The Genetics of Mating Song Evolution Underlying Rapid Speciation: Linking Quantitative Variation to Candidate Genes for Behavioral Isolation

Mingzi Xu[1] and Kerry L. Shaw
Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853
ORCID IDs: 0000-0001-7947-9138 (M.X.); 0000-0001-8300-0425 (K.L.S.)

ABSTRACT Differences in mating behaviors evolve early during speciation, eventually contributing to reproductive barriers between species. Knowledge of the genetic and genomic basis of these behaviors is therefore integral to a causal understanding of speciation. Acoustic behaviors are often part of the mating ritual in animal species. The temporal rhythms of mating songs are notably species-specific in many vertebrates and arthropods and often underlie assortative mating. Despite discoveries of mutations that disrupt the temporal rhythm of these songs, we know surprisingly little about genes affecting naturally occurring variation in the temporal pattern of singing behavior. In the rapidly speciating Hawaiian cricket genus *Laupala*, the striking species variation in song rhythms constitutes a behavioral barrier to reproduction between species. Here, we mapped the largest-effect locus underlying interspecific variation in song rhythm between two *Laupala* species to a narrow genomic region, wherein we find no known candidate genes affecting song temporal rhythm in *Drosophila*. Whole-genome sequencing, gene prediction, and functional annotation of this region reveal an exciting and promising candidate gene, the putative cyclic nucleotide-gated ion channel-like gene, for natural variation in mating behavior. Identification and molecular characterization of the candidate gene reveals a nonsynonymous mutation in a conserved binding domain, suggesting that ion channels are important targets of selection on rhythmic signaling during establishment of behavioral isolation and rapid speciation.

KEYWORDS Interspecific variation; behavioral barrier; genetic architecture; cyclic nucleotide-gated ion channel; *Laupala*

SPECIATION can arise from divergence in reproductive phenotypes (Coyne and Orr 2004). Divergent mating behaviors can result in reproductive barriers by causing assortative mating within incipient species. It is well documented that mating behaviors and morphologies diverge early in the speciation process, suggesting an explanation for why prezygotic barriers evolve sooner than postzygotic barriers in the origin of species (*e.g.*, Mendelson 2003; Sánchez-Guillén *et al.* 2014). Moreover, some of the most rapid speciation rates known, such as those in Lake Victoria cichlid fish (Seehausen *et al.* 2008), Hawaiian *Laupala* crickets (Mendelson and Shaw 2005), Baltic Sea European flounders (Momigliano *et al.* 2017), and a putative case in Galapagos finches (Lamichhaney *et al.* 2018), occur when species diverge in mating behaviors and associated structures. Thus, studying the genetics and evolution of behavioral barriers can contribute to an emerging general principle of speciation.

Because evolution is a genetic process, characterizing the genetic architecture and identifying genes involved in behavioral barriers is crucial to understanding targets of selection and establishing causal links among genes, pathways, and mating behaviors in the early stages of speciation. In animals, however, courtship is often complex and multimodal, involving many traits (*e.g.*, Greenspan and Ferveur 2000; Rundus *et al.* 2010; Starnberger *et al.* 2014; Ullrich *et al.* 2016; Mowles *et al.* 2017). Accordingly, it can be difficult to isolate specific behaviors for genetic analysis. Perhaps because of these complexities, we have a limited understanding of the evolutionary genetics of mating behaviors that contribute to reproductive barriers despite its general importance in speciation.

doi: https://doi.org/10.1534/genetics.118.301706
Manuscript received October 16, 2018; accepted for publication January 11, 2019; published Early Online January 15, 2019.
Supplemental material available at Figshare: https://doi.org/10.25386/genetics.7505762.
[1]Corresponding author: Department of Neurobiology and Behavior, Cornell University, 215 Tower Rd., Ithaca, NY 14853. E-mail: mx52@cornell.edu

While some progress has been made in understanding the genetic basis of natural variation in visual and olfactory signals, such as cuticular hydrocarbons, sex pheromones, and body coloration [*e.g.*, Gleason *et al.* 2005, 2009; Kronforst *et al.* 2006; Sæther *et al.* 2007; Lassance *et al.* 2010, 2013; Merrill *et al.* 2011; Niehuis *et al.* 2011; Bay *et al.* 2017; also reviewed by Groot *et al.* (2016)], many organisms use acoustic signals involving rhythmic neuromuscular behaviors for which we still have a very limited genetic understanding. Even in *Drosophila*, where acoustic behavior is expressed widely in courtship, we lack a gene-based understanding of natural variation (but see Gleason and Ritchie 2004; Ding *et al.* 2016). Rhythmic, temporal patterns of such mating "songs" are often species-specific and known components of reproductive barriers among species of insects, fish and amphibians (Gerhardt and Huber 2002; Hartbauer and Römer 2016; Barkan *et al.* 2017; Smith *et al.* 2018). The rhythmic elements of song are a result of regularly patterned motor output, products of localized, neural circuits called central pattern generators (CPGs; Chagnaud and Bass 2014; Katz 2016; Schöneich and Hedwig 2017). Compared with other rhythmic mating behaviors such as courtship dance, song rhythms are easy to isolate and measure.

To date, genetic studies of song rhythm variation have revealed a polygenic genetic architecture in insects, including fruit flies, lacewings, crickets, grasshoppers, and moths (Shaw 1996; Williams *et al.* 2001; Henry *et al.* 2002; Gleason and Ritchie 2004; Saldamando *et al.* 2005; Shaw *et al.* 2007; Ellison *et al.* 2011; Limousin *et al.* 2012). However, the causal genes underlying natural variation remain elusive in most cases. Eleven candidate genes that regulate interpulse interval in *Drosophila melanogaster*, including ion channel genes, transcription factors, and transcription/translation regulators, have been identified through experimentally generated mutations [reviewed by Gleason (2005), also see Turner *et al.* (2013) and Fedotov *et al.* (2014, 2018)]. These discoveries offer insight into the types of genes capable of modulating song rhythmicity in naturally occurring systems and thus are reasonable candidate genes for interspecific variation in other singing insects.

Here, we investigate the genetic and genomic basis of natural variation in pulse rate (the inverse of pulse duration, Figure 1) of the male mating song of the endemic Hawaiian cricket *Laupala*. In *Laupala*, a rapid radiation has resulted in 38 morphologically and ecologically similar, but acoustically distinctive species (Otte 1994; Mendelson and Shaw 2005). Similar to most crickets, males sing long-range "calling" songs to attract females. Male songs of *Laupala* are characterized by simple trains of pulses delivered at species-specific rates (Otte 1994; Shaw 2000). Evidence shows that divergent pulse rate partially mediates mate choice (Shaw and Herlihy 2000; Oh and Shaw 2013) and constitutes a reproductive barrier between species (Mendelson and Shaw 2002).

We focus on two closely related species, the slow-calling *Laupala paranigra* (0.71 pulses per second; pps) and the fast-calling *Laupala kohalensis* (3.72 pps; Shaw *et al.* 2007; Figure 1), who have diverged roughly 0.43 MYA on the Big Island of Hawaii (Mendelson and Shaw 2005). Roughly 74% of the difference in pulse rate between these species is due to eight small- to medium-effect, additive QTL (Shaw *et al.* 2007). In this study, we isolate the largest-effect QTL, examine its inheritance and phenotypic effects, fine-map its location, and test hypotheses about the underlying causal gene or genes. Specifically, we hypothesized that the focal QTL region would include one of several candidate genes for interpulse interval variation in *D. melanogaster* [Turner *et al.* 2013; Fedotov *et al.* 2014, 2018; also reviewed by Gleason (2005)]. Although the specific forms of fly and cricket songs differ, both have species-specific, rhythmic features resulting from regular contractions of thoracic muscles that drive wing movements. Alternatively, the causal gene underlying the focal QTL in our study may be a previously unknown gene. Based on the functional categories of *Drosophila* candidate genes and biological processes involved in singing, we hypothesize that the pulse rate QTL falls in one of three functional categories: ion transportation, neural modulation and development, and locomotion and muscle development. It may also be the case that regulatory genes affecting the expression of genes belonging to these categories are responsible for the pulse rate variation. We test these hypotheses by conducting whole-genome sequencing (WGS), gene prediction, and functional annotation in the genomic region linked to the focal QTL.

Materials and Methods

Breeding design

The breeding design included two stages: isolating the focal QTL (QTL4) in near isogenic lines [NILs; for details see Wiley *et al.* (2012), Ellison and Shaw (2013) and Supplemental Material, File S1] and generation of NIL mapping populations. We created three replicate NILs (4B, 4C, and 4E) by selectively backcrossing individuals with *L. paranigra* (the slow singer) alleles at the marker linked to QTL4 to *L. kohalensis* (the fast singer) for four generations (Figure 2). The fourth-generation backcross individuals were then intercrossed to generate NIL lines, where nonrecurrent QTL4 homozygotes were maintained through intercrossing thereafter. Mapping populations were generated by backcrossing one 7–9th generation NIL male with one *L. kohalensis* female from the same recurrent line used to generate the NILs, with replication in each NIL (replicates: NIL4B: $n = 3$, NIL4C: $n = 2$, NIL4E: $n = 1$). The offspring were intercrossed, resulting in three, two, and one NIL-*L. kohalensis* backcross F_2 (hereafter, F_2) mapping families for 4B, 4C, and 4E, respectively (denoted as family 4B.1, 4B.2, 4B.3, 4C.5, 4C.9, and 4E.1). We maintained the *L. kohalensis* recurrent and NIL lines alongside the mapping populations.

All crickets were reared individually in 120 ml specimen cups with a piece of moist tissue and fed *ad libitum* Organix organic chicken and brown rice dry cat food (Castor & Pollux

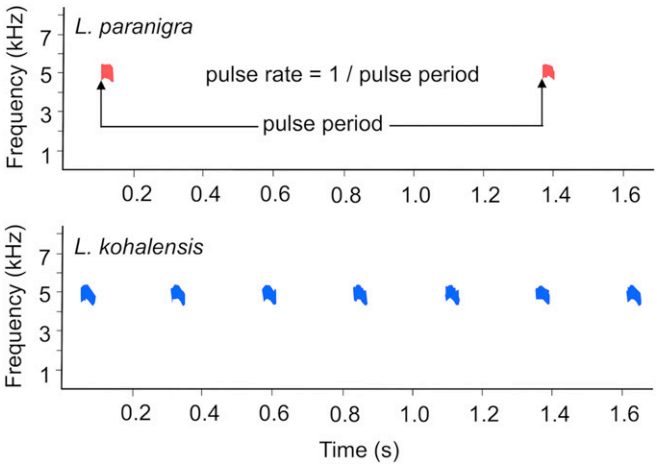

Figure 1 Sonograms of the male calling songs of the slow-calling *Laupala paranigra* and the fast-calling *Laupala kohalensis* showing measurement of pulse rate.

Natural Petworks, Clackamas, OR) twice per week in a rearing room held at 20.0° and a 12:12 hr light:dark cycle.

Phenotyping

We recorded male songs with an Olympus WS-852 digital stereo recorder (Olympus Imaging Corp., Tokyo, Japan) during daylight hours in a temperature-controlled room (20.3 ± 0.01°, mean ± SE, $n = 824$). Digital sound files were analyzed using RavenPro 1.4 (http://ravensoundsoftware.com). Pulse period was measured as the time differential between the beginnings of two consecutive pulses (Figure 1). Mean pulse period was calculated from five independent pulse period measurements from a single song bout; mean pulse period was transformed to pulse rate (pps) by taking the inverse of the pulse period.

