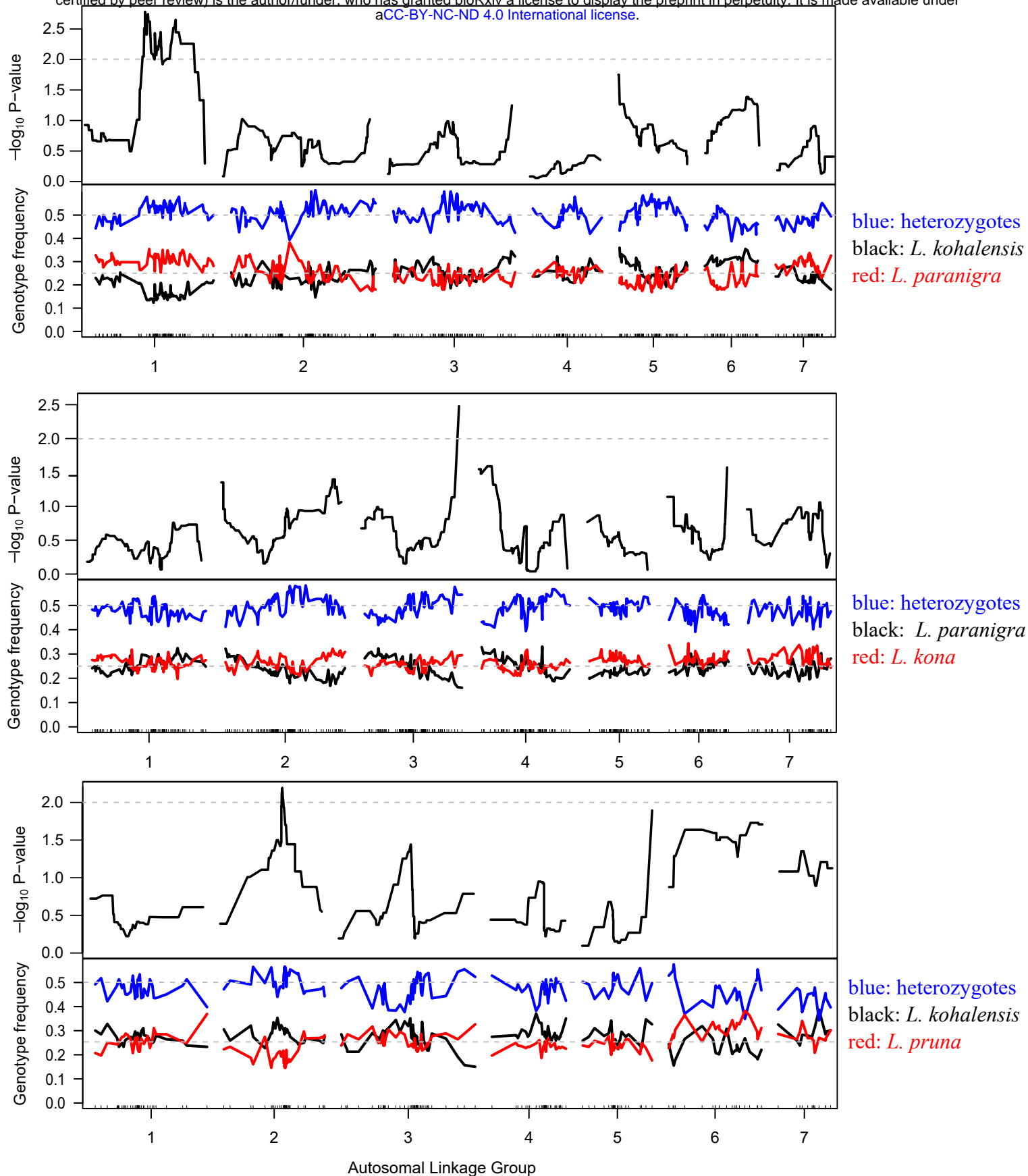
Figure 3. Segregation distortion. For each of the seven autosomal linkage groups within the three comprehensive linkage maps (from top to bottom: ParKoh, ParKon, PruKoh), a sliding window of the negative log-transformed P-values for the χ^2-square test for deviation from a 1:2:1 segregation ratio is shown across markers with black lines in the top panel. In the panel below, the trace of the frequency of heterozygote genotypes (blue lines) and homozygote genotypes for both parental alleles (black and red lines, respectively) is shown. For each intercross, dashed grey lines indicate $P = 0.01$ (top) or expected allele frequencies based on 1:2:1 inheritance (bottom).
