PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Blankers T, Oh KP, Shaw KL. 2019 Parallel genomic architecture underlies repeated sexual signal divergence in Hawaiian *Laupala* crickets. *Proc. R. Soc. B* **286**: 20191479.
http://dx.doi.org/10.1098/rspb.2019.1479

Received: 24 June 2019
Accepted: 13 September 2019

Subject Category:
Evolution

Subject Areas:
genomics, evolution, behaviour

Keywords:
parallel evolution, speciation, quantitative trait loci, sexual communication, *Laupala*

Author for correspondence:
Thomas Blankers
e-mail: thomasblankers@gmail.com

†Present address: Evolutionary and Population Biology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4674254.

THE ROYAL SOCIETY PUBLISHING

Parallel genomic architecture underlies repeated sexual signal divergence in Hawaiian *Laupala* crickets

Thomas Blankers†, Kevin P. Oh and Kerry L. Shaw

Department of Neurobiology and Behavior, Cornell University, Ithaca NY, USA

TB, 0000-0002-1893-8537

When the same phenotype evolves repeatedly, we can explore the predictability of genetic changes underlying phenotypic evolution. Theory suggests that genetic parallelism is less likely when phenotypic changes are governed by many small-effect loci compared to few of major effect, because different combinations of genetic changes can result in the same quantitative outcome. However, some genetic trajectories might be favoured over others, making a shared genetic basis to repeated polygenic evolution more likely. To examine this, we studied the genetics of parallel male mating song evolution in the Hawaiian cricket *Laupala*. We compared quantitative trait loci (QTL) underlying song divergence in three species pairs varying in phenotypic distance. We tested whether replicated song divergence between species involves the same QTL and whether the likelihood of QTL sharing is related to QTL effect size. Contrary to theoretical predictions, we find substantial parallelism in polygenic genetic architectures underlying repeated song divergence. QTL overlapped more frequently than expected based on simulated QTL analyses. Interestingly, QTL effect size did not predict QTL sharing, but did correlate with magnitude of phenotypic divergence. We highlight potential mechanisms driving these constraints on cricket song evolution and discuss a scenario that consolidates empirical quantitative genetic observations with micro-mutational theory.

1. Introduction

Parallel phenotypic evolution offers a unique opportunity to explore the possible routes from genotype to phenotype and associated variations in fitness. Common genetic mechanisms involved in parallel phenotypic evolution offer strong support for evolution by natural selection [1,2], because it is unlikely that stochastic processes such as mutation and drift would repeatedly act on the same genes. Identifying common genetic mechanisms also allows us to assess the predictability of evolutionary changes more broadly. Are the same genomic loci, genes, alleles or nucleotides involved in the repeated evolution of the same trait [3–6]? The answers to these fundamental evolutionary questions have wide-ranging applications, including in medicine [7] and biodiversity management [8].

Current evidence suggests that parallel phenotypic evolution, i.e. repeated divergence of the same phenotype in independent species pairs (see also [9]), can be explained by shared (parallel) genetic mechanisms [3,10]. Theoretical work has shown that parallel genetic evolution is more likely when selection is strong, allelic effect sizes are high, adaptation involves few loci and when selection favours the same phenotypes in diverging populations [11–14]. Accordingly, several traits that experience strong ecological selection pressures and are controlled by one or few genetic loci have been found to evolve repeatedly by changes in the same gene, e.g. wing pattern mimicry in butterflies by *optix* [15] and heavy metal tolerance in populations of *Silene vulgaris* [16]. Conversely, parallel phenotypic evolution of traits controlled by many loci of small effect and mostly evolving through soft rather than hard selective sweeps

[17,18], should be less likely to share genetic bases. Accordingly, population genetic [19–23] and simulation studies [24] have supported the hypothesis of an independent genetic basis for polygenic adaptation.

Nonetheless, some studies examining the genetic architecture of repeatedly evolving quantitative traits have reported reuse of quantitative trait loci (QTL) [9,25,26], genes [27–29] and even nucleotides [30,31]. With further study, these empirical findings may represent a more general pattern for two reasons. First, ancestral polymorphisms at QTL may increase the likelihood of genetic parallelism by presenting opportunity for repeated recruitment of the same loci in parallel divergence events. Second, while many possible genetic trajectories might cause similar phenotypic changes, they probably carry differential fitness consequences, making some trajectories more common than others. This seems especially true when some allelic variants or allele combinations suffer fewer constraints from, e.g. negative pleiotropy [10,32,33].

It is also worth noting that QTL of large effect may, in fact, be owing to the spatially clustered accumulation of many small effect quantitative trait nucleotides (QTN) that effectively inherit as a single QTL of large effect [34,35]. This would result in fewer detected QTL in empirical studies and increases the likelihood of finding shared QTL between independent divergence events, even if different mutations are involved. Theory suggests that a synergy between migration, selection and drift in a background of genomic rearrangements can foster the consolidation of many independently segregating QTN into QTL of large effect [24,36]. In this context, fewer adaptive QTL means fewer maladaptive allelic combinations segregate following recombination during divergence. Much of life's complexity in both plants and animals is quantitative and evolves by many genetic changes of small effect (i.e. polygenic genetic architectures; [37,38]), yet empirical studies of the mechanisms and genetic architecture underlying replicated polygenic divergence are rare in the empirical literature. Particularly underappreciated is the parallel genetic evolution of divergent sexual behaviours, despite their prominent role in repeatedly driving natural variation across the animal kingdom [39,40].

Here, we examine the extent of parallel genetic divergence underlying repeated cases of mating song evolution in the Hawaiian swordtail cricket genus *Laupala* (figure 1*a*). The 38 species of *Laupala* are morphologically cryptic, flightless, forest-dwellers with geographical ranges endemic to single islands within the archipelago. Many species will interbreed under laboratory conditions, but all closely related as well as sympatric species have evolved conspicuous differences in the male mating song that conspecific females prefer, and that correspond to hypothesized species boundaries [42]. Male songs are simple pulse trains produced by single wing strokes at regular intervals [43]. The diversification of male mating song in *Laupala* is quantitative in nature, largely owing to pulse rates ranging from relatively slow to relatively fast.