Genotyping

We obtained SNP-based genotypes from F_2 mapping populations for linkage map estimation and QTL mapping using genotyping-by-sequencing (GBS; Elshire *et al.* 2011). Sequencing was conducted on the Illumina HiSeq 2000 platform at the Genomic Diversity Facility at Cornell University (see Supplemental File S1 for details of DNA extraction, library preparation, and sequencing). Sequencing reads were demultiplexed with fastq-multx v.1.3.2, base calls at the ends of reads with Phred scrore <30 were trimmed, and reads <50 bases long were removed with fastq-mcf v.1.04.636 (Aronesty 2011). Processed reads were aligned to the *L. kohalensis* genome reference (NCBI genome assembly ASM231320v1; Blankers *et al.* 2018a) using Bowtie2 v.2.2.6 (Langmead and Salzberg 2012) with default parameter settings. We called single nucleotide polymorphism (SNP) variants within each F_2 family, allowing for a maximum of two mismatches per mapped read using FreeBayes v.0.9.12-2-ga830efd (Garrison and Marth 2012). The resulting SNP markers were filtered using VCFtools 0.1.15 (Danecek *et al.* 2011) and vcffilter in vcflib v.1.0.0 (Garrison 2012). We retained bi-allelic SNP markers that fulfill the following criteria: (1) <20% missing data per family, (2) minor allele frequency ≥2.5%, (3) genotype depth ≥5, (4) Phred scaled variant quality ≥30, and (5) strand balance probability for reference and alternative alleles >0.0001. Because we did not have the sequences of the parental NIL and *L. kohalensis* individuals from families 4C.5, 4C.9, and 4E.1, F_2 genotypes were called using the *L. kohalensis* genome reference as the *L. kohalensis* parent; the alternative allele was assigned to the NIL parent. For replicate 4B where we additionally sequenced parental NIL and *L. kohalensis* individuals, we further retained SNPs only if the parental *L. kohalensis* base call was the same as the *L. kohalensis* genome reference and the parental NIL was homozygous for the alternative allele. F_2 genotypes for 4B families were called using genotypes of the NIL and *L. kohalensis* grandparents.

Because GBS is a reduced-representation sequencing method that does not reveal contiguous sequence information of a genomic region, we conducted WGS from two male intercross offspring of BC_4 in line 4E (Figure 2). We selected males whose pulse rates suggested that they were homozygous in the introgressed QTL4 region for either *L. paranigra* or *L. kohalensis* alleles. The depth of coverage for WGS was roughly 18× and 12× for the slow- and fast-singing males, respectively. We identified alternative homozygous SNPs and indels (insertions and deletions) between *L. paranigra* and *L. kohalensis* across the introgressed QTL4 region, and used the presence of such SNPs and indels as an indication of increased probability of causation when we evaluated *Drosophila* song candidate genes and the predicted *Laupala* genes in this region as the potential causal gene. Paired-end sequencing with an insert size of 200 bp was conducted on the Illumina HiSeq 2000 platform. Parameter values for quality control, read mapping, and variant calling were as used for GBS data. Two sets of SNPs in the final WGS data were of interest: (1) SNPs yielding a homozygous genotype in the fast-calling male that was identical to the reference, and likewise homozygous for the alternative allele in the slow-calling male; and (2) SNPs with no read mapped for the fast-calling male, but with a homozygous alternative genotype (to the reference) for the slow-calling male. SNPs in the first set were filtered with the same criteria as were GBS SNPs except that we allowed no missing genotypes and minor allele frequency ≥24%. For SNPs in the second set where no SNP calls could be made for the fast-calling male, the missing data, genotype depth, and alternative variant quality filters were only applied to the slow-calling male. After quality filtering, SNPs in the second set made up 1.4% of total WGS SNPs. Indels were filtered by the same criteria as SNPs.

Linkage mapping

Linkage mapping for autosomal linkage groups was conducted with genotypes of both F_2 males and females in Joinmap 4 (Van Ooijen 2006), excluding markers that deviated from a segregation ratio of 1:2:1 (Benjamini–Hochberg adjusted $P < 0.05$) and whose mean depth of coverage

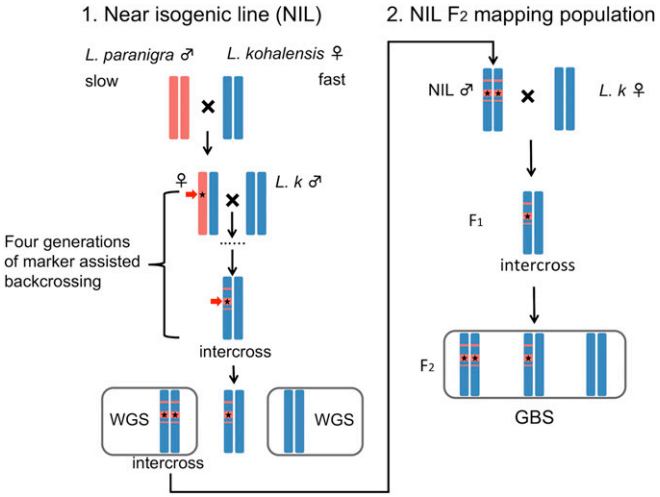

Figure 2 Schematic figure of the two-step breeding design for QTL fine-mapping of male song pulse rate variation between the slow-calling *Laupala paranigra* and the fast-calling *Laupala kohalensis* on linkage group 5 (represented by red and blue bars). In step 1, following the introgression of pure species isofemale lines, near isogenic lines (NILs) were achieved through four generations of marker assisted backcrossing (indicated by the red arrow) selecting for individuals carrying the *L. paranigra* allele at the genetic marker linked to QTL4 in Shaw *et al.* (2007) (indicated by the black star) and one generation of intercross. Three independent NIL replicates (NIL4B, 4C, and 4E) were established after the intercross. In step 2, seventh to ninth generation NIL males were backcrossed to *L. kohalensis* females to generate segregating F_2 mapping populations within each NIL replicate. Individuals used for genotyping-by-sequencing and whole-genome sequencing were indicated with GBS and WGS, respectively. In replicate 4B only, the NIL and *L. kohalensis* grandparents of the F_2 mapping populations were also sequenced by GBS (not labeled in the figure).

was <20. When there are multiple markers on the same scaffold, only one marker was retained per 200 kb. Individuals with ≥20% missing genotype were excluded from linkage mapping.

For each family, markers were grouped using an independent logarithm of the odds (LOD) threshold of 4. Linkage group 5 (LG5) was easily identified by shared markers with previous studies (Blankers *et al.* 2018a,b). Maps were estimated using the regression algorithm with Kosambi mapping function (details and parameter values are available in File S1).

QTL mapping

Males with no missing phenotypes and <20% missing genotypes on LG5 were used for QTL mapping. We performed standard interval mapping (SIM), two-QTL scan, and multiple QTL mapping (MQM) in each family. MQM was conducted using forward selection with backward elimination. LOD thresholds for SIM were calculated from 20,000 permutations using the maximum likelihood method and LOD thresholds for main and interaction terms in MQM were calculated from 1000 permutations using Haley–Knott regression. We used Haley–Knott regression instead of multiple imputation due to computational constraints. We estimated effect sizes and 1.5-LOD support confidence intervals of significant QTL (see File S1 for further details). All QTL mapping analyses were conducted in R/qtl v.1.39-5 (Broman *et al.* 2003). Because all two-QTL scan results were consistent with the final MQM results, we only report results from SIM and MQM.

For families where the focal QTL was detected on LG5, we tested whether the phenotypic distribution of F_2 males deviates from Mendelian segregation with chi-squared tests. If the primary additive QTL4 has been cleanly isolated, we expect F_2 individuals to show 1:2:1 phenotypic segregation. We binned the F_2 phenotype data by dividing the range of the phenotypic values evenly into three bins between the maximum and the minimum pulse rate. The Bonferroni-corrected significance level was 0.017.

Identification of D. melanogaster song candidate homologs

To test the hypothesis that the *Laupala* QTL 4 region contains one or more homologs of *D. melanogaster* song candidate genes, we first identified homologs of *Drosophila* song candidate genes in the *L. kohalensis* genome using a reciprocal blast strategy. We identified 11 genes in *Drosophila* with experimentally confirmed effects on mean interpulse interval from two sources: a review yielding six genes (Gleason 2005), and a literature search using two combinations of three search terms, including "*Drosophila*," "courtship song" and "gene" or "*Drosophila*," "courtship song" and "genetic," from 2005 to present using Google Scholar, yielding five additional genes (Turner *et al.* 2013; Fedotov *et al.* 2014, 2018; Table 1). Sequences of these 11 genes were blasted against the *L. kohalensis* reference genome using tblastx in Blast+ (Camacho *et al.* 2009), and a new *Laupala cerasina* transcriptome assembled herein (see File S1) was blasted against the *D. melanogaster* protein database using tblastn at an E-value cutoff of 10.

We then obtained the conserved domains of these genes from the conserved domain database at NCBI (Marchler-Bauer *et al.* 2015) using default settings and confirmed the detected domains using the HomoloGene database at NCBI (Table S1). We inferred homology between *L. kohalensis* sequence and a described *D. melonagaster* gene if (1) the *L. kohalensis* and the *D. melanogaster* genes were mutual best blast hits in the reciprocal blast, (2) the matched sequences were in synteny, and (3) at least half of the conserved domains of a given *D. melanogaster* gene contain blast hits matching the *L. kohalensis* query. In cases where the best blast hit in one direction is not the best hit in the other direction but was within the top five best matches, we designated the sequence with the longest match, the most conserved domains and with complete synteny as the *Laupala* homolog.

We evaluated the likelihood of putative *Drosophila* homologs as the causal gene underlying the focal QTL by two criteria: (1) its hosting scaffold resides within the 1.5-LOD confidence interval of QTL4, and if so, (2) the gene and the putative regulatory region 20 kb up- and downstream from

Table 1 Experimentally verified candidate genes in *Drosophila melanogaster* that affect mean interpulse interval of the courtship song, the location and linkage group assignment of their putative homologs in the *Laupala kohalensis* genome, and homology and synteny between the *Drosophila* candidate genes and putative *Laupala* homologs

Gene	Abbreviation	Putative homolog location	Linkage group	# CD in gene	# CD with alignments	% CD coverage	E-value	Alignment synteny
cacophony	*cac*	S000496:650433-736176	4	8	8	100.0	0.00E+00	Yes
cacophony	*cac*	S000703:661226-714877	x	8	8	93.4	9.39E−157	Partial
CG15630	CG15630	S003101:147058-154537	5	2	2	48.4	8.44E−02	Yes
CG6746	CG6746	S001095:330297-327934	Unknown	1	1	97.7	1.04E−46	Yes
fruitless	*fru*	S000605:104164-163750	2	2	2	85.6	2.33E−145	Yes
maleless	*mle*	S003080:411383-438168	3	3	3	100.0	0.00E+00	Yes
Myocyte enhancer factor 2	*Mef2*	S001895:95733-110761	7	2	2	97.3	4.72E−74	Yes
paralytic	*para*	S002404:402957-487077	3	19	16	63.4	0.00E+00	Yes
Selenophosphate synthetase 2	*Sps2*	S005967:125411-126211	Unknown	2	2	67.1	3.14E−64	Yes
slowpoke	*slo*	S000206:909-126847	3	3	3	49.8	0.00E+00	Yes
Syntrophin-like 1	*Syn1*	S000296:874002-902774	x	4	3	65.3	6.17E−108	Yes
CG34460	CG34460	–	–	1	0	0.0		–

The number of conserved domains (CDs), the proportion of total CD length covered by significant alignments, the E value and alignment synteny were from a tblastx between the *D. melanogaster* candidate genes and the *L. kohalensis* genome

the gene contain at least one WGS SNP or indel between the alternative homozygote males from line 4E. To decrease the potential for a false negative, we used a 20 kb up- and downstream region as the size of the putative regulatory region, rather than the conventional 5 kb window. We assigned scaffolds hosting *Drosophila* homologs to *Laupala* linkage groups estimated in this and another study where linkage maps of three intercrosses of *Laupala* species pairs were constructed (Blankers *et al.* 2018a). We then examined whether any scaffold hosting a putative *Drosophila* homolog resides within the 1.5-LOD confidence intervals of the major QTL on LG5. Putative homologs that could not be assigned to a linkage group (*i.e.*, their hosting scaffolds are not in any available linkage maps) were evaluated solely on the basis of whether they contained WGS SNPs or indels.

