





Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus* grasshoppers

Zachary J. Nolen^{1,2}  | Burcin Yildirim¹ | Iker Irisarri^{3,4,5}  | Shanlin Liu^{6,7} |
Clara Groot Crego¹  | Daniel Buchvaldt Amby⁶ | Frieder Mayer⁸ |
M. Thomas P. Gilbert⁹ | Ricardo J. Pereira^{1,6} 

¹Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, München, Germany

²Department of Biology, Lund University, Lund, Sweden

³Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

⁴Department of Organismal Biology (Systematic Biology), Uppsala University, Uppsala, Sweden

⁵Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Goettingen, Campus Institute Data Science, Goettingen, Germany

⁶Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

⁷College of Plant Protection, China Agricultural University, Beijing, China

⁸Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

⁹GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

Correspondence

Ricardo J. Pereira, Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany.
Email: ricardojn.pereira@gmail.com

Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 658706

Abstract

Theoretical and empirical studies have shown that species radiations are facilitated when a trait under divergent natural selection is also involved in sexual selection. It is yet unclear how quick and effective radiations are where assortative mating is unrelated to the ecological environment and primarily results from sexual selection. We address this question using sympatric grasshopper species of the genus *Chorthippus*, which have evolved strong behavioural isolation while lacking noticeable ecomorphological divergence. Mitochondrial genomes suggest that the radiation is relatively recent, dating to the mid-Pleistocene, which leads to extensive incomplete lineage sorting throughout the mitochondrial and nuclear genomes. Nuclear data shows that hybrids are absent in sympatric localities but that all species have experienced gene flow, confirming that reproductive isolation is strong but remains incomplete. Demographic modelling is most consistent with a long period of geographic isolation, followed by secondary contact and extensive introgression. Such initial periods of geographic isolation might facilitate the association between male signaling and female preference, permitting the coexistence of sympatric species that are genetically, morphologically, and ecologically similar, but otherwise behave mostly as good biological species.

KEYWORDS

diversification, gene flow, hybridization, sexual selection, speciation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Speciation begins when populations accumulate genetic differences that result in reproductive barriers (Mayr, 1963), such as incompatible mating signals that prevent interbreeding, incompatible ecological interactions that render hybrids unfit in either extrinsic parental habitat, or incompatible gene interactions that cause intrinsic hybrid dysfunction (Presgraves, 2010). Although multiple barriers are expected to become coupled during the speciation process (Butlin & Smadja, 2018), behavioural barriers usually evolve early during divergence, leading taxonomists to rank their importance “far ahead of all others” (Mayr, 1963). Although sexual selection is generally accepted as an important driving force of such early behavioural incompatibilities, it is thought to result in assortative mating only when acting in concert with divergent natural selection (Servedio & Boughman, 2017). Thus, it is still unclear how efficient species radiations are when assortative mating is unrelated to the ecological environment, and which demographic scenarios may favour the evolution of assortative mating in the first place.

Emblematic examples of species radiations driven by divergent natural selection, termed “ecological radiations” (Nosil, 2012; Schluter, 2000), have been a major focus of evolutionary studies because they shed light onto the selective forces and the demographic scenarios that result in stable behavioural incompatibilities in the face of homogenizing gene flow. For example, in radiations of sympatric species of Darwin finches (Grant & Grant, 2011), cichlid fishes (Carleton & Kocher, 2001), and butterflies (Jiggins, 2017), consistent phenotype-environment correlations suggest the repeated action of natural selection driven by different ecological niches (seed sizes, water clarity, or antipredator strategies) shaping trait values (beak shape, opsins, and colour pattern). Because such ecologically relevant traits also function as mating cues, behavioural isolation may emerge as a byproduct of divergent natural selection (Langerhans & Riesch, 2013; Servedio et al., 2011; Smadja & Butlin, 2011).

Nevertheless, species radiations may also occur without the clear evidence of divergent natural selection. In these so-called “nonecological” radiations (Mayer et al., 2010; Rundell & Price, 2009), sympatric species do not occupy different ecological niches (Gillespie et al., 2001), and did not diverge in ecologically relevant traits (Gillespie et al., 2020; Schluter, 2000). Instead, sexual selection is assumed to be the primary driving force in the divergence between populations (Panhuis et al., 2001; Ritchie, 2007). Although the sexual selective pressures (e.g., female preference) are similar in sister taxa at early stages of divergence, the order that sexual traits (e.g., male signalling) arise and become fixed is stochastic, resulting in independent coevolution of signal and preference within each lineage, and causing behavioural isolation upon secondary contact (i.e., “mutation-order divergence” by sexual selection; Mendelson et al., 2014). For example, in radiations of crows (Vijay et al., 2016), frogs (Yang et al., 2019), and field crickets (Gray & Cade, 2000), closely related species differ most in sexually selected traits (plumage, colour pattern, or songs) that do not obviously affect signaling efficiency in their current environments, but that are tightly associated with