With the results of two new QTL experiments, and integrating results from two previous studies [42,44], we compare the location and effect size of pulse rate QTL that contribute to the differences between the species of three independently diverged species pairs (six species in total). Two forms of evidence support the assertion that mating song has evolved independently between species of each pair. First, a molecular phylogeny shows that each species pair shares an exclusive common ancestor ([41]; figure 1*a*). Indeed, one species pair (*Laupala cerasina* × *Laupala eukolea*) is part of the *cerasina* species group, which is some 3.5 Myr diverged from the *pacifica* species group to which the other two pairs belong (*Laupala paranigra* × *Laupala kona* and *Laupala pruna* × *Laupala kohalensis*). Second, in the case of the *pacifica* group pairs, a shared most recent common ancestor to all four species was probably endemic to the Big Island of Hawaii, dating the divergence to 0.45 Mya [41]. Despite this close relationship, distinctive song divergence has occurred within each species pair (figure 1*a*), resulting in differences in pulse rates that are either slow–medium (*L. paranigra* × *L. kona*) or medium–fast (*L. pruna* × *L. kohalensis*). Thus, based on current evidence, the comparisons of species differences are made not only among phylogenetically independent pairs of species (independent contrasts) but between species characterized by distinctive outcomes involving the same trait. Within this framework, we test two predictions emerging from the theoretical and empirical work discussed above: (i) when the potential for numerous, small genetic changes underlies phenotypic evolution, as is expected for quantitative traits, repeated divergence events in such traits are unlikely to share a common genetic basis (e.g. [32,45]); and (ii) when loci are shared among divergence events, they will be those of relatively larger phenotypic effect [11,13].

2. Methods

(a) Samples

Comparative QTL mapping was done combining results from a previously published QTL study involving an interspecies cross between *L. cerasina* and *L. eukolea* (hereafter referred to as 'CerEuk') [42], and from two new QTL analyses involving *L. kona* and *L. paranigra* (KonPar), and, *L. pruna* and *L. kohalensis* (PruKoh). A combination of wild-caught and laboratory-reared individuals was used for the parental crosses. For the new analyses presented here, we crossed one male (*L. kona* or *L. pruna*) and one female (*L. paranigra* or *L. kohalensis*) of each species pair, then individually sibling-mated multiple male and female first-generation hybrid offspring to generate 263 (KonPar) and 193 (PruKoh) second-generation hybrid (F_2) males. Details about the other crosses are in [46] and [42].

(b) Genotypes

DNA from parental and all second-generation hybrid offspring (KonPar: $n = 263$; PruKoh: $n = 193$; CerEuk: $n = 230$ and ParKoh $n = 178$) was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) and sequenced following the genotype-by-sequencing (GBS) protocol [47] with *Pst* I, obtaining 100 bp reads on the Illumina HiSeq 2000 (Illumina, San Diego, CA, USA). After quality control and processing of the raw reads, genotypes were called using an approach combining FreeBayes [48] and the Genome Analysis Toolkit [49,50] after which genotypes were filtered based on stringent criteria and only retaining bi-allelic single nucleotide polymorphisms (SNPs) with less than 10% missing data. Further details for genotype calling and filtering have been described previously [42,46].

(c) Phenotypes

Songs were recorded in a temperature controlled, anechoic chamber (for KonPar, temperature range: 19.9–20.7°C; mean ± s.d.: 20.5 ± 0.15°C; for PruKoh, temperature range: 19.3–20.8°C; mean ± s.d.: 20.2 ± 0.34°C) with a Sony Pro cassette recorder and condenser microphone from screen/plastic chambers. The