Lastly, to investigate if any scaffold within the confidence interval of QTL4 contains divergent paralogs of the *Drosophila* candidate genes that escaped detection by the method above, we blasted the sequences of conserved domains in the *Drosophila* candidate genes against scaffolds within the confidence interval of the QTL using an E-value cutoff of 1E−5. A diverged paralog would be expected to contain at least one significant blast hit in the conserved domains. Conversely, failure to see a significant blast hit indicates the absence of orthologs or paralogs of *Drosophila* candidate genes.

Gene prediction and functional annotation

To identify *L. kohalensis* candidate genes causing pulse rate variation on LG5, we annotated scaffolds within and flanking the 1.5-LOD confidence interval of the major QTL in 4C.9 (the most robust QTL identified and the family with the highest sample size). We chose the 1.5-LOD confidence interval by convention, yet because the LOD profile of the final QTL model drops sharply on both sides of the major QTL peak, the scaffolds included in the 1.5-LOD confidence interval are identical to those up to a 5-LOD confidence interval. We therefore consider these scaffolds sufficiently inclusive for our candidate gene search. Gene prediction was done using the Maker pipeline (Cantarel *et al.* 2008) with available RNA, EST, and protein evidence (Danley *et al.* 2007; Bailey *et al.* 2013; Zeng *et al.* 2013; Berdan *et al.* 2016). Detailed information on gene prediction can be found in the File S1. Briefly, we performed two rounds of training with SNAP (Korf 2004) and Augustus-3.2.3 (Stanke and Morgenstern 2005) with the 10 longest scaffolds and the five focal scaffolds in the *L. kohalensis* reference genome in a bootstrap manner to obtain *Laupala*-specific gene model predictions. The second-round gene model outputs from both SNAP and Augustus were used to predict gene structures on the five focal scaffolds in the third round.

To examine predicted genes, we performed functional annotations in Blast2GO 4.1 (Götz *et al.* 2008). Specifically, we blasted sequences of all predicted genes against the NCBI nonredundant protein database using an E-value cutoff of 1E−4 as well as InterPro (Finn *et al.* 2016) with the default setting, and inferred the potential gene identity with the automatic annotation description function and manual curation. We then annotated gene ontologies (Ashburner *et al.* 2000) for predicted genes using NCBI and InterPro blast results as well as UniProt (Apweiler *et al.* 2004) and KEGG databases (Kanehisa *et al.* 2016) with default parameter values in Blast2GO. To interrogate further the identities of any remaining predicted genes lacking an annotation, we reran blast for these sequences using a relaxed E-value cutoff of 100.

To further implicate candidate genes as the cause of pulse rate variation, we tallied WGS SNPs and indels in the QTL4 region (within annotated genes and 5 kb up- and downstream) from the two alternative homozygous 4E males using

SnpEff (Cingolani *et al.* 2012) and manually, respectively. We evaluated the possibility of a causal role for a given predicted gene in pulse rate variation by three criteria: (1) the gene maps to a scaffold within the 1.5-LOD confidence interval, (2) the gene has at least one WGS SNP or indel within the coding or (putative) regulatory region, and (3) orthologs or paralogs of the gene play a functional role in rhythmic movements in other organisms. For the most promising candidate gene fulfilling all three criteria, we confirm the identity of the predicted gene by both a more focused sequence alignment using protein products of representative members of the indicated gene family in Exonerate 2.2.0 (Slater and Birney 2005) and phylogenetic inference using the amino acid sequences of the conserved domains in these proteins in PhyML 3.0 (Guindon and Gascuel 2003). We further determined WGS SNPs and indels as regulatory, intronic, synonymous, or nonsynonymous. For details of the gene identity inference and SNP/indel effect annotation, see File S1. We evaluated the effects of any resulting amino acid substitutions in PROVEAN Protein (Choi and Chan 2015).

Data availability

All DNA and RNA sequences, the *L. cerasina* transcriptome, and the associated metadata are available at NCBI (DNA and RNA sequences: PRJNA509479; transcriptome: GHDK00000000). Phenotypic data, SNP genotypes, bioinformatic and QTL mapping scripts, and metadata (in the readme file) are available for download at https://github.com/MingziXu/QTL4-male-fine-mapping-scripts. File S1 contains additional method details and supplemental results; Table S1 contains information about conserved domains in the 11 *D. melanogaster* candidate genes for variation in interpulse interval of courtship song; Table S2 contains summary statistics of the linkage maps; Table S3 contains results from MQM on linkage groups other than LG5; Table S4 contains frequency distributions of the F_2 male pulse rates and chi-squared test results for Mendelian segregation ratios in 4C and 4E families; Table S5 shows results of functional annotation of the 66 predicted genes on the five focal scaffolds within and flanking the 1.5-LOD (same as 5-LOD) support confidence interval of the major QTL peak in 4C.9; Table S6 contains information about conserved domains in CNG and HCN channels used for protein tree construction; Table S7 contains descriptions of top blast hits and the E-values for unannotated genes in Table S5 from a blast using relaxed E-value cutoff of 100; Figure S1 shows the phenotypic distribution in NIL4B; Figure S2 shows homology and synteny of linkage groups among six F_2 families; Figure S3 shows the linkage maps of (a) NIL4C and NIL4E and (b) NIL4B; Figure S4 shows the LOD profile of MQM and SIM models in NIL4B; Figure S5 shows synteny between the linkage map of 18 markers showing potential double recombination patterns on LG5 and the linkage map for LG5 used for QTL mapping in 4C.9; Figure S6 shows alignment between protein sequence of the termite *Zootermopsis nevadensis* cyclic nucleotide-gated olfactory channel-like protein and the putative *Laupala Cngl* homolog. Supplemental material available at Figshare: https://doi.org/10.25386/genetics.7505762.

Results

Phenotypic distribution of L. kohalensis line and NILs

Males from the parental *L. kohalensis* line showed expected *L. kohalensis* pulse rates ($n = 24$, 3.80 ± 0.11 pps, mean $\pm$ SD) and males from the three replicates of NIL showed expected slower pulse rates by substituting the *L. paranigra* allele at QTL4 (Figure 3 and Figure S1; NIL4B: $n = 4$, 3.12 ± 0.13 pps; NIL4C: $n = 7$, 3.07 ± 0.05 pps; NIL4E: $n = 7$, 2.95 ± 0.04 pps).

Linkage mapping

The linkage maps of the six families contained seven autosomal linkage groups where marker orders were highly consistent (Figure 4 and Figure S2). Among all linkage groups, LG5 had the highest number of shared markers among replicates (Figure S2 and Figure S3). Family 4C.9 had the highest number of markers and the highest marker density of all families (Figure 4 and Table S2).

Families 4C.5, 4C.9 and 4E.1 shared a large and dense group of markers in the central part of LG5 that was absent in 4B families (Figure 4, red). We discuss results from 4C/4E and 4B families separately (see below).

QTL mapping in NIL4C and 4E

Using SIM, we estimated the location of a large effect ($\sim$87.25% of F_2 variance) QTL segregating in 4C.9 on LG5. With MQM, we additionally estimated the position of a smaller effect QTL ($\sim$1% of F_2 variance) on LG5 (Figure 5 and Table 2). Together, these two QTL explain 88.3% of the phenotypic variance for pulse rate in 4C.9; no other QTL were detected in this family. Similarly, both a major QTL, identified with both SIM and MQM, and a minor QTL, identified using MQM, were localized on LG5 in 4C.5 (Figure 5 and Table 2), which together explain 87.8% of the variation segregating for pulse rate in this F_2 mapping population (Table 2). In 4C.5, we also detected four additional, small-effect QTL on LG1, LG2, LG3, and LG7 (Table S3). Lastly, we detected a single QTL on LG5 in 4E.1 using SIM and MQM, comparable in size ($\sim$81% of segregating F_2 variance) to those identified in 4C.9 and 4C.5. Several smaller-effect QTL were also identified in 4E.1, localized to LG1, LG3, LG4, and LG7 (Table S3). Concomitant with the introgression pattern revealed by QTL mapping, F_2 generation males from 4C and 4E families each exhibited the classical Mendelian 1:2:1 phenotypic segregation pattern for pulse rate (Figure 3 and Table S4). Hereafter, we refer to the large-effect QTL on LG5 identified in 4C.9, 4C.5, and 4E.1 as the "major" QTL.

The locations of the major QTL on 4C.5, 4C.9, and 4E.1 are highly consistent with each other, as in each family the same SNP (at base location 224,758 on scaffold S000353) is the marker that exhibits the highest LOD score by SIM. In MQM,

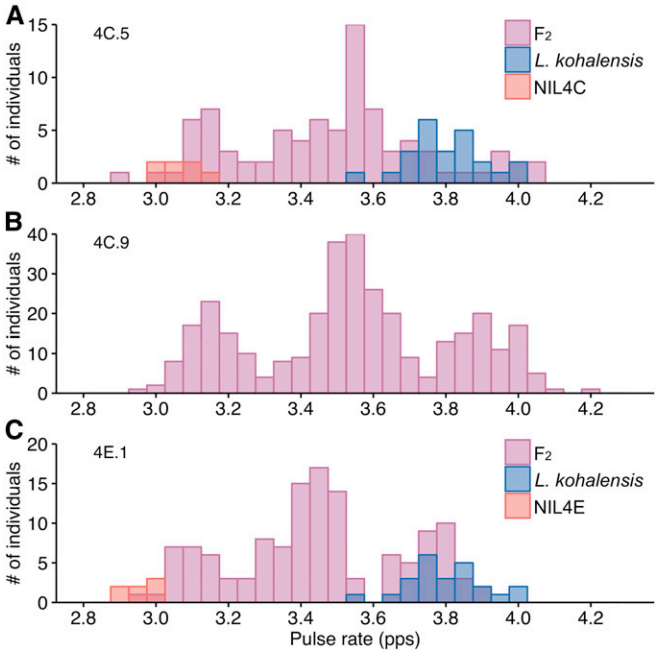

Figure 3 Phenotypic distribution of F_2 males in (A) 4C.5, (B) 4C.9, and (C) 4E.1. The phenotypic distributions of parental *L. kohalensis* and NIL individuals are shown in (A) and (C).