female preference. However, the operational difficulty of classifying a radiation as nonecological comes from the impossibility of measuring all ecological attributes of species, which means that ecological differences can escape detection when they are difficult to study, when their relative contribution to reproductive isolation is small, or when they are restricted to certain life stages (Rundell & Price, 2009). Although these systems have received less attention from evolutionary biologists relative to well-studied ecological radiations, nonecological radiations can lead to similar burst of species that maintain behavioural barriers, despite lacking geographic and ecological isolation. The highest rate of speciation recorded in arthropods occurs in *Laupala* Hawaiian crickets (Mendelson & Shaw, 2005), where species remain reproductively isolated due to pulse rate of male courtship songs associated to female preference for specific pulse rates (Otte, 1994), suggesting that sexual selection on these traits may be the driver of this radiation. Although these systems are common in nature, the demographic conditions associated to their radiation remain poorly understood. In particular, it is still unclear how rapid these radiations can be, if behavioural isolation alone can prevent hybridization in sympatry, and whether behavioural isolation evolved in the face of continuous gene flow or during historical periods of geographic isolation.

Chorthippus grasshoppers are ideally suited to answer such questions and to provide general insights into whether sexual selection can maintain species boundaries without a primary role of environmental factors (Mayer et al., 2010). This genus comprises around 230 nominal species (Cigliano et al., 2020), including several species groups of morphologically similar species such as the *biguttulus* group (Harz, 1975; Figure 1) that conform to several expectations for species mainly separated by sexual selection (Panhuis et al., 2001). First, these species strongly differ in male songs and in co-evolving female acoustic preferences (Perdeck, 1958; von Helversen & von Helversen, 1994), as well as in cuticular hydrocarbons (Finck et al., 2016), all genetically inherited. Second, this coupling of male song and female preference is sufficient to result in prezygotic isolation where species are sympatric (von Helversen & von Helversen, 1994), so that hybrids are rarely found in nature (Baur, 2006; Jacobs, 1963; Perdeck, 1958). Third, there is little genetic differentiation between species (Hawllitschek et al., 2017; Mason et al., 1995), suggesting rapid divergence. Finally, there is no reduction in F1 hybrid viability or fertility caused by Dobzhansky-Muller incompatibilities (Perdeck, 1958; Saldamando, Tatsuta, & Butlin, 2005; but see Finck & Ronacher, 2017). Yet, hybrid fitness is reduced in F1 males due to an intermediate song that is not appealing to either parental female nor to their F1 sisters, leaving them “behaviourally sterile” (Gottsberger & Mayer, 2007, 2019). It is still unclear whether character displacement is implicated in the evolution of assortative mating.

Although *Chorthippus* spp. offer an unusual opportunity to understand how sexual selection can lead to the radiation of sympatric species, evolutionary genetic studies have been rare in this genus. Phylogeographic studies have been restricted to the former species *Chorthippus parallelus* (reviewed in Butlin, 1998; Hewitt, 1993),