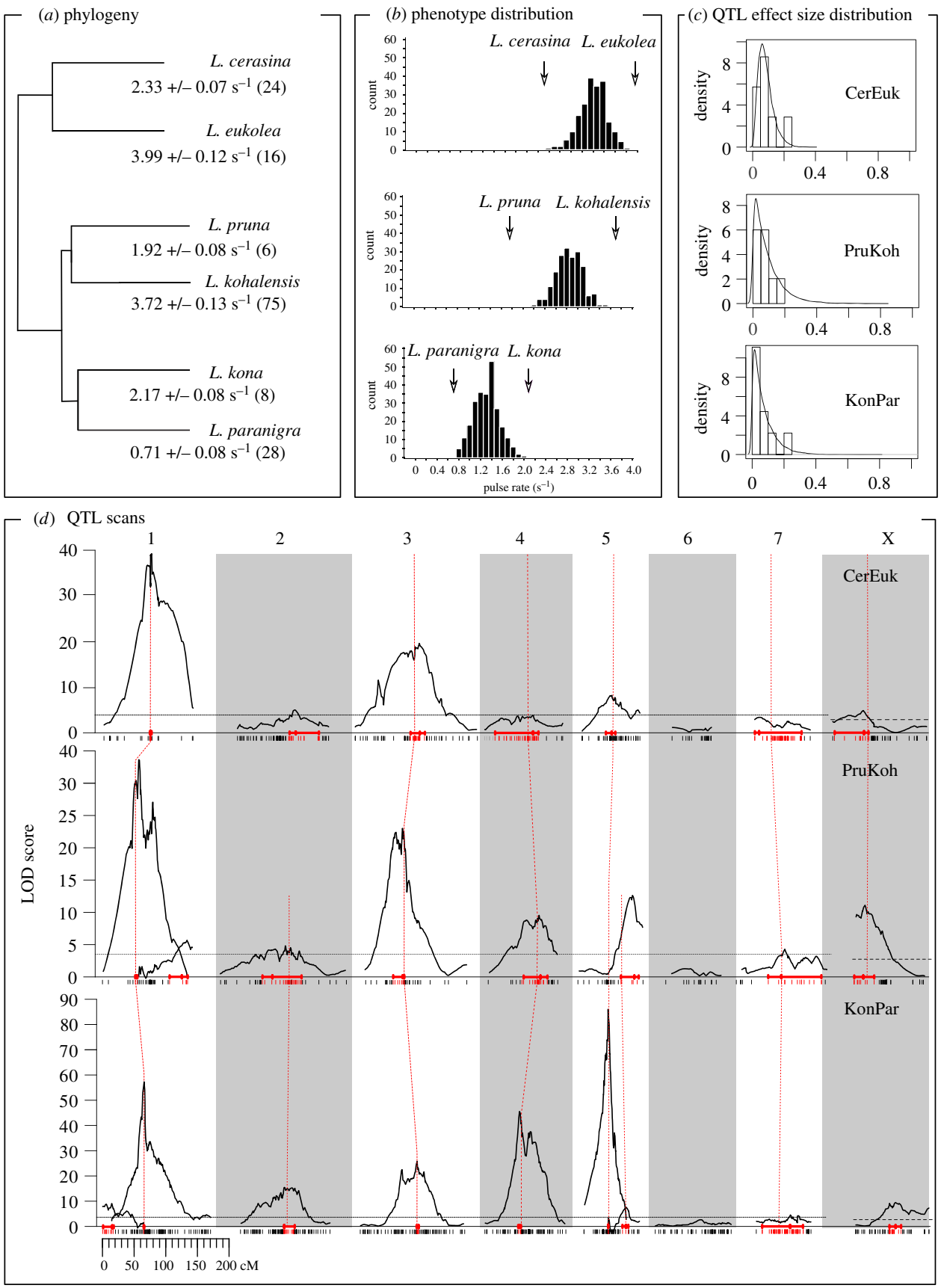

Figure 1. Phenotypic divergence and QTL scan. (*a*) Phylogenetic relationships (based on [41]) for the three species pairs in this study along with species averages and standard deviations for the male pulse rate (sample sizes for song recordings in parentheses; intervening species not shown). (*b*) Phenotypic distribution of the F_2 progeny used for QTL mapping. Data sources: CerEuk (top, from [42]), PruKoh (middle; this study) and KonPar (bottom; this study). Arrows indicate the average phenotypic value of the parental lines of the intercross families. (*c*) Probability density functions (solid lines) of the best fit exponential (KonPar, PruKoh) and gamma distribution (CerEuk) and the density histogram (bars) for the effect size distribution. (*d*) QTL scan. The strength of linkage (LOD score) of the markers with pulse rate variation across the eight linkage groups is shown for the multiple QTL model (black lines). Horizontal lines indicate the significance threshold at 5% FDR. Red bold horizontal lines on the *x*-axis show the location of the QTL peak and the limits of the 95% Bayesian credible interval. Red dashed lines connect markers close to or at the QTL peak that are shared between one or more crosses (see also the electronic supplementary material, tables S2 & S3 and figure S1). (Online version in colour.)

pulse period (the beginning of one pulse to the beginning of the following pulse) was measured from five independent periods (accurate to 10 ms) and transformed to pulse rates (the inverse of pulse period). Recording and quantification of song variation followed established procedures [51].

(d) Linkage maps

The linkage maps and detailed methods regarding their assembly have been described previously [42,46]. In short, marker loci (ancestry informative loci not deviating significantly from 1 : 2 : 1 or 1 : 1 ratios for autosomal or X-linked inheritance, respectively) were grouped into linkage groups (LGs) using JoinMap 4.0 [52] and MapMaker 3.0 [53,54] at log-of-odds (LOD) > 5. Then, for each LG, we used markers that had identical order in JoinMap and MapMaker to create a high confidence scaffolding map. The scaffolding map was then filled-out in MapMaker by iteratively adding markers at log-likelihood thresholds of 3.0, 2.0 and then 1.0, randomly permutating marker orders in seven marker windows to explore alternative orders between each round.

(e) Quantitative trait loci scan

The first goal of this study was to find the number and location of genomic loci underlying pulse rate variation and estimate their contribution to phenotypic divergence. The QTL scans and comparative QTL analyses were done in R [55] and custom codes are available at github.com/thomasblankers/QTLreuse. We detected QTL using single-QTL scans followed by multiple QTL mapping (MQM) in r/QTL as described previously [42]. Thresholds for including additional QTL at false-positive error rate lower than 0.05 were based on 1000 permutations. QTL effect sizes were calculated from the MQM models using drop-one term ANOVA. Based on the total detected QTL and their effect sizes estimated from the MQM models, we estimated the true number of loci (and power to detect QTL) following the equations in [56] using the script QTL_power_detect.r available at https://github.com/thomasblankers/statistics/.

(f) Quantitative trait loci overlap

We tested whether any of the detected QTL in the three independent contrasts (CerEuk, KonPar and PruKoh) overlapped. For each of the detected QTL, we calculated 95% Bayesian credible intervals (BCIs). QTL reuse was inferred when BCIs overlapped between any two (referred to as 'double QTL' from hereon) or all three (triple QTL) species pairs. QTL present in one cross that did not overlap with any other QTL were dubbed 'unique QTL'. To test whether the observed extent of QTL overlap would be expected by chance, we first approximated the probability that we would find six overlapping QTL by chance. The probability was approximated using 100 000 simulations where at each iteration two confidence intervals of s cM (average confidence interval size) were placed on an LG of length l cM (average LG length), n times (where n is the number of detected QTL) using custom code (https://github.com/thomasblankers/qtlreuse). This is similar to an approach using the hypergeometric probability density test in [57], except that the latter method is an exact test statistic, but does not account for the fact that QTL cannot span across LGs.

To more realistically reflect the conditions of our comparative QTL experiment, we also simulated the entire analysis (linkage mapping, QTL mapping, QTL overlap) 1000 times and retained quantiles on the extent of QTL overlap when QTL effects are randomly distributed across the genome. We first obtained a simulated linkage map with eight LGs (one of which was X-linked) of lengths varying between 150 cM and 75 cM and randomly placed markers on that map averaging a marker every 2 cM. We then randomly distribute 16 'true' loci with their effect sizes drawn from an exponential distribution with mean 0.047 pulses s^{-1} (corresponding to the overall mean effect size across all QTL detected in PruKoh, KonPar and CerEuk) on that map. We simulated a QTL cross object using the simulated map and phenotypes for 200 F_2 individuals. Using 'stepwiseqtl', we performed an automated version of MQM to identify and localize the randomly drawn QTL effects and calculated BCIs for each of the identified QTL. Through 1000 iterations of two simulated QTL scans per iteration, we tested if any one or more of the BCIs in the two simulated QTL scans overlapped and counted the number of overlapping BCIs in each iteration. Any BCI overlap here is purely stochastic as the effect size and location of the QTL in the two parallel simulated crosses are randomly drawn from a hypothetical distribution of QTL. However, the linkage map, the number of 'true' loci and the effect size distribution are based on our observations for the three crosses studied here (see Results).