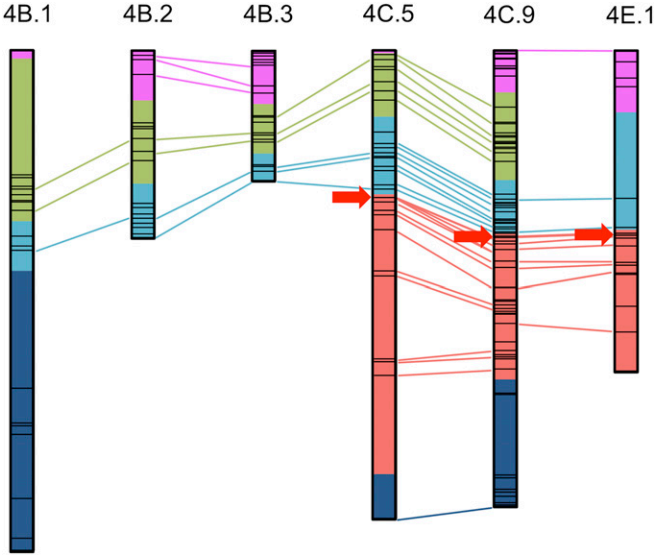

Figure 4 Linkage maps of linkage group 5 in six F_2 mapping families. Different clusters of markers based on their presence and absence in different families are color coded and markers shared between two families are connected with colored lines. The red arrows indicate the positions of the major QTL from the final multiple QTL models. Linkage maps with marker names can be found in Figure S3.

the marker on scaffold S000353 has the highest LOD score in 4C.9 and 4E.1; the SNP at base location 297,109 on scaffold S001839, the scaffold immediately adjacent to S000353, has the highest LOD score in 4C.5 (Table 3). Further, the 1.5-LOD confidence intervals of the major QTL are narrow and largely overlap in 4C and 4E families (Figure 5 and Table 3), containing seven (including five mapping to the same position), five (including two mapping to the same position), and two scaffolds in 4C.5, 4C.9, and 4E.1, respectively.

The major QTL in 4C and 4E families also have similar phenotypic effects (Table 2). In all three families, the phenotypic distributions at the markers with the highest LOD scores showed the expected slow pulse rates for the *L. paranigra* genotype (AA, Figure 5, 4C.5: 3.18 ± 0.02, 4C.9: 3.17 ± 0.01, 4E.1: 3.12 ± 0.01 pps), fast pulse rates for the *L. kohalensis* genotype (BB, 4C.5: 3.86 ± 0.03, 4C.9: 3.90 ± 0.01, 4E.1: 3.75 ± 0.02 pps), and intermediate pulse rates for heterozygotes (4C.5: 3.52 ± 0.02, 4C.9: 3.54 ± 0.01, 4E.1: 3.43 ± 0.01 pps). There was little phenotypic overlap between the genotypes (Figure 5). The phenotypic effect of a single allele at the major QTL on LG5 was predominantly additive in 4C and 4E families, contributing 10–12% of the phenotypic difference between the original parent (*L. kohalensis* and *L. paranigra*) phenotypes (Table 2).

QTL mapping in NIL4B

In both 4B.1 and 4B.2, we detected a minor effect QTL in the beginning of LG5 (Figure S4 and Table 2), but no QTL were detected in the central part of the linkage group near the major QTL found in 4C and 4E families. The introgressed region in 4B families did not include markers within the 1.5-LOD confidence intervals of the major QTL in 4C and 4E families (Figure 4 and Figure 5) and all markers within the confidence interval of 4C and 4E were homozygous for *L. kohalensis* genotypes in 4B families. No QTL were detected on LG5 in 4B.3. We also detected several moderate-effect and primarily additive QTL on LG1 and LG7 in 4B families (Table S3).

Putative homologs of D. melanogaster song candidate genes

Reciprocal blast identified putative homologs for 10 out of 11 candidate genes in the *L. kohalensis* genome (Table 1). These genes had several identifiable conserved domains with high coverage and extensive synteny across matched regions, and in all cases, we identified transcripts that mapped to these genes in the *Laupala* transcriptome. In contrast, CG34460 (Fedotov *et al.* 2014) had no significant match in the *L. kohalensis* genome to the conserved domain region, nor a *Laupala* transcript match, at an E-value cutoff of 10.

Putative *Laupala* homologs of 10 *Drosophila* genes were distributed across six different linkage groups (Table 1). Only one putative homolog (CG15630) groups with LG5 in *Laupala*. However, the scaffold hosting the putative homolog locates between scaffold S001650 and S000820, outside the confidence intervals of both the major and the minor QTL on LG5 (Figure S5 and Table 3). We were unable to assign the putative homologs of CG6746 and *Sps2*, residing on scaffolds S001095 and S005967, respectively, to any linkage group. However, we detected no alternative homozygous WGS SNPs or indels between parental lines within the putative gene or the extended 20 kb putative regulatory regions for these two

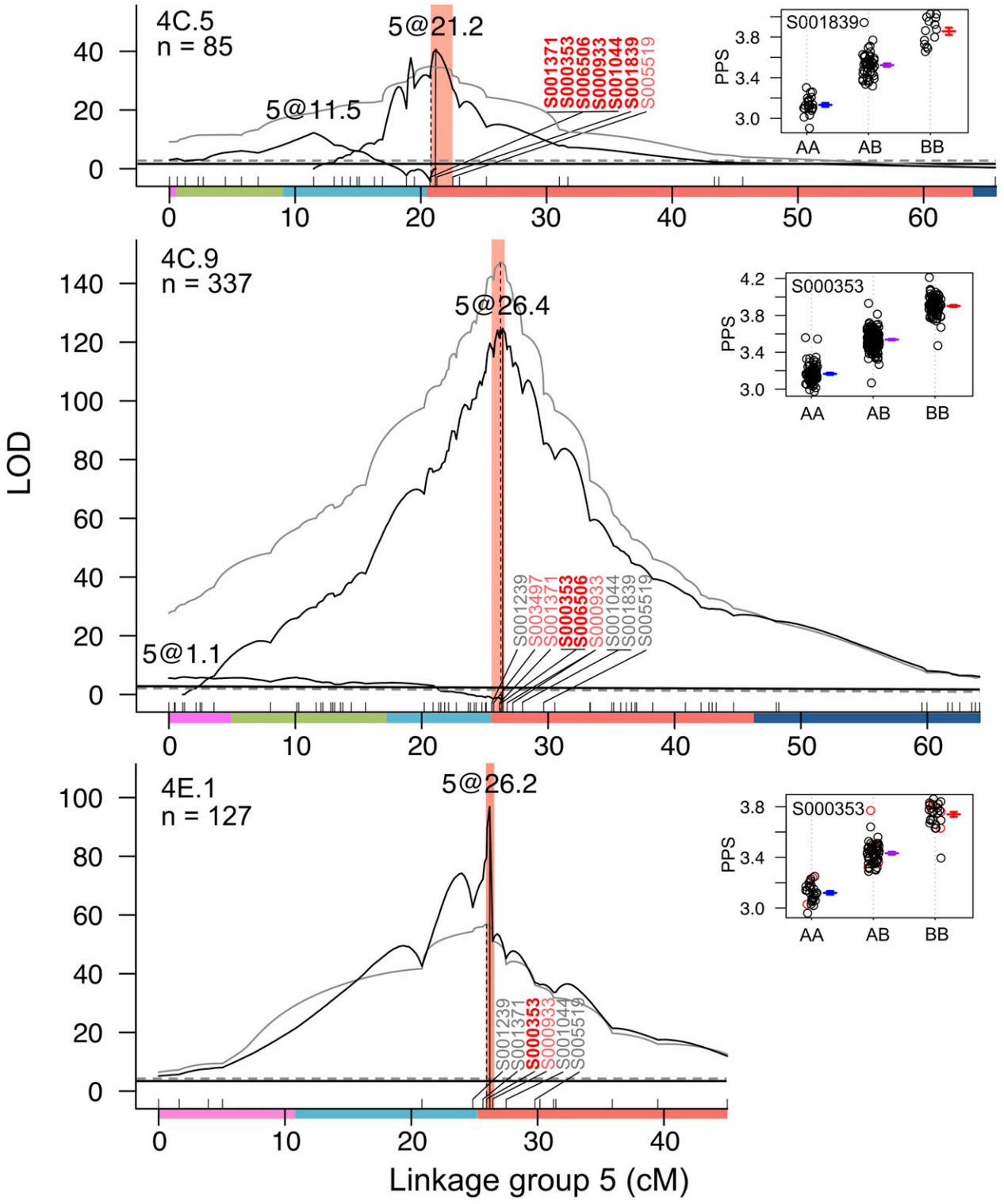

Figure 5 LOD profiles of the final multiple QTL mapping models (MQM, shown in black) and standard interval mapping models (SIM, shown in gray) for 4C and 4E families. The shaded areas indicate 1.5-LOD support confidence intervals, the vertical lines indicate the locations of the peaks and the horizontal lines indicate the significance thresholds for QTL; solid and dashed lines are from MQM and SIM, respectively. Clusters of markers on the linkage group are represented by colored bars on the linkage maps and the same color scheme from Figure 4 is used. Markers within and flanking the 1.5-LOD support confidence intervals are labeled in red, markers with the highest LOD scores from the MQM and SIM are in bold, and markers shared among families but are outside the 1.5-LOD confidence intervals are labeled in gray. The top right panels show phenotypes of F_2 males at the SNP marker with the highest LOD score from the MQM. "A" denotes *L. paranigra* genotype and "B" denotes *L. kohalensis* genotype. Open black circles represent an individual with existing genotype and open red circles represent an individual with imputed genotype. The colored horizontal lines and error bars represent mean ± SE.

Table 2 Standard interval mapping and multiple QTL mapping results on linkage group 5 for five F_2 mapping populations showing the location, LOD score, 1.5-LOD support confidence interval, and three measures of the phenotypic effect for each QTL

				Standard interval mapping			Multiple QTL mapping					
Linkage group	QTL class	Family	Sample size	QTL location (cM)	LOD score (LOD threshold)	1.5-LOD CI (cM)	QTL location (cM)	LOD score (LOD threshold)	1.5-LOD CI (cM)	Additive effect size (pps)	% Species difference explained	% F_2 variance explained
5	Major	4C.9	337	26.20	147.36 (3.23)	25.80−26.75	26.40	124.64 (3.14)	25.80−26.60	0.34 ± 0.01	11.30	87.25
5	Major	4E.1	127	25.96	56.88 (3.15)	24.86–26.40	26.20	96.83 (2.84)	25.96–26.40	0.32 ± 0.01	10.63	81.17
5	Major	4C.5	85	20.70	34.85 (3.30)	19.60−22.54	21.20	40.67 (3.09)	20.80−21.60	0.35 ± 0.02	11.63	86.07
5	Minor	4C.5	85	–	–	–	11.47	12.25 (3.09)	10.60–12.40	0.04 ± 0.02	1.33	3.02
5	Minor	4B.1	83	17.55	4.79 (3.29)	8.00−21.15	17.94	6.70 (3.71)	10.00−25.00	0.07 ± 0.01	2.33	19.85
5	Minor	4C.9	337	–	–	–	1.09	5.97 (3.14)	0.00–11.65	0.04 ± 0.01	1.33	1.01
5	Minor	4B.2	63	–	–	–	15.00	4.15 (3.23)	0.00−26.35	0.11 ± 0.02	3.65	21.62

Note that the proportion of phenotypic difference between the two parental species explained by a single allele is capped at 50%. QTL effects were estimated from the multiple QTL models. 4B.3 does not have any significant QTL, hence, was not shown in the table.

homologs. Moreover, no scaffold within the confidence interval of the major or minor QTL on LG5 contained significant blast hits to conserved domains of *D. melanogaster* candidate genes.