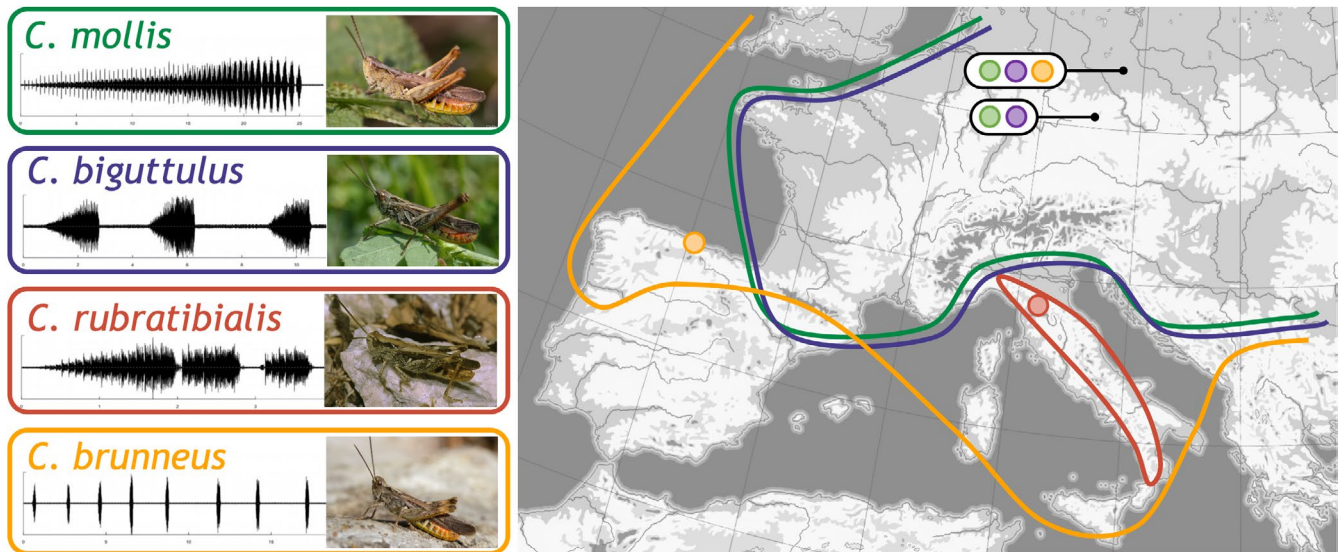


FIGURE 1 Grasshopper species of the genus *Chorthippus* have similar distribution, morphology and habitat, but are highly divergent in male mating song (shown by oscillograms) and in associated female preference. Map depicts approximate species distributions (Ragge et al., 1990) and sampling sites [Colour figure can be viewed at wileyonlinelibrary.com]

where two subspecies, *parallelus* and *erythropus*, have been shown to diverge in allopatry during the Pleistocenic glaciation, and established recent secondary contact, forming a narrow hybrid zone (Hewitt, 2000). This species has recently been reassigned to the genus *Pseudochorthippus* (Cigliano et al., 2020; Defaut, 2012), based on divergence in a fragment of the CO1 mitochondrial gene (Vedenina & Mugue, 2011); this hypothesis remains to be tested using nuclear markers. Attempts to study the evolution of *Chorthippus* species of the *biguttulus* group using mitochondrial markers have failed to provide a phylogenetic hypothesis due to unusually high levels of intraspecific diversity (Hawliitschek et al., 2017; Mason et al., 1995). Another difficulty for genetic studies is their large genome (10 Gb; Gosalvez et al., 1980) which remains challenging to assemble using modern genome sequencing methods. Such difficulties can be overcome by high throughput sequencing of transcriptomes, which offer a reliable, reference-free approach for phylogenetic and population genomic methods (Gayral et al., 2013).

In this study, we used four species of the *biguttulus* group (*C. mollis*, *C. biguttulus*, *C. brunneus*, and *C. rubratibialis*), and its close relative *Pseudochorthippus parallelus*, to understand the demographic history of a species radiation where sexual selection is hypothesized to be the main driver of speciation. First, by assembling full mitochondria and 12,735 independent nuclear genes from transcriptome data, we test whether this radiation is recent, and whether nominal species form monophyletic groups (i.e., showing that the evolution of behavioural isolation coincides with the evolution of strong genome wide differentiation). Second, using single-nucleotide polymorphisms (SNPs), we test whether allopatric and sympatric species experienced frequent hybridization and genetic introgression. Finally, using demographic modeling, we test whether speciation has occurred in the face of continuous gene flow or with periods of historical isolation.

2 | MATERIALS AND METHODS

2.1 | Specimen and genomic sampling

To understand the radiation of species that mainly differ in mating behaviour, we focused on four species from the *biguttulus* group (Ragge et al., 1990): the three highly sympatric species *C. mollis* (herein, *Cmol*), *C. biguttulus* (*Cbig*), and *C. brunneus* (*Cbru*); and *C. rubratibialis* (*Crub*), which is only sympatric with *Cbru* (Figure 1; Table S1 for sampling details and accession numbers). Published RNAseq data from *Cmol* ($n = 20$ individuals), *Cbig* ($n = 18$), and *Cbru* ($n = 20$) were obtained from Berdan et al. (2015, NCBI BioProject PRJNA284873). These populations were sampled in localities in which each species pair is sympatric or allopatric in relation to each other, allowing to test whether gene flow follows current geographic overlap between species. Additionally, new RNAseq data was collected for *Crub* ($n = 16$), and for the two parapatric subspecies of *Pseudochorthippus parallelus* (*P. p. parallelus* (*Pppar*), $n = 5$, and *P. p. erythropus* (*Ppery*), $n = 5$), until recently named *Chorthippus parallelus*, to serve as a phylogenetic outgroup of the radiation and to avoid potential mapping biases arising from using any of the ingroup species.