(g) Parallel quantitative trait loci effects

There is considerable variation in the magnitude of phenotypic divergence between parental lines across the three species crosses analysed here, as well as a fourth cross, between *L. kohalensis* and *L. paranigra*, examined in a previous study [44]. The latter cross is not independent of the KonPar and PruKoh crosses but provides an interesting contrast with respect to increasing time since divergence and phenotypic distance (figure 1*a*). The phenotypic difference between *L. kohalensis* and *L. paranigra* is 3.01 pulses per second, whereas parental lines of KonPar, CerEuk and PruKoh differ by 1.46, 1.66 and 1.64 pulses s^{-1}, respectively. We asked whether the magnitude of phenotypic divergence was related to the (detected or estimated total) number of QTL, which is expected if pulse rate evolution occurs by accumulating variants in additional genomic regions, or by increasing the average QTL effect size, which is expected if pulse rate evolution occurs by accumulating variants in genomic regions where other causal variants already reside or by selection favouring variants of increasingly larger phenotypic effect in those regions.

To further examine the genetic basis of repeated pulse rate evolution, we asked whether QTL of larger effect are more likely to be shared, following theoretical research that suggests genetic parallelism is more likely when phenotypic effects are larger [11–13]. We tested whether QTL effect size (calculated above) was predicted by the extent of sharing (factor with three levels: 'unique', 'double' or 'triple'), after accounting for species-specific effects (factor with three levels: 'KonPar', 'PruKoh' and 'CerEuk') using an analysis of covariance.

3. Results

(a) Sequencing and linkage mapping

For the linkage maps, 650 (KonPar) and 325 (PruKoh), SNP markers were grouped into eight LGs at a LOD threshold of 5. Using a combination of JoinMap [52] and MapMaker [53,54] to iteratively order markers within LGs at decreasing stringency of linkage criteria, we obtained two collinear maps (electronic supplementary material, figure S1). The total map lengths were 887.3 cM (KonPar) and 1056.3 cM (PruKoh), corresponding to an average marker spacing of 1.37 cM and 3.25 cM, respectively (electronic supplementary material, figure S1). The average sequencing depth (± s.d.) per linkage map marker across individuals was 38.1× (± 23.8), and 41.8× (± 29.3) for KonPar and PruKoh, respectively. More detailed results on the linkage maps (and collinearity) have been published previously [46].

Table 1. QTL results for PruKoh. (QTL were mapped using 193 F_2 individuals. LG, linkage group; LOD, log-of-odds. A and B alleles denote *L. pruna* and *L. kohalensis* alleles, respectively. QTL effects are shown as the estimated effect in pulses s^{-1} derived from the multiple QTL model and as the per cent variance explained relative to the parental difference (1.64 pulses s^{-1}). All QTL effects are significantly different from zero (*$p < 0.05$; †$p < 0.01$). X-linked marker genotype classes are A0 and B0 because males are hemizygous.)

					genotypes				
LG	position	LOD	nearest scaffold	marker location	AA	AB	BB	effect (pulses s^{-1})	% of parental difference
1	54.0	38.03	S000873	55.4	2.56	2.82	3.02	0.202	12.29†
1	127.5	6.28	S002077	127.5	2.77	2.84	2.92	0.054	3.28*
2	104.0	5.29	S001532	104.5	2.79	2.82	2.93	0.048	2.92*
3	61.5	25.97	S000726	61.5	2.69	2.85	2.96	0.131	8.01†
4	81.0	10.67	S002090	81.0	2.80	2.82	2.92	0.076	4.65†
5	87.0	14.17	S007270	87.8	2.77	2.81	3.00	0.097	5.93†
7	70.0	4.71	S000677	70.1	2.77	2.83	2.98	0.055	3.34*
X	14.0	12.44	S003455	17.5	2.78	—	2.90	0.066	4.02†

Table 2. QTL results for KonPar. (QTL were mapped using 263 F_2 individuals. LG, linkage group; LOD, log-of-odds. A and B alleles denote *L. paranigra* and *L.kona* alleles, respectively. QTL effects are shown as the estimated effect in pulses s^{-1} derived from the multiple QTL model and as the per cent variance explained relative to the parental difference (1.46 pulses s^{-1}). All QTL effects are significantly different from zero (*$p < 0.05$; †$p < 0.01$) except for the LG5 x LG1 interaction (which increases overall QTL model fit, but the slope the relationship between genotype interaction and phenotypic effect size is not significant).)

					genotypes				
LG	position	LOD	nearest scaffold	marker location	AA	AB	BB	effect (pulses s^{-1})	% of parental difference
1	12.7	9.65	S002179	12.7	1.25	1.33	1.34	0.035	2.39*
1	58.1	62.88	S002151	58.1	1.17	1.33	1.44	0.112	7.64†
2	60.0	16.70	S000517	60.1	1.27	1.30	1.39	0.048	3.26†
3	79.1	28.08	S001173	79.1	1.27	1.29	1.41	0.070	4.80†
4	48.0	49.93	S001965	48.8	1.22	1.28	1.47	0.096	6.57†
5	32.5	94.61	S004383	32.5	1.07	1.34	1.55	0.217	14.88†
5	57.6	8.06	S000770	57.6	1.18	1.31	1.49	0.019	1.29†
7	47.0	4.65	S001185	46.9	1.29	1.33	1.31	0.023	1.58*
X	55.0	10.11	S00105	53.4	1.27	—	1.36	0.025	1.72*
LG5 x LG1	(interaction)	6.66	—	—	—	—	—	0.002	0.13
LG1 x LG4	(interaction)	10.04	—	—	—	—	—	−0.025	−1.71*

(b) Is divergence in replicate species pairs associated with similar genetic architectures?