Annotation of the QTL region

The 1.5-LOD confidence interval of the major QTL on LG5 contained five scaffolds in 4C.9 (Table 3). Gene predictions resulted in a total of 66 genes on four of the five scaffolds (Table S5). One scaffold (S006506, 51.9 kb total length) contained no predicted gene, had no mapped transcript or EST, and had no blast hit against any protein databases. Among the 66 predicted genes, 38 had significant blast hits against the nonredundant protein database, 21 of which had gene ontology annotations (Table S5). The annotated genes have functions pertaining metabolism and biosynthesis, cell cycle regulation, cell differentiation, DNA repair, immune response, development, cell signaling, and transmembrane transportation. Among these functional categories, three are potentially relevant to pulse rate variation: ion transmembrane transportation (Table S5, genes #20, #22, #26), synaptic vesicle transportation (gene #23), and muscle development and locomotion (gene #26). All these genes locate on scaffolds with the top two LOD scores.

Evaluation of the predicted genes

We further evaluated the likelihood of annotated genes as the causal gene underlying pulse rate variation using WGS SNPs and indels. The two males used for WGS had song pulse rates of 3.69 and 2.98 pps, respectively. Between these two males, we identified a total of 1664 SNPs, distributed among four of the five scaffolds (S006506 contains no WGS SNP) within the major QTL confidence interval. Among 1664 SNPs, 416 located within 28 (of the 66) predicted genes or in the putative regulatory regions (Table S5). A single-nucleotide deletion and a single-nucleotide insertion were found on scaffolds S000933 and S003497, respectively, both of which intergenic and neither was in the putative regulatory region.

Among 38 predicted genes with annotations, only one gene (Table S5, gene #20) fulfilled all three criteria for potential causal genes. The gene between base locations 312,523 and 384,104 on scaffold S001371 (with the second highest LOD score; Table 3) matches cyclic nucleotide-gated ion channel-like gene (*Cngl*), whose protein product is a distinct member of the cyclic nucleotide-gated ion channel (CNG) family (Figure 6) that contains ion channels gated by cAMP or cGMP. Closely related ion channels gated by cyclic nucleotides are known to be involved in rhythm generation, *e.g.*, leech heartbeat, mammalian locomotion, and respiratory networks (Thoby-Brisson *et al.* 2000; Yamada *et al.* 2005; Herrmann *et al.* 2015; Calabrese *et al.* 2016; Zhu *et al.* 2016). Among the putative *Cngl* homologs we identified among animals (Table S6), the putative *Laupala Cngl* is most similar to the cyclic nucleotide-gated olfactory channel-like gene in termites (Figure 6 and Figure S6), then to AgaP_AGAP003349-PB in mosquitos, and then to *Cngl* in *D. melanogaster* (Figure 6).

The putative *Laupala Cngl* gene contains a nonsynonymous SNP from WGS in the exon coding the cNMP binding domain (Figure 7 and Figure S6). This domain is highly conserved among identified homologs of *Cngl* in insects (Figure 7). This SNP causes the eighth amino acid of the cNMP binding domain to change from glutamic acid in *L. kohalensis* to valine in *L. paranigra* (Figure 7, red box). The putative homologous proteins from termite, mosquito and fruit fly all have glutamic acid at the eighth amino acid. The effect of the amino acid substitution between *L. kohalensis* and *L. paranigra* is predicted to be deleterious (PROVEAN score, −6.70).

Discussion

Dissecting the genetic causes of variation underlying the evolution of reproductive barriers is integral to understanding speciation (Coyne and Orr 2004; Shaw and Mullen 2011). Mating behaviors can be powerful "speciation phenotypes" because their divergence is an effective means to curtail gene flow between incipient species and maintain reproductive boundaries. Moreover, divergence in mating behaviors often

Table 3 Names, map positions, LOD scores, and types of markers within and flanking the 1.5-LOD confidence intervals of the major QTL from multiple QTL mapping models in 4C and 4E families

Family	Marker name	Map position (cM)	LOD score	Marker Type
4C.5	S001371_680831	20.724	30.98	Flanking
4C.5	S000353_224758	20.724	30.98	Flanking
4C.5	S006506_7146	20.724	30.98	Flanking
4C.5	S000933_697595	20.724	30.98	Flanking
4C.5	S001044_55790	20.724	30.98	Flanking
4C.5	5@21.2	21.200	40.67	Peak
4C.5	S001839_297109	21.282	39.86	CI
4C.5	S005519_323363	22.543	27.10	Flanking
4C.9	S003497_180469	25.692	115.01	Flanking
4C.9	S001371_680831	26.147	121.44	CI
4C.9	S000353_224758	26.298	122.88	CI
4C.9	S006506_7146	26.298	122.88	CI
4C.9	5@26.4	26.400	124.65	Peak
4C.9	S000933_1324863	26.754	117.47	Flanking
4E.1	S001239_386983	24.865	62.52	Flanking
4E.1	S001371_680831	25.956	72.17	Flanking
4E.1	S000353_224758	25.956	79.52	Flanking
4E.1	5@26.2	26.200	96.83	Peak
4E.1	S000933_697595	26.440	50.99	Flanking

Markers whose names start with "S" are SNPs and markers whose names start with "5@" are phantom markers from simulations in multiple QTL models. Marker type is categorized as peak (the marker closest to the QTL peak location), CI (markers within the confidence interval), and flanking (markers immediately outside the confidence interval boundaries on both sides). Because the LOD profile in 4E.1 drops much more sharply on the left side of the peak than on the right side, we also included two additional markers outside the left boundary of the confidence interval that have higher LOD score than the immediate right flanking marker.

coincides with the most rapid speciation events (Mendelson and Shaw 2005; Seehausen *et al.* 2008; Momigliano *et al.* 2017; Lamichhaney *et al.* 2018). Yet evidence linking genetic variation in specific mating behaviors to reproductive boundaries in nature is rare.

The Hawaiian cricket *Laupala* has experienced extremely rapid speciation, with the most closely related species showing distinctive song differences, supporting a role for acoustic behavior in the origin of species (Otte 1994; Mendelson and Shaw 2005). The species-specific acoustic signals of *Laupala* males and corresponding female preferences have been found to contribute to reproductive barriers between species (Mendelson and Shaw 2002; Oh *et al.* 2013). Because of the ease of generating laboratory hybrids and isolating specific mating behaviors, *Laupala* crickets are a promising system to identify causal genes underlying behavioral barriers and speciation. Here, we (1) isolate the largest-effect QTL for a pulse rate difference between *L. kohalensis* and *L. paranigra* through selective introgression, (2) fine-map the focal QTL and produce a physical map of the QTL region, (3) test if any *Drosophila* candidate genes for song interpulse interval variation map to the focal QTL region in *Laupala*, and (4) identify strong candidate genes underlying the focal QTL.

Multiple lines of evidence show that we successfully moved the largest-effect QTL from the slow pulsing species (*L. paranigra*) into the genomic background of the fast-pulsing species (*L. kohalensis*), isolating variation at this locus from the polygenic background of pulse rate differences. First, pooling across all families, we found more and a higher density of markers on LG5 than on other linkage groups (Table S2), consistent with selective introgression of QTL4 into NILs. Second, we identified a major QTL on LG5 in 4C and 4E families that explained nearly all F_2 phenotypic variation (>80%; Table 2) and exhibited iconic 1:2:1 Mendalian segregation (Figure 3 and Table S4). Although additional QTL were identified in 4C.5 and 4E.1, each effect was minor (<∼5%) in comparison with the major QTL (Table S3), suggesting successful isolation of a single major locus in replicates 4C and 4E. Third, the major QTL peak in 4C and 4E families are linked to the same SNP marker, with similar estimated phenotypic effect sizes, suggesting that the major QTL in 4C.5, 4C.9, and 4E.1 are the same locus. We also note that the effect size of the major QTL in each family (4C.5, 4C.9, and 4E.1) explains roughly 10% of the species difference, similar to a previous estimate on this linkage group in an F_2 mapping study using AFLP markers (Shaw *et al.* 2007). Thus, the major QTL in 4C.5, 4C.9, and 4E.1 and QTL4 in the previous mapping study (Shaw *et al.* 2007) appear to be the same locus.

Phenotypic data from replicates 4C and 4E exhibited the classical 1:2:1 Mendelian segregation pattern (Figure 5 and Table S4), demonstrating that the major QTL contributing to natural variation of behavioral divergence follows a simple Mendelian segregation rule. Our result here suggests that the genetic underpinning of a complex behavioral trait can be understood element by element with simple Mendelian inheritance rules.

Mapping results testified that we have substantially improved resolution, power, and precision in localizing the major locus. Previous efforts with AFLP markers produced an average marker spacing between 5 and 8 cM (Shaw *et al.* 2007). In comparison, we achieved a median marker spacing of 1.67 cM for all linkage groups and 0.80 cM for LG5 (Table S2), increasing the map resolution by roughly 10-fold on LG5. Moreover, LOD scores and confidence intervals associated with the focal QTL on LG5 were vastly improved, and the distance between the QTL peaks and the closest markers have decreased from 3–5 to 0.1–0.2 cM in the current study (Table 3).

In contrast to results in 4C and 4E families, there were no segregating markers within the confidence interval region of the major QTL in any of the three 4B families (all were homozygous for the *L. kohalensis* genotype). Concomitantly, no major QTL were detected in any 4B family. The likely explanation is that recombination may have occurred between the *L. paranigra* markers used for selective backcrossing and QTL4. Linkage mapping data and genotypes of NIL4B individuals suggested a breakpoint between QTL4 and scaffold S001239, the first scaffold immediately outside the confidence interval of the major QTL (Figure S3A). Replicate 4B families thus serve as a (accidental) negative control, by exhibiting both the lack of a major QTL and any segregating markers within the confidence interval of the major QTL. The