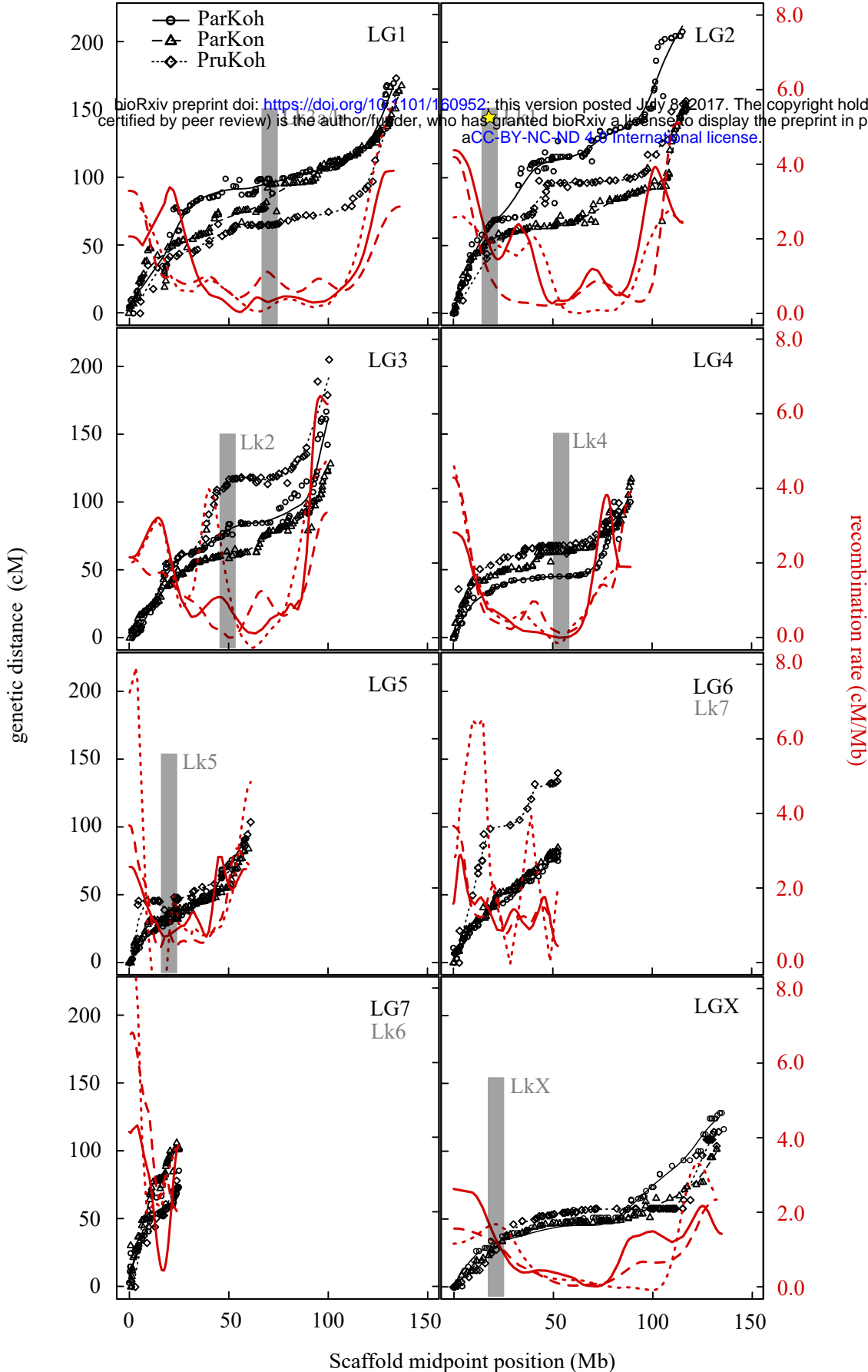
Figure 4. Recombination map. Black symbols and lines indicate the relationship between the physical distance (scaffold midposition) in million base pairs on the x-axis and the genetic distance in centimorgans for each of the 8 linkage groups on the left y-axis. Open dots represent the dense ParKoh linkage map, triangles and diamonds that of the ParKon and PruKoh cross, respectively. Black lines (ParKoh: solid, ParKon: dashed, PruKoh; dotted) indicate the fitted smoothing spline (10 degrees of freedom). The red lines (same stroke style annotation) show the first order derivative of the fitted splines and represent the variation in recombination rate (in cM per Mp, on the right y-axis) as a function of physical distance. Grey bars indicate the approximate location of male song rhythm QTL peaks. The yellow star (LG 2) indicates the co-localizing female preference QTL peak (Shaw & Lesnick 2009). The numbering of the linkage groups in the 2009 study is given in grey text.