In line with previous results for pulse rate variation in *Laupala*, our findings support a polygenic, additive genetic architecture. Second-generation laboratory-reared hybrids have phenotypic values (i.e. pulse rates) that are intermediate relative to the parental lines (figure 1*b*). Previously, we detected seven QTL for CerEuk ($n = 230$; figure 1*b* top panel; electronic supplementary material, table S1). Our current QTL analyses revealed eight QTL for PruKoh and nine QTL for KonPar and effect size distributions follow either an exponential or gamma distribution (figure 1*c*,*d* and table 2). In addition, two significant interactions were found for KonPar (but not for any of the other crosses), involving QTL on LG1 and LG4, and on LG1 and LG5 (table 1). In all three crosses, multiple loci of small (less than 5% of the parental difference) to moderate (less than 17% of the parental difference) effect were found on the same six autosomal (all but LG6) and the X-linked LGs (tables 1 and 2; electronic supplementary material, table S1).

Using the framework in [56], we estimated the total number of QTL in each cross (assuming an exponential model and estimating the detection threshold, θ, for QTL from the data). The predicted total number of QTL was 12.05 (95% confidence interval: 5.79–21.72; $\theta = 0.0135$) for KonPar and 19.78 (9.05–36.83; $\theta = 0.0418$) for PruKoh. Based on the data in Table 3 in [42], the total number of QTL for CerEuk is estimated at 16.60 (7.14–32.12).

(c) Are the same quantitative trait loci regions involved?

To account for the uncertainty associated with estimating QTL peak locations, we considered QTL to be shared between two ('double' QTL) or all three ('triple' QTL) species pairs if the 95% BCI for QTL on the homologous LGs overlapped. Out of a total of 12 QTL detected in the three species crosses combined, we found four triple QTL and four double QTL (two between PruKoh and KonPar, one each between those and CerEuk) and four QTL unique to a single cross (figure 1; electronic supplementary material, table S4 and figures S2–S4). Thus, between any contrast of two species pairs, five (CerEuk and both other crosses) or six QTL (KonPar and PruKoh) are shared.

The chance of finding six overlapping confidence intervals of average size 21.8 cM for a total of seven detected QTL that are each located on a separate LG of average size 121.5 cM is 0.004. If there are a total of nine QTL, this probability increases to 0.025. To account for the fact that there are more possible QTL than the ones we detected, we calculated the probability that the number of shared QTL (six between KonPar and PruKoh, five between CerEuk and both KonPar and PruKoh) would be observed by chance, given the true number of QTL (between 12 and 20) and the detected QTL (between seven and nine). Ninety-five per cent of the simulated comparative QTL experiments resulted in fewer than four shared QTL and only 1% resulted in six shared QTL. In addition, we found that shared QTL in simulations were typically those with relatively broad confidence intervals, while in the data we observe overlap even between QTL with relatively narrow confidence intervals (e.g. LG1 and LG3; figure 1*d*).

Despite significant overlap in QTL locations, phenotypic effects of shared QTL varied among the crosses. We inspected the relationship between phenotypic effect size, the number of QTL, and the magnitude of phenotypic divergence between the parental species. The phenotypic difference was unrelated to the number of detected QTL ($r^2 = 0.01$, $F_{1,2} = 0.03$, $p = 0.8862$) or the estimated true number of loci ($r^2 = 0.10$, $F_{1,2} = 0.23$, $p = 0.6797$), but was significantly associated with the average effect size (in pulses s^{-1}) across all detected QTL ($r^2 = 0.96$, $F_{1,2} = 68.01$, $p = 0.0144$; figure 2*a*). Furthermore, we found no association between effect size and QTL sharing ($F_{2,19} = 1.41$, $p = 0.2680$; figure 2*b*). *Post hoc* tests further revealed that average effect sizes did not differ significantly among any categories of QTL sharing. This indicates that QTL reuse is not biased towards QTL of larger effect.

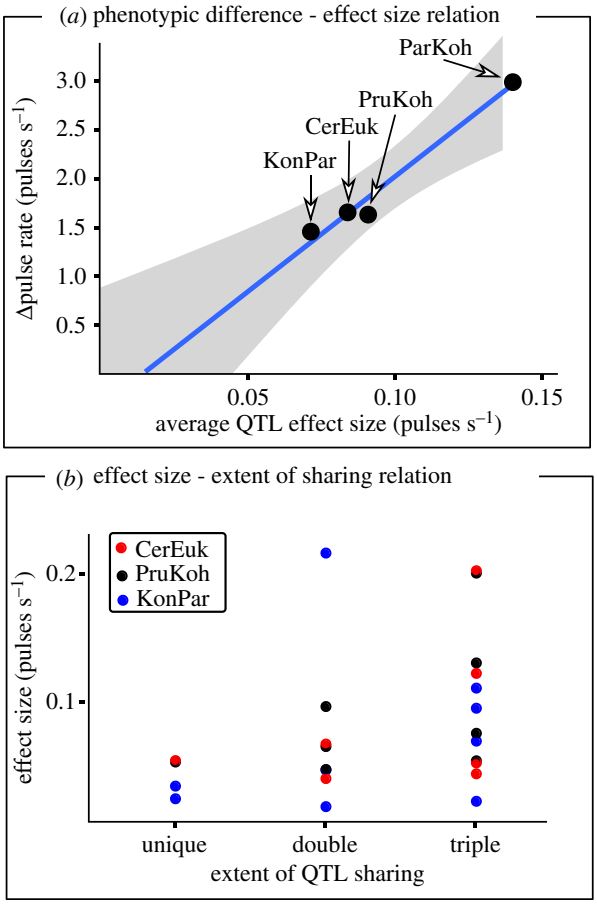

Figure 2. Shared QTL effects. (*a*) Association between phenotypic divergence (the difference between parental pulse rates) and the average effect size of detected QTL across (from lowest to highest) KonPar, CerEuk, PruKoh and ParKoh. The latter data are from Table 3 in [44]. (*b*) The effect size of QTL (dots) across families (colours) is not related to the extent of QTL sharing. However, QTL not shared among any species pairs tend to be of smaller effect than those that are shared across two or all three species pairs (not significant). (Online version in colour.)

4. Discussion

In the evolution of quantitative traits, multiple genetic trajectories can achieve the same net phenotypic effect. However, repeated use of the same genetic elements underlying parallel phenotypic evolution suggests that there are only few feasible genetic pathways towards phenotypic change. Thus far, theoretical and empirical research primarily has focused on the recurrent use of major effect loci [3,10,58,59]. This focus leaves important gaps in our knowledge about genetic parallelism, because phenotypes controlled by many genes of small effect represent most of animal and plant diversity and are central to quantitative genetic theory [37,60,61].