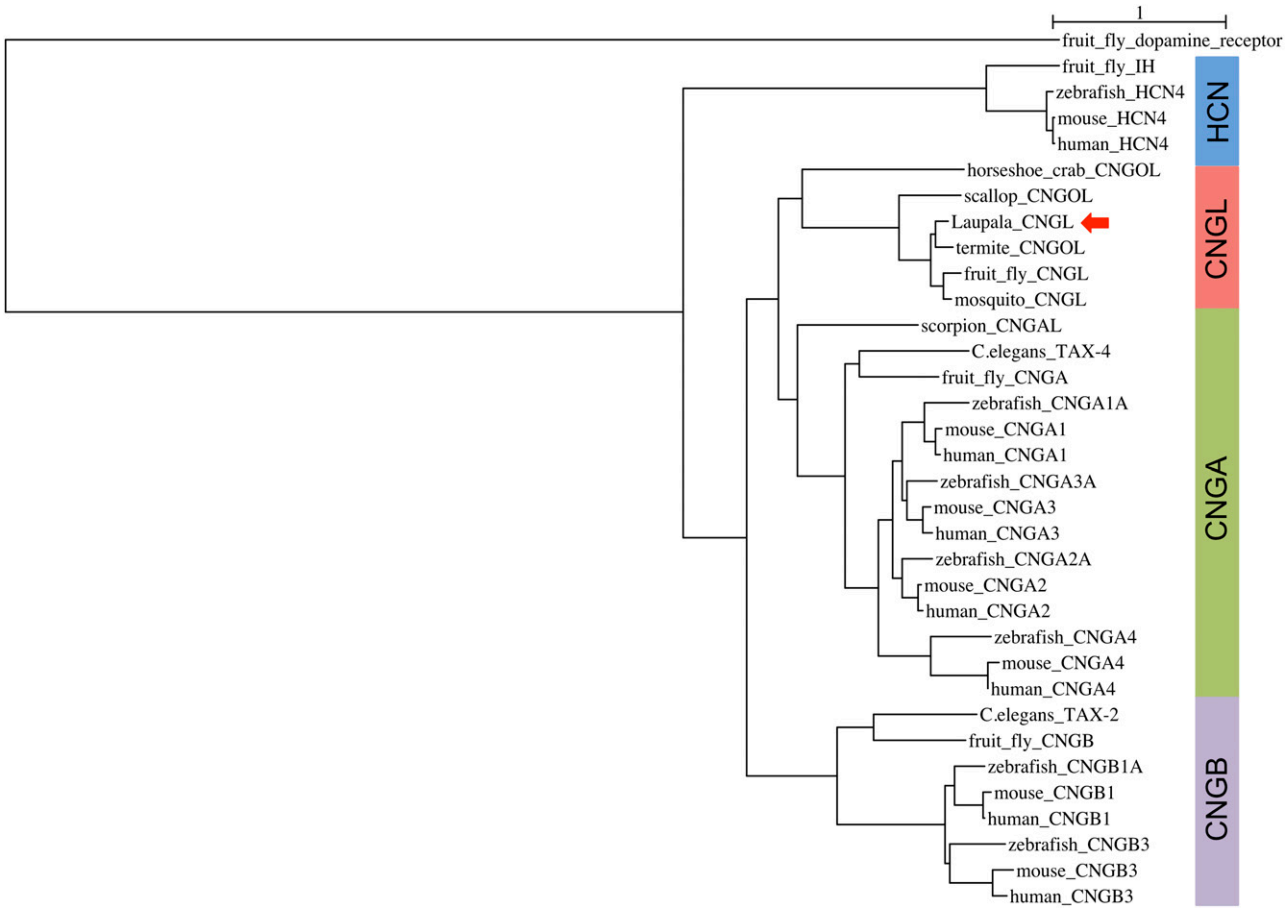

Figure 6 Maximum likelihood tree of amino acid sequences of conserved domains in cyclic nucleotide-gated ion channels subunit A1-4, B1, and B3, and hyperpolarization-activated cyclic nucleotide-gated ion channel subunit 4 from human (*Homo sapiens*), house mouse (*Mus musculus*), zebrafish (*Danio rerio*), and fruit fly (*Drosophila melanogaster*), homologous protein TAX-2 and TAX-4 in roundworm (*Caenorhabditis elegans*), as well as amino acid sequences of conserved domains in putative cyclic nucleotide-gated ion channel-like proteins identified in this study from scallop (*Mizuhopecten yessoensis*), horseshoe crab (*Limulus polyphemus*), scorpion (*Centruroides sculpturatus*), termite (*Zootermopsis nevadensis*), mosquito (*Anopheles gambiae*), fruit fly (*Drosophila melanogaster*), and cricket (*Laupala kohalensis*) using the LG model. Detailed information of the conserved domains represented in this phylogeny can be found in Table S6. The red arrow indicates the location of the putative *Laupala* CNGL protein.

observed phenotypic segregation in F_2 individuals in 4B was thus likely due to effects of QTL on LG1, LG7, and other minor QTL that may not have been detected in the QTL models. The lack of the 1:2:1 phenotypic segregation pattern in 4B families is consistent with this explanation.

We found no evidence to support the hypothesis that the major QTL on LG5 is homologous to a candidate gene for variation in interpulse interval of courtship song in *Drosophila*. None of the putative *Drosophila* homologs we localized in the *Laupala* genome map within the confidence intervals of either the major or the minor QTL on LG5 (Table 1). The two putative homologs for genes CG6746 and *Sps2* to which we could not assign a linkage group were also unlikely to be the causal gene because the gene sequence and the 20-kb flanking regions for each were completely homozygous between the two males we sequenced. The confidence interval region is also unlikely to contain duplicated copies of *Drosophila* song candidate genes because scaffolds within the confidence interval do not contain any significant blast hit against conserved domains in the 11 candidate genes. Therefore, it is unlikely that we missed mapping a *Drosophila* song candidate gene to LG5 because of the distant evolutionary relationship between crickets and flies. These results suggest that the causal gene underlying the major QTL on LG5 is likely a novel gene for song temporal pattern regulation.

To further probe for the causal gene, we conducted gene prediction and functional annotation of five focal scaffolds. The genomic region coverage estimation suggested that the linkage map has saturated the genomic region within the confidence interval of the major QTL (File S1). Thus, the causal gene is likely to be located on one of the focal scaffolds we annotated. We predicted 66 genes on scaffolds within the confidence interval. Twenty-eight of these 66 genes could not be annotated at an E-value cutoff of 1E−4, although just

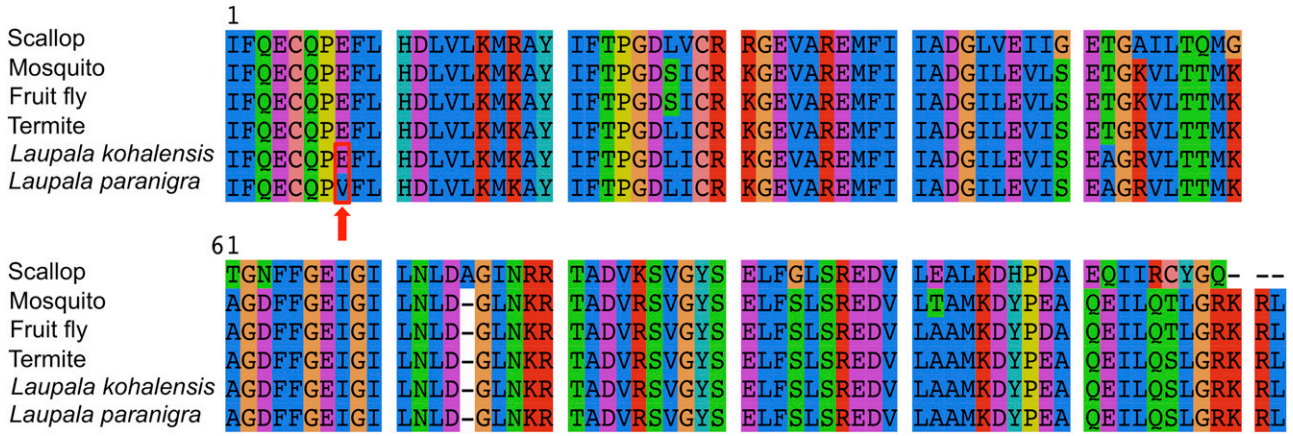

Figure 7 Multiple alignment of amino acid sequences of the cNMP binding domain within putative homologous proteins of the cyclic nucleotide-gated ion channel-like protein in *D. melanogaster*. The red arrow and the red box indicate amino acid substitution caused by the nonsynonymous SNP in the putative *cngl* on scaffold S001371. *Laupala kohalensis* and *L. paranigra* sequences were modified from the termite cNMP domain sequence according to the Exonerate protein2genome alignment (available in Figure S6).

6 out of 66 could not be annotated at a relaxed E-value cutoff of 100 (Table S7). These unannotated genes were, nonetheless, retained as potential causal genes.

Of the 38 annotated genes, only four (Table S4; genes #20, #22, #23, and #26) fall within our anticipated functional categories (ion transportation, neural modulation and development, and locomotion and muscle development) for the causal gene. Each of these genes resides on the scaffold with either the highest or the second highest LOD score. We acknowledge that by relying on SNPs and short indels in our evaluation of potential candidate genes, we may have missed other types of sequence variation in the predicted genes. The putative calcium release-activated calcium channel 1 gene (*Orai1*) and the putative signal peptide peptidase-like protein 3 gene (*Sppl3*), both on the highest LOD scaffold (S000353), have functions related to store-operated calcium entry into nerve and muscle cells, which is necessary for normal muscle function (Venkiteswaran and Hasan 2009; Prakriya and Lewis 2015). Specifically, in *D. melanogaster*, *Orai* has been shown to encode an integral component of the flight CPG and is required for both normal development and functioning of the flight circuit (Venkiteswaran and Hasan 2009; Pathak *et al.* 2015). In vertebrate cell lines, *SPPL3* can enhance the association between *Orai1* and the intracellular Ca^{2+} sensor gene *STIM1* through its protease independent function (Makowski *et al.* 2015). These two genes, however, do not contain any WGS SNPs or indels. The putative synaptic vesicle 2-related protein gene (also known as SVOP), also located on the highest LOD scaffold, contains potential regulatory SNPs, has gene ontology annotation pertaining transmembrane transportation, and is expressed in developing nervous system (Janz *et al.* 1998; Logan *et al.* 2005). However, the specific function of this gene remains elusive and we could not find evidence directly linking this gene to the regulation of rhythmic behavior in either vertebrates or invertebrates.

Finally, a putative homolog of the *Drosophila* cyclic nucleotide-gated ion channel-like gene (*Cngl*) is the most promising candidate, the only one to fulfill all three criteria for a causal role in pulse rate variation. This gene resides on a scaffold immediately adjacent to the peak scaffold in 4C and 4E families. Members of the CNG gene family have been primarily described to function in signal transduction in the visual and olfactory sensory systems (Podda and Grassi 2014). However, the CNGL proteins have distinctive conserved domain architectures from all other CNG proteins (Table S6) and form a distinct clade from other CNG family channels (Figure 6). CNGL proteins are also longer, contain a unique section of C terminus sequence, and are expressed primarily in the brain, thoracic ganglia (wherein lies the CPG for pulse song in *Drosophila*; Clyne and Miesenböck 2008; von Philipsborn *et al.* 2011), and tubular fibers of muscles, as opposed to the photoreceptor, taste, and olfactory neurons, as in other CNG proteins (Miyazu *et al.* 2000). These distinctions suggest that *Cngl* may have functionally diverged from other CNG genes. In addition, members in the closely related hyperpolarization-activated cyclic nucleotide-gated ion channel (HCN) gene family (Figure 6) are known to modulate rhythmic behaviors across broad taxa and contexts, including rhythmic activity of the pyloric and gastric mills in crustaceans (Zhu *et al.* 2016), cardiac pacemaking in both invertebrates and vertebrates (Herrmann *et al.* 2015; Calabrese *et al.* 2016), and rhythmic activity of the respiratory network in mammals (Thoby-Brisson *et al.* 2000). CNG and HCN channels are structurally similar, share two conserved domains (Table S6), are both gated by intracellular cyclic nucleotides, and their genes exhibit a high level of sequence similarity, suggesting shared evolutionary history (Craven and Zagotta 2006). We therefore suggest that *Cngl* may have evolved to perform physiological functions in the CPGs of the wing muscles in insects.