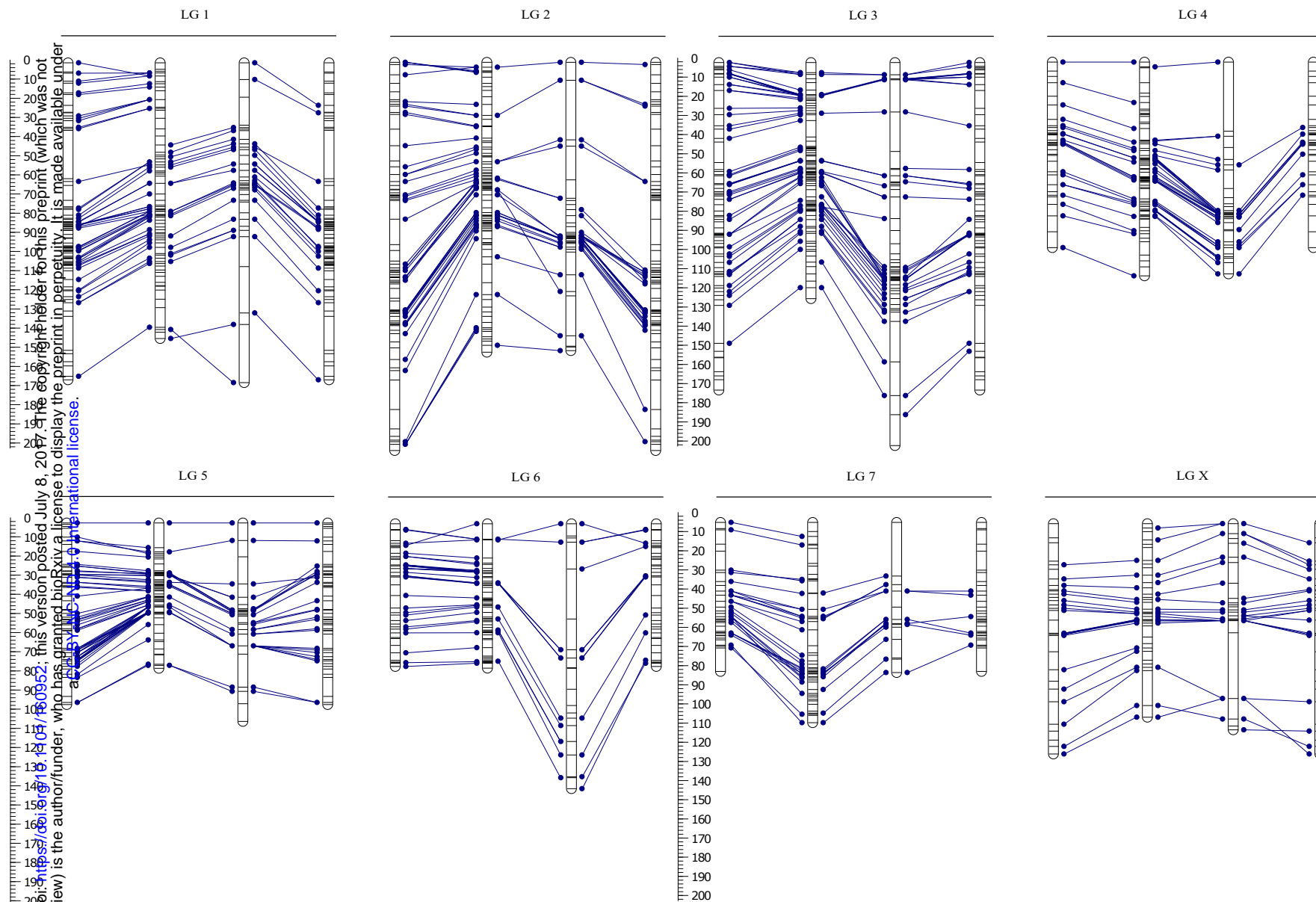
Figure S1. Comprehensive linkage maps. Bars represent linkage groups (LG) for ParKoh, ParKon, and ParKoh. Lines within the bars indicate marker positions. The scale on the left measures marker spacing in centimorgans (cM). Blue lines connect markers on the same scaffold between the different maps. The map for ParKoh is shown twice to facilitate comparison across all three maps.