Here, we asked whether the genetic architecture underlying mating song divergence between Hawaiian swordtail cricket species is shared among phylogenetically independent species pairs. The *Laupala* radiation is a powerful study system to address the genetics of parallel phenotypic evolution. The 38 morphologically and ecologically cryptic species, which display conspicuous species boundaries in the wild but can interbreed in the laboratory, have evolved conspicuous differences in mating signals and preferences on short evolutionary timescales. Recent, repeated evolution in this significant barrier to gene flow provides a fascinating natural context to study the evolutionary genetics of sexual selection and speciation.

We find that repeated and independent interspecific divergences of *Laupala* mating songs involve broadly similar genetic architectures (figure 1) and have more QTL in common than would be expected by chance. Therefore, we reject the hypothesis that repeated episodes of divergence in polygenic traits travel along unique evolutionary genetic trajectories, providing important evidence for genetic constraints on polygenic evolution of courtship song. Our result agrees with previous studies in this system in that

pulse rate divergence in *Laupala* is caused by many, additive loci of small to moderate effect [44,62,63], but goes further to show common QTL underlying these changes. Our study further shows that effect size does not predict whether a QTL is likely to be reused in independent, parallel divergence events (figure 2; see also [9]). Consequently, we also reject the hypothesis based on recent theoretical work [13] that loci of relatively larger effect are more likely to diverge in parallel than those of smaller effect.

(a) Genomic constraints on polygenic evolution

Our results indicate genomic constraints to polygenic sexual signal evolution in *Laupala* crickets. First, we found that at least five QTL overlapped between any species pair. Our simulations suggest that the probability of observing five or six QTL in common between crosses is less than 5%, given the size of LGs and QTL confidence intervals, the number of QTL, and their effect size. These results thus provide evidence that QTL from overlapping regions of homologous LGs are recruited repeatedly as species diverged in pulse rate. With current data, while we *can* conclude that these changes occur within the same QTL regions, at least some of which span just a few cM, whether they have occurred in the same genes must await identification of the specific genes involved.

Second, comparing QTL effect sizes between species pairs sharing QTL revealed that when phenotypic divergence increases over time, this is achieved not by recruiting additional QTL, but rather by larger effects at existing QTL. Some caution in the interpretation of our findings is necessary here, as most of the resolution to detect the relationships is the result of a single cross, that of the most phenotypically diverged *L. kohalensis* and *L. paranigra*. We found that average effect size, across all QTL, scales linearly with the magnitude of the pulse rate difference (figure 2*a*). Conversely, we found no relationship between the average QTL effect size and the number of detected QTL that explain species differences in pulse rate. Thus, elaboration of pulse rate differences is associated with larger effects of the same QTL, rather than with additional QTL elsewhere in the genome. Effect size variation among shared QTL suggests that these regions contain distinct, but closely linked, variants or combinations of variants. Because average QTL effect size in each cross and the associated total phenotypic difference increase with increasing divergence times, a reasonable hypothesis is that QTL regions harbour multiple, tightly linked variants of small effect that have accumulated over time.

There are two leading candidate explanations for the observed molecular constraints. First, there is evidence that song rhythm QTL in *Laupala* fall in regions enriched for genes with neuronal and motor rhythmic functions [42], suggesting that pulse rate differentiation might be controlled by clusters of closely linked genes. This would mean that there are spatial limits on where in the genome variation is generated (through de novo mutations and recombination of standing genetic variation), on which selection can act [2,3,58,64]. This will inevitably result in more parallel genetic architectures, because repeated phenotypic evolution would be powered by (the same or different) substitutions in the same cluster. Interestingly, multiple substitutions may accumulate across genes within a cluster, and jointly contribute to the phenotypic effect of the QTL [34,35], especially if genomic rearrangements are favourable under the selection-migration balance [24,36].

Second, pulse rate and pulse rate preference QTL co-localize in the *Laupala* genome [65–67], suggesting that trait-preference coevolution in *Laupala* is driven by tight genomic linkage. Similar to linkage of multiple song genes in tight clusters, linkages between song and preference genes would create differential fitness among the possible routes of male song evolution and those routes that result in the highest mating success of males and females would be favoured. This follows because divergence in sexual phenotypes often requires trait-preference coevolution for the system to remain functional during evolutionary change, i.e. coordinated divergence in signalling and receiver traits [61,68]. It is thus conceivable that linkage or pleiotropy underlying the genes of male and female traits in *Laupala* are contributing to the patterns of QTL sharing that we observed.

Another source of constraint on the molecular pathways towards phenotypic change is when adaptation results mostly from standing genetic variation. Species sharing a common ancestor are also likely to share these variants. By contrast, de novo variants are less likely to be shared [4,18]. Our current data do not allow us to distinguish between the two alternatives. However, given the rapid evolution of *Laupala* species and their pulse rates on the one hand and the limited levels of pulse rate variation segregating within species on the other hand, both are probably important. Although parallel genetic mechanisms are more likely when standing genetic variation is involved, both theoretical and empirical work on polygenic trait evolution more often suggest independent genetic mechanisms [10,32,45]. By contrast, our finding of overlapping QTL for repeated pulse rate evolution contributes to a growing body of literature showing that parallel polygenic architectures may be more likely than previously thought [6,9], either because mutations arise de novo in narrow genomic regions [3,35] or because recombination, migration, and selection result in clustering of causal variants in the genome [24].

(b) Consolidating empirical observations with theory

Based on theory, we hypothesized that QTL sharing between species pairs would be both limited and more likely for QTL of large effect [11,13]. By contrast, we found substantial QTL overlap that was unrelated to the magnitude of the QTL effect (figure 2*b*). This suggests phenotypic effect sizes do not predict the likelihood of genetic parallelism in polygenic evolution, underscoring the assertion that parallel evolution of polygenic traits remains poorly understood.