Most intriguingly, our annotation suggests a functional consequence for the nonsynonymous SNP in the cNMP binding domain between *L. kohalensis* and *L. paranigra*. The cNMP binding domain functions to bind intracellular cAMP

or cGMP to modulate channel opening. Thus, the amino acid substitution could potentially cause a shift in sensitivity of this ion channel to the two types of cNMP. We have not yet identified a transcript for the putative *Cngl* gene; future transcriptomes from both *L. kohalensis* and *L. paranigra* will offer crucial expression data to test functional differences of this gene between the two parental species.

Interestingly, although the QTL 4 confidence interval does not contain any *D. melanogaster* candidate genes, like *cacophony* or *slowpoke*, the most intriguing candidate for pulse rate difference in *Laupala* is also an ion channel gene. This result indicates that song temporal pattern evolution in crickets and flies may have taken genetically diverse, but functionally conserved routes.

In summary, identifying targets of selection underlying behavioral divergence is crucial to understanding the early stage of speciation (Nosil and Schluter 2011). Toward this goal, genes identified from divergent lineages offer us more direct inference about genetic routes toward reproductive barriers in natural populations than candidate genes identified from mutational studies. Our annotation of the genomic region and identification of a promising candidate gene offers rare insights into the type of genetic variation, protein coding or regulatory, and its functional categorization underlying the evolution of a behavioral barrier in a rapidly speciating group of nonmodel organisms. Only with data from a diverse range of taxa can we begin to understand the general principles and mechanisms behind the evolution of reproductive barriers and rapid speciation.

Acknowledgments

We thank Ben Weaver, Alex Thomas, Eric Cole, and McKenzie Laws for assistance in cricket rearing; Cornell Institute for Genomic Diversity for advice on molecular procedures; and Aure Bombarely, Thomas Blankers, Linlin Zhang, and Cornell BioHPC laboratory for assistance in bioinformatics. Thomas Blankers and three anonymous reviewers provided comments that improved the manuscript. This project is funded by National Science Foundation grant 1257682 to K.L.S.

Author contributions: M.X. and K.L.S. designed and performed research. M.X. analyzed data. M.X. wrote the paper. K.L.S. edited the paper.

Literature Cited

Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann *et al.*, 2004 UniProt: the universal protein knowledgebase. Nucleic Acids Res. 32: D115–D119. https://doi.org/10.1093/nar/gkh131

Aronesty, E., 2011 *ea-utils: Command-Line Tools for Processing Biological Sequencing Data*. Expression Analysis, Durham, NC.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. Nat. Genet. 25: 25–29. https://doi.org/10.1038/75556

Bailey, N. W., P. Veltsos, Y. F. Tan, A. H. Millar, M. G. Ritchie *et al.*, 2013 Tissue-specific transcriptomics in the field cricket *Teleogryllus oceanicus*. G3 (Bethesda) 3: 225–230. https://doi.org/10.1534/g3.112.004341

Barkan, C. L., E. Zornik, and D. B. Kelley, 2017 Evolution of vocal patterns: tuning hindbrain circuits during species divergence. J. Exp. Biol. 220: 856–867.

Bay, R. A., M. E. Arnegard, G. L. Conte, J. Best, N. L. Bedford *et al.*, 2017 Genetic coupling of female mate choice with polygenic ecological divergence facilitates stickleback speciation. Curr. Biol. 27: 3344–3349.e4. https://doi.org/10.1016/j.cub.2017.09.037

Berdan, E. L., T. Blankers, I. Waurick, C. J. Mazzoni, and F. Mayer, 2016 A genes eye view of ontogeny: de novo assembly and profiling of the *Gryllus rubens* transcriptome. Mol. Ecol. Resour. 16: 1478–1490. https://doi.org/10.1111/1755-0998.12530

Blankers, T., K. P. Oh, A. Bombarely, and K. L. Shaw, 2018a The genomic architecture of a rapid island radiation: recombination rate variation, chromosome structure, and genome assembly of the Hawaiian cricket Laupala. Genetics 209: 1329–1344. https://doi.org/10.1534/genetics.118.300894

Blankers, T., K. P. Oh, and K. L. Shaw, 2018b The genetics of a behavioral speciation phenotype in an island system. Genes (Basel) 9: 346. https://doi.org/10.3390/genes9070346

Broman, K. W., H. Wu, G. A. Churchill, and Ś. Sen, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890. https://doi.org/10.1093/bioinformatics/btg112

Calabrese, R. L., B. J. Norris, and A. Wenning, 2016 The neural control of heartbeat in invertebrates. Curr. Opin. Neurobiol. 41: 68–77. https://doi.org/10.1016/j.conb.2016.08.004

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421. https://doi.org/10.1186/1471-2105-10-421

Cantarel, B. L., I. Korf, S. M. Robb, G. Parra, E. Ross *et al.*, 2008 MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18: 188–196. https://doi.org/10.1101/gr.6743907

Chagnaud, B. P., and A. H. Bass, 2014 Vocal behavior and vocal central pattern generator organization diverge among toadfishes. Brain Behav. Evol. 84: 51–65.

Choi, Y., and A. P. Chan, 2015 PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31: 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. https://doi.org/10.4161/fly.19695

Clyne, J. D., and G. Miesenböck, 2008 Sex-specific control and tuning of the pattern generator for courtship song in *Drosophila*. Cell 133: 354–363. https://doi.org/10.1016/j.cell.2008.01.050

Coyne, J. A., and H. A. Orr, 2004 Speciation. Sinauer Associates, Sunderland, MA.

Craven, K. B., and W. N. Zagotta, 2006 CNG and HCN channels: two peas, one pod. Annu. Rev. Physiol. 68: 375–401. https://doi.org/10.1146/annurev.physiol.68.040104.134728

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danley, P. D., S. P. Mullen, F. Liu, V. Nene, J. Quackenbush *et al.*, 2007 A cricket gene index: a genomic resource for studying neurobiology, speciation, and molecular evolution. BMC Genomics 8: 109. https://doi.org/10.1186/1471-2164-8-109

Ding, Y., A. Berrocal, T. Morita, K. D. Longden, and D. L. Stern, 2016 Natural courtship song variation caused by an intronic

retroelement in an ion channel gene. Nature 536: 329–332. https://doi.org/10.1038/nature19093

Ellison, C. K., and K. L. Shaw, 2013 Additive genetic architecture underlying a rapidly evolving sexual signaling phenotype in the Hawaiian cricket genus *Laupala*. Behav. Genet. 43: 445–454. https://doi.org/10.1007/s10519-013-9601-2

Ellison, C. K., C. Wiley, and K. L. Shaw, 2011 The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*. J. Evol. Biol. 24: 1110–1119. https://doi.org/10.1111/j.1420-9101.2011.02244.x

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379. https://doi.org/10.1371/journal.pone.0019379

Fedotov, S. A., J. V. Bragina, N. G. Besedina, L. V. Danilenkova, E. A. Kamysheva *et al.*, 2014 The effect of neurospecific knockdown of candidate genes for locomotor behavior and sound production in *Drosophila melanogaster*. Fly (Austin) 8: 176–187. https://doi.org/10.4161/19336934.2014.983389

Fedotov, S. A., J. V. Bragina, N. G. Besedina, L. V. Danilenkova, E. A. Kamysheva *et al.*, 2018 Gene CG15630 (fipi) is involved in regulation of the interpulse interval in *Drosophila* courtship song. J. Neurogenet. 32: 15–26. https://doi.org/10.1080/01677063.2017.1405000

Finn, R. D., T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork *et al.*, 2016 InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45: D190–D199. https://doi.org/10.1093/nar/gkw1107

Garrison, E., 2012 Vcflib: A C++ library for parsing and manipulating VCF files. GitHub https://github.com/ekg/vcflib.

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

Gerhardt, H. C., and F. Huber, 2002 *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. University of Chicago Press, Chicago.

Gleason, J. M., 2005 Mutations and natural genetic variation in the courtship song of *Drosophila*. Behav. Genet. 35: 265–277. https://doi.org/10.1007/s10519-005-3219-y

Gleason, J. M., and M. G. Ritchie, 2004 Do quantitative trait loci (QTL) for a courtship song difference between *Drosophila simulans* and *D. sechellia* coincide with candidate genes and intraspecific QTL? Genetics 166: 1303–1311. https://doi.org/10.1534/genetics.166.3.1303

Gleason, J. M., J.-M. Jallon, J.-D. Rouault, and M. G. Ritchie, 2005 Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between *Drosophila simulans* and *D. sechellia*. Genetics 171: 1789–1798. https://doi.org/10.1534/genetics.104.037937

Gleason, J. M., R. A. James, C. Wicker-Thomas, and M. G. Ritchie, 2009 Identification of quantitative trait loci function through analysis of multiple cuticular hydrocarbons differing between *Drosophila simulans* and *Drosophila sechellia* females. Heredity 103: 416–424. https://doi.org/10.1038/hdy.2009.79

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj *et al.*, 2008 High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–3435. https://doi.org/10.1093/nar/gkn176

Greenspan, R. J., and J. F. Ferveur, 2000 Courtship in Drosophila. Annu. Rev. Genet. 34: 205–232. https://doi.org/10.1146/annurev.genet.34.1.205

Groot, A. T., T. Dekker, and D. G. Heckel, 2016 The genetic basis of pheromone evolution in moths. Annu. Rev. Entomol. 61: 99–117. https://doi.org/10.1146/annurev-ento-010715-023638

Guindon, S., and O. Gascuel, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52: 696–704. https://doi.org/10.1080/10635150390235520

Hartbauer, M., and H. Römer, 2016 Rhythm generation and rhythm perception in insects: the evolution of synchronous choruses. Front. Neurosci. 10: 223. https://doi.org/10.3389/fnins.2016.00223

Henry, C. S., M. L. M. Wells, and K. E. Holsinger, 2002 The inheritance of mating songs in two cryptic, sibling lacewing species (Neuroptera: Chrysopidae: Chrysoperla), pp. 269–289 in *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation*. Springer, Dordrecht, The Netherlands.

Herrmann, S., S. Schnorr, and A. Ludwig, 2015 HCN channels—Modulators of cardiac and neuronal excitability. Int. J. Mol. Sci. 16: 1429–1447.