The QTL on LG5 and other shared QTL that show varying effect sizes depending on the species involved also reveal important qualities of quantitative trait evolution. The QTL underlying differences in more distantly related species that also have more divergent phenotypes, on average, show higher effect sizes than the orthologous QTL in more closely related species. Therefore, as we argued above, it is conceivable that QTL regions accumulate causal variants. There are several mechanisms that ultimately lead to a clustered genetic architecture, including genomic rearrangements [36] and a competitive advantage of linked variants (that have relatively larger joint effects and thus higher fitness) over unlinked small-effect variants [24]. The QTL that have accumulated multiple, tightly linked, small

effect variants will appear as if they are single variants of moderate to large effect. This may explain, in part, why larger effect loci are observed relatively frequently, while quantitative genetics theory predicts phenotypic evolution proceeds with small steps (i.e. adaptive walks) and can resolve the apparent contradiction between micro-mutational theory and ubiquitous observations of major effect loci.

Data accessibility. The datasets supporting the conclusions of this article are archived online: the genome assembly and whole genome sequences are available on NCBI's GenBank, under project number PRJNA392944; GBS alignments for variant calling are available on NCBI's GenBank, under project number PRJNA429815. All variant files, linkage maps, QTL objects and custom code are archived in the Dryad repository https://doi.org/10.5061/dryad.6fd53f8 [69].

Authors' contributions. K.L.S. and T.B. conceived the study, K.P.O. and K.L.S. performed husbandry and the experiments, T.B. analysed the data, T.B. and K.L.S. wrote the manuscript with input from K.P.O.

Competing interests. We declare we have no competing interests.

Funding. This project was funded by the National Science Foundation, grant no. IOS1257682 'The Genomic Architecture underlying Behavioral Isolation and Speciation'.

Acknowledgements. We are thankful to the members of the Shaw laboratory, in particular, Mingzi Xu for her input on linkage mapping and QTL analysis. We also thank the Cornell NBB Department, in particular, Michael Sheehan and Sarah Miller for helpful discussion about this study. We further thank Karl Broman and the r/QTL Google group as well as Biostars and Stackoverflow communities for guidance during bioinformatics and data analysis and Daniel Ortiz-Barrientos and two anonymous reviewers for their helpful comments on a previous version of this manuscript.

References

1. Schluter D, Nagel LM. 1995 Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301. (doi:10.1086/285799)
2. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. (doi:10.1111/j.1558-5646.2011.01289.x)
3. Martin A, Orgogozo V. 2013 The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235–1250. (doi:10.1111/evo.12081)
4. Schluter D, Clifford EA, Nemethy M, McKinnon JS. 2004 Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822. (doi:10.1086/383621)
5. Wood TE, Burke JM, Rieseberg LH. 2005 Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157–170. (doi:10.1007/s10709-003-2738-9)
6. Roda F *et al.* 2013 Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution* **67**, 2515–2529. (doi:10.1111/evo.12136)
7. Nesse RM, Ganten D, Gregory TR, Omenn GS. 2012 Evolutionary molecular medicine. *J. Mol. Med.* **90**, 509–522. (doi:10.1007/s00109-012-0889-9)
8. Templeton AR. 2017 Measuring biodiversity and monitoring ecological and evolutionary processes with genetic and genomic tools. In *Routledge handbook of philosophy of biodiversity* (eds J Garson, A Plutynski, S Sarkar), pp. 251–265, Oxford, UK: Routledge.
9. Conte GL, Arnegard ME, Best J, Chan YF, Jones FC, Kingsley DM, Schluter D, Peichel CL. 2015 Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. *Genetics* **201**, 1189–1200. (doi:10.1534/genetics.115.182550)
10. Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012 The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B* **279**, 5039–5047. (doi:10.1098/rspb.2012.2146)
11. Orr HA. 2005 The probability of parallel evolution. *Evolution* **59**, 216–220. (doi:10.1111/j.0014-3820.2005.tb00907.x)
12. Chevin LM, Martin G, Lenormand T. 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213–3231. (doi:10.1111/j.1558-5646.2010.01058.x)
13. MacPherson A, Nuismer SL. 2017 The probability of parallel genetic evolution from standing genetic variation. *J. Evol. Biol.* **30**, 326–337. (doi:10.1111/jeb.13006)
14. Thompson KA, Osmond MM, Schluter D. 2019 Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* **3**, 129–141. (doi:10.1002/evl3.106)
15. Reed RD *et al.* 2011 *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141. (doi:10.1126/science.1208227)
16. Schat H, Vooijs R, Kuiper E. 1996 Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of *Silene vulgaris*. *Evolution* **50**, 1888–1895. (doi:10.1111/j.1558-5646.1996.tb03576.x)
17. Messer PW, Petrov DA. 2013 Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* **28**, 659–669. (doi:10.1016/j.tree.2013.08.003)
18. Barrett RDH, Schluter D. 2008 Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. (doi:10.1016/j.tree.2007.09.008)
19. Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L. 2011 SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp). *Mol. Ecol.* **20**, 545–559. (doi:10.1111/j.1365-294X.2010.04952.x)
20. Deagle BE, Jones FC, Chan YF, Absher DM, Kingsley DM, Reimchen TE. 2011 Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. B* **279**, 1277–1286. (doi:10.1098/rspb.2011.1552)
21. Kautt AF, Elmer KR, Meyer A. 2012 Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Mol. Ecol.* **21**, 4770–4786. (doi:10.1111/j.1365-294X.2012.05738.x)
22. Westram AM, Galindo J, Alm Rosenblad M, Grahame JW, Panova M, Butlin RK. 2014 Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol. Ecol.* **23**, 4603–4616. (doi:10.1111/mec.12883)
23. Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. 2016 Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**, 287–305. (doi:10.1111/mec.13332)
24. Yeaman S, Whitlock MC. 2011 The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911. (doi:10.1111/j.1558-5646.2011.01269.x)
25. Streisfeld MA, Rausher MD. 2009 Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytol.* **183**, 751–763. (doi:10.1111/j.1469-8137.2009.02929.x)
26. Ferris KG, Rushton T, Greenlee AB, Toll K, Blackman BK, Willis JH. 2015 Leaf shape evolution has a similar genetic architecture in three edaphic specialists within the *Mimulus guttatus* species complex. *Ann. Bot.* **116**, 213–223. (doi:10.1093/aob/mcv080)
27. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012 The molecular diversity of adaptive convergence. *Science* **335**, 457–461. (doi:10.1126/science.1212986)
28. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522. (doi:10.1126/science.1250939)
29. Yeaman S *et al.* 2016 Convergent local adaptation to climate in distantly related conifers. *Science* **353**, 1431–1433. (doi:10.1126/science.aaf7812)