Janz, R., K. Hofmann, and T. C. Südhof, 1998 SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles. J. Neurosci. 18: 9269–9281. https://doi.org/10.1523/JNEUROSCI.18-22-09269.1998

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, 2016 KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44: D457–D462. https://doi.org/10.1093/nar/gkv1070

Katz, P. S., 2016 Evolution of central pattern generators and rhythmic behaviours. Philos. Trans. R. Soc. Lond. B Biol. Sci. 371: 20150057. https://doi.org/10.1098/rstb.2015.0057

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59. https://doi.org/10.1186/1471-2105-5-59

Kronforst, M. R., L. G. Young, D. D. Kapan, C. McNeely, R. J. O'Neill *et al.*, 2006 Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. Proc. Natl. Acad. Sci. USA 103: 6575–6580. https://doi.org/10.1073/pnas.0509685103

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant *et al.*, 2018 Rapid hybrid speciation in Darwin's finches. Science 359: 224–228. https://doi.org/10.1126/science.aao4593

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923

Lassance, J.-M., A. T. Groot, M. A. Liénard, B. Antony, C. Borgwardt *et al.*, 2010 Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature 466: 486–489. https://doi.org/10.1038/nature09058

Lassance, J.-M., M. A. Liénard, B. Antony, S. Qian, T. Fujii *et al.*, 2013 Functional consequences of sequence variation in the pheromone biosynthetic gene pgFAR for *Ostrinia* moths. Proc. Natl. Acad. Sci. USA 110: 3967–3972. https://doi.org/10.1073/pnas.1208706110

Limousin, D., R. Streiff, B. Courtois, V. Dupuy, S. Alem *et al.*, 2012 Genetic architecture of sexual selection: QTL mapping of male song and female receiver traits in an acoustic moth. PLoS One 7: e44554. https://doi.org/10.1371/journal.pone.0044554

Logan, M. A., M. R. Steele, and M. L. Vetter, 2005 Expression of synaptic vesicle two-related protein SVOP in the developing nervous system of *Xenopus laevis*. Dev. Dyn. 234: 802–807. https://doi.org/10.1002/dvdy.20618

Makowski, S. L., Z. Wang, and J. L. Pomerantz, 2015 A protease-independent function for SPPL3 in NFAT activation. Mol. Cell. Biol. 35: 451–467. https://doi.org/10.1128/MCB.01124-14

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz *et al.*, 2015 CDD: NCBI's conserved domain database. Nucleic Acids Res. 43: D222–D226. https://doi.org/10.1093/nar/gku1221

Mendelson, T. C., 2003 Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). Evolution 57: 317–327. https://doi.org/10.1111/j.0014-3820.2003.tb00266.x

Mendelson, T. C., and K. L. Shaw, 2002 Genetic and behavioral components of the cryptic species boundary between Laupala cerasina and L. kohalensis (Orthoptera: Gryllidae), pp. 301–310 in *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation*. Springer, Dordrecht, The Netherlands.

Mendelson, T. C., and K. L. Shaw, 2005 Sexual behaviour: rapid speciation in an arthropod. Nature 433: 375–376. https://doi.org/10.1038/433375a

Merrill, R. M., B. Van Schooten, J. A. Scott, and C. D. Jiggins, 2011 Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. Proc. R. Soc. Lond. B Biol. Sci. 278: 511–518. https://doi.org/10.1098/rspb.2010.1493

Miyazu, M., T. Tanimura, and M. Sokabe, 2000 Molecular cloning and characterization of a putative cyclic nucleotide-gated channel from *Drosophila melanogaster*. Insect Mol. Biol. 9: 283–292. https://doi.org/10.1046/j.1365-2583.2000.00186.x

Momigliano, P., H. Jokinen, A. Fraimout, A.-B. Florin, A. Norkko *et al.*, 2017 Extraordinarily rapid speciation in a marine fish. Proc. Natl. Acad. Sci. USA 114: 6074–6079. https://doi.org/10.1073/pnas.1615109114

Mowles, S. L., M. Jennions, and P. R. Backwell, 2017 Multimodal communication in courting fiddler crabs reveals male performance capacities. R. Soc. Open Sci. 4: 161093. https://doi.org/10.1098/rsos.161093

Niehuis, O., J. Büllesbach, A. K. Judson, T. Schmitt, and J. Gadau, 2011 Genetics of cuticular hydrocarbon differences between males of the parasitoid wasps *Nasonia giraulti* and *Nasonia vitripennis*. Heredity (Edinb) 107: 61–70. https://doi.org/10.1038/hdy.2010.157

Nosil, P., and D. Schluter, 2011 The genes underlying the process of speciation. Trends Ecol. Evol. 26: 160–167. https://doi.org/10.1016/j.tree.2011.01.001

Oh, K. P., G. L. Conte, and K. L. Shaw, 2013 Founder effects and the evolution of asymmetrical sexual isolation in a rapidly-speciating clade. Curr. Zool. 59: 230–238. https://doi.org/10.1093/czoolo/59.2.230

Oh, K. P., and K. L. Shaw, 2013 Multivariate sexual selection in a rapidly evolving speciation phenotype. Proc. R. Soc. Lond. B Biol. Sci. 280: 20130482.

Otte, D., 1994 *The Crickets of Hawaii: Origin, Systematics and Evolution*. The Orthopterists' Society, Philadelphia.

Pathak, T., T. Agrawal, S. Richhariya, S. Sadaf, and G. Hasan, 2015 Store-operated calcium entry through Orai is required for transcriptional maturation of the flight circuit in *Drosophila*. J. Neurosci. 35: 13784–13799. https://doi.org/10.1523/JNEUROSCI.1680-15.2015

Podda, M. V., and C. Grassi, 2014 New perspectives in cyclic nucleotide-mediated functions in the CNS: the emerging role of cyclic nucleotide-gated (CNG) channels. Pflügers Arch. 466: 1241–1257.

Prakriya, M., and R. S. Lewis, 2015 Store-operated calcium channels. Physiol. Rev. 95: 1383–1436. https://doi.org/10.1152/physrev.00057.2003

Rundus, A. S., R. D. Santer, and E. A. Hebets, 2010 Multimodal courtship efficacy of *Schizocosa retrorsa* wolf spiders: implications of an additional signal modality. Behav. Ecol. 21: 701–707. https://doi.org/10.1093/beheco/arq042

Sæther, S. A., G.-P. Sætre, T. Borge, C. Wiley, N. Svedin *et al.*, 2007 Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318: 95–97. https://doi.org/10.1126/science.1141506

Saldamando, C., S. Miyaguchi, H. Tatsuta, H. Kishino, J. Bridle *et al.*, 2005 Inheritance of song and stridulatory peg number divergence between *Chorthippus brunneus* and *C. jacobsi*, two naturally hybridizing grasshopper species (Orthoptera: Acrididae). J. Evol. Biol. 18: 703–712. https://doi.org/10.1111/j.1420-9101.2004.00838.x

Sánchez-Guillén, R. A., A. Córdoba-Aguilar, A. Cordero-Rivera, and M. Wellenreuther, 2014 Rapid evolution of prezygotic barriers in non-territorial damselflies. Biol. J. Linn. Soc. Lond. 113: 485–496. https://doi.org/10.1111/bij.12347

Schöneich, S., and B. Hedwig, 2017 Neurons and networks underlying singing behaviour, pp. 141–153 in *The Cricket as a Model Organism: Development, Regeneration, and Behavior*. Springer, Tokyo.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. Mrosso *et al.*, 2008 Speciation through sensory drive in cichlid fish. Nature 455: 620–626. https://doi.org/10.1038/nature07285

Shaw, K. L., 1996 Polygenic inheritance of a behavioral phenotype: interspecific genetics of song in the Hawaiian cricket genus *Laupala*. Evolution 50: 256–266. https://doi.org/10.1111/j.1558-5646.1996.tb04489.x

Shaw, K. L., 2000 Interspecific genetics of mate recognition: inheritance of female acoustic preference in Hawaiian crickets. Evolution 54: 1303–1312. https://doi.org/10.1111/j.0014-3820.2000.tb00563.x

Shaw, K. L., and D. P. Herlihy, 2000 Acoustic preference functions and song variability in the Hawaiian cricket *Laupala cerasina*. Proc. R. Soc. Lond. B Biol. Sci. 267: 577–584. https://doi.org/10.1098/rspb.2000.1040

Shaw, K. L., and S. P. Mullen, 2011 Genes *vs.* phenotypes in the study of speciation. Genetica 139: 649–661. https://doi.org/10.1007/s10709-011-9562-4

Shaw, K. L., Y. M. Parsons, and S. C. Lesnick, 2007 QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*. Mol. Ecol. 16: 2879–2892. https://doi.org/10.1111/j.1365-294X.2007.03321.x

Slater, G. S., and E. Birney, 2005 Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31. https://doi.org/10.1186/1471-2105-6-31

Smith, M. E., K. K. Weller, B. Kynard, Y. Sato, and A. L. Godinho, 2018 Mating calls of three prochilodontid fish species from Brazil. Environ. Biol. Fishes 101: 327–339. https://doi.org/10.1007/s10641-017-0701-3

Stanke, M., and B. Morgenstern, 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33: W465–W467. https://doi.org/10.1093/nar/gki458

Starnberger, I., D. Preininger, and W. Hödl, 2014 The anuran vocal sac: a tool for multimodal signalling. Anim. Behav. 97: 281–288. https://doi.org/10.1016/j.anbehav.2014.07.027

Thoby-Brisson, M., P. Telgkamp, and J. M. Ramirez, 2000 The role of the hyperpolarization-activated current in modulating rhythmic activity in the isolated respiratory network of mice. J. Neurosci. 20: 2994–3005. https://doi.org/10.1523/JNEUROSCI.20-08-02994.2000

Turner, T. L., P. M. Miller, and V. A. Cochrane, 2013 Combining genome-wide methods to investigate the genetic complexity of courtship song variation in *Drosophila melanogaster*. Mol. Biol. Evol. 30: 2113–2120. https://doi.org/10.1093/molbev/mst111

Ullrich, R., P. Norton, and C. Scharff, 2016 Waltzing Taeniopygia: integration of courtship song and dance in the domesticated Australian zebra finch. Anim. Behav. 112: 285–300. https://doi.org/10.1016/j.anbehav.2015.11.012

Van Ooijen, J. W., 2006 *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*, p. 33. Kyazma BV, Wageningen, The Netherlands.

Venkiteswaran, G., and G. Hasan, 2009 Intracellular Ca^{2+} signaling and store-operated Ca^{2+} entry are required in *Drosophila* neurons for flight. Proc. Natl. Acad. Sci. USA 106: 10326–10331. https://doi.org/10.1073/pnas.0902982106

von Philipsborn, A. C., T. Liu, Y. Y. Jai, C. Masser, S. S. Bidaye *et al.*, 2011 Neuronal control of *Drosophila* courtship song. Neuron 69: 509–522. https://doi.org/10.1016/j.neuron.2011.01.011

Wiley, C., C. K. Ellison, and K. L. Shaw, 2012 Widespread genetic linkage of mating signals and preferences in the Hawaiian

cricket *Laupala*. Proc. Biol. Sci. 279: 1203–1209. https://doi.org/10.1098/rspb.2011.1740

Williams, M. A., A. G. Blouin, and M. A. F. Noor, 2001 Courtship songs of *Drosophila pseudoobscura* and *D. persimilis*. II. Genetics of species differences. Heredity 86: 68–77. https://doi.org/10.1046/j.1365-2540.2001.00811.x

Yamada, R., H. Kuba, T. M. Ishii, and H. Ohmori, 2005 Hyperpolarization-activated cyclic nucleotide-gated cation channels regulate auditory coincidence detection in nucleus laminaris of the chick. J. Neurosci. 25: 8867–8877. https://doi.org/10.1523/JNEUROSCI.2541-05.2005

Zeng, V., B. Ewen-Campen, H. W. Horch, S. Roth, T. Mito *et al.*, 2013 Developmental gene discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for the cricket *Gryllus bimaculatus*. PLoS One 8: e61479. https://doi.org/10.1371/journal.pone.0061479

Zhu, L., A. I. Selverston, and J. Ayers, 2016 Role of I_h in differentiating the dynamics of the gastric and pyloric neurons in the stomatogastric ganglion of the lobster, *Homarus americanus*. J. Neurophysiol. 115: 2434–2445. https://doi.org/10.1152/jn.00737.2015

Communicating editor: C. Peichel