30. Graves JL *et al*. 2017 Genomics of parallel experimental evolution in *Drosophila*. *Mol. Biol. Evol.* **34**, 831–842.
31. Tennessen JA, Akey JM. 2011 Parallel adaptive divergence among geographically diverse human populations. *PLoS Genet.* **7**, e1002127. (doi:10.1371/journal.pgen.1002127)
32. Elmer KR, Meyer A. 2011 Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**, 298–306. (doi:10.1016/j.tree.2011.02.008)
33. Marques DA, Jones FC, Di Palma F, Kingsley DM, Reimchen TE. 2018 Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.* **2**, 1128–1138. (doi:10.1038/s41559-018-0581-8)
34. Huxley J. 1942 *Evolution: The modern synthesis*. London, UK: George Allen and Unwin.
35. Stern DL. 2000 Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091. (doi:10.1111/j.0014-3820.2000.tb00544.x)
36. Yeaman S. 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl Acad. Sci. USA* **110**, E1743–E1751. (doi:10.1073/pnas.1219381110)
37. Mackay TFC. 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339. (doi:10.1146/annurev.genet.35.102401.090633)
38. Orr HA. 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127. (doi:10.1038/nrg1523)
39. Coyne JA, Orr AH. 2004 *Speciation*. Sunderland, MA: Sinauer Associates.
40. Ritchie MG, Phillips SDF. 1998 The genetics of sexual isolation. In *Endless forms: species and speciation* (eds DJ Howard, SH Berlocher), pp. 291–308. New York, NY: Oxford University Press.
41. Mendelson TC, Shaw KL. 2005 Rapid speciation in an arthropod. *Nature* **433**, 375–376. (doi:10.1038/433375a)
42. Blankers T, Oh KP, Shaw KL. 2018 The genetics of a behavioral speciation phenotype in an island system. *Genes* **9**, E346. (doi:10.3390/genes9070346)
43. Otte D. 1994 *The crickets of Hawaii: origin, systematics, and evolution*. Philadelphia, PA: Orthoptera Society/Academy of Natural Sciences of Philadelphia.
44. Shaw KL, Parsons YM, Lesnick SC. 2007 QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*. *Mol. Ecol.* **16**, 2879–2892. (doi:10.1111/j.1365-294X.2007.03321.x)
45. Agrawal AA. 2017 Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *Am. Nat.* **190**, S1–S12. (doi:10.1086/692111)
46. Blankers T, Oh KP, Bombarely A, Shaw KL. 2018 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. (doi:10.1534/genetics.118.300894)
47. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES. 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379. (doi:10.1371/journal.pone.0019379)
48. Garrison E, Marth G. 2012 Haplotype-based variant detection from short-read sequencing. arXiv Prepr arXiv12073907.
49. Van der Auwera GA *et al*. 2013 From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 1–11. (doi:10.1002/0471250953.bi1110s43)
50. DePristo MA *et al*. 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. (doi:10.1038/ng.806)
51. Shaw KL. 1996 Polygenic inheritance of a behavioral phenotype: interspecific genetics of song in the Hawaiian cricket genus *Laupala*. *Evolution* **50**, 256–266. (doi:10.1111/j.1558-5646.1996.tb04489.x)
52. van Ooijen JW. 2006 *Joinmap 4, software for the calculation of genetic linkage maps in experimental populations*. Wageningen, The Netherlands: Kyazma B.V.
53. Lander E, Green P, Abrahamson J, Barlow A, Daly M, Lincoln S, Newburg L. 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181. (doi:10.1016/0888-7543(87)90010-3)
54. Lincoln SE, Daly MJ, Lander ES. 1993 Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: a tutorial and reference manual. A Whitehead Inst Biomed Res Tech Rep. 78–79.
55. R Development Core Team R. 2016 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
56. Otto SP, Jones CD. 2000 Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* **156**, 2093–2107.
57. Moyle LC, Nakazato T. 2008 Comparative genetics of hybrid incompatibility: sterility in two *Solanum* species crosses. *Genetics* **179**, 1437–1453. (doi:10.1534/genetics.107.083618)
58. Stern DL. 2013 The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764. (doi:10.1038/nrg3483)
59. Stern DL, Orgogozo V. 2008 The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177. (doi:10.1111/j.1558-5646.2008.00450.x)
60. Robertson A. 1967 The nature of quantitative genetic variation. In *Heritage from Mendel* (ed. A Brink), pp. 265–280. Madison, WI: University of Wisconsin Press.
61. Lande R. 1981 Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725. (doi:10.1073/pnas.78.6.3721)
62. Ellison CK, Wiley C, Shaw KL. 2011 The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*. *J. Evol. Biol.* **24**, 1110–1119. (doi:10.1111/j.1420-9101.2011.02244.x)
63. Oh KP, Fergus DJ, Grace JL, Shaw KL. 2012 Interspecific genetics of speciation phenotypes: song and preference coevolution in Hawaiian crickets. *J. Evol. Biol.* **25**, 1500–1512. (doi:10.1111/j.1420-9101.2012.02531.x)
64. Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010 Convergence in pigmentation at multiple levels: mutations, genes and function. *Phil. Trans. R. Soc. B* **365**, 2439–2450. (doi:10.1098/rstb.2010.0104)
65. Shaw KL, Lesnick SC. 2009 Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl Acad. Sci. USA* **106**, 9737–9742. (doi:10.1073/pnas.0900229106)
66. Wiley C, Ellison CCK, Shaw KLK. 2012 Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proc. R. Soc. B* **279**, 1203–1209. (doi:10.1098/rspb.2011.1740)
67. Xu M, Shaw KL. In press. Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proc. R. Soc. B* (doi:10.1098/rspb.2019.1607.R1)
68. Kirkpatrick M. 1982 Sexual selection and the evolution of female mate choice. *Evolution* **36**, 1–12. (doi:10.1111/j.1558-5646.1982.tb05003.x)
69. Blankers T, Oh KP, Shaw KL. 2019 Data from: Parallel genomic architecture underlies repeated sexual signal divergence in Hawaiian *Laupala* Crickets. Dryad Digital Repository. (https://doi.org/10.5061/dryad.6fd53f8)