Tissue samples were preserved in RNAlater (Qiagen). To sample the largest variety of transcripts possible, we extracted total RNA from the whole body after removing the head and digestive track of each individual to avoid overrepresentation of eye pigments and gut contaminants. We homogenized the samples using ceramic beads (1.4/2.8 mm, Precellys) and the standard Tri-Reagent protocol (Sigma). Resuspended RNA pellets were purified with RNAeasy Mini columns (Qiagen), and final sample integrity and quantity was assessed with an Agilent 2100 BioAnalyser. Enrichment for mRNA, library construction and paired end Illumina HiSeq2500 sequencing was performed by the BGI Group at a high coverage for all the *Pppar* and *Ppery* reference samples and at a lower coverage for *Crub*.

2.2 | Transcriptome assembly and mapping

We constructed a complete reference transcriptome for each of the 5 *Ppar* and 5 *Ppery* samples separately. We used TRINITY DE NOVO ASSEMBLER v2.2.0 (Grabherr et al., 2011), which used Trimmomatic for trimming and filtering the raw reads. We used the default parameters and a kmer of 31. We retained contigs with kmer coverage >2, supported by >3 reads, and with >400 bp in length, retaining the longest isoform per gene. We evaluated individual assemblies by computing the contig N50 based on transcript contigs representing the top 90% of expressed transcripts. We assessed the completeness of each individual assembly, first by estimating the length of our assembled genes relative to the complete transcriptome of *Locusta migratoria* (Wang et al., 2014) using BLAST, and second by estimating the percentage of complete, fragmented and missing genes relative to the single copy genes expected for metazoans using BUSCO (Waterhouse et al., 2018).

We then constructed a single transcriptome reference for *Pseudochorthippus* representing the maximum variability observed in this species. First, we assigned genes to the mitochondrial genome through BLASTN searches of the assembled contigs against the mitogenome of *Chorthippus chinensis* (GenBank accession number NC_011095.1; Liu & Huang, 2008) and keeping the longest contigs. Second, we compared the 10 individual assemblies using the Markov Cluster algorithm (MCL), using the pipeline implemented in ORTHOMCL (Li et al., 2003) to distinguish “orthologous genes” identified across all individuals from “singleton genes” identified only in some individuals, keeping the longest sequence for each gene. Third, for the “orthologous genes”, we identified genes potentially involved in duplications, based on the distribution of coverage of whole genome reads from *Ppar* (NCBI BioProject PRJNA280771) mapped onto our reference transcriptome (median coverage = 1.5 \times). We labelled all genes with coverage up to quantile 80% (<5 \times) as “single-copy genes”, and the remaining as “multicopy genes” (>5 \times). Fourth, for the “single-copy genes”, we identified untranslated regions (UTRs) and open reading frames (ORFs) using TRANSDCODER v3.0.1 (Haas & Papanicolaou, 2012). To expedite bioinformatic analyses, all reference genes were separated by 500 Ns and organized in four artificial chromosomes: (a) single-copy genes (with and without ORFs); (b) multicopy genes; (c) singleton genes; and (d) mitochondrial genes. We also produced auxiliary files containing the coordinates of each protein coding gene and the coordinates of synonymous positions (i.e., untranslated regions (UTRs) and third codon positions).

For assessing the genetic variability in the 84 individuals across the transcriptome, we processed raw sequencing reads using the PALEOMIX BAM pipeline v1.2.9 (Schubert et al., 2014) with default settings. This process first removed adapters and trimmed low quality bases with AdapterRemoval (Lindgreen, 2012). Overlapping paired reads were collapsed into a single consensus read. Trimmed and collapsed reads were then mapped to the reference transcriptome using BWA-MEM, discarding unmapped reads and reads below a mapping quality threshold of 30 (Li & Durbin, 2009). In the final step, local realignment was performed around indels using GATK INDELREALIGNER

(McKenna et al., 2010). We recorded the number filtered nucleotides, read length and coverage to assess potential mapping biases between *Chorthippus* spp.

Because the mitochondrial genome evolves faster than the nuclear genome, we also assembled one reference mitochondrial genome for one individual of each of the four *Chorthippus* species. We used 10 million pairs of filtered RNAseq reads in MIRA v4.0.2 (Chevreux et al., 2004) and MITOBIM v1.8 (Hahn et al., 2013) with the default parameters and using the mitochondrial genome from the transcriptome of *Pseudochorthippus parallelus* as a reference. The initial assemblies were annotated with MITOS revision 656 (Bernt et al., 2013) to ensure correct gene orders. Protein-coding genes were individually aligned with orthologous sequences from other Acrididae available at NCBI (Table S2) in order to identify possible indels within the ORFs, which were considered misassemblies and therefore corrected in the initial assemblies. Corrected mitochondrial genomes were then used as references to re-map the reads from all the individuals from the corresponding species with BWA-MEM as explained above. The corrections introduced to the mitochondrial references were confirmed by a higher proportion of aligned reads compared to the initial assemblies. The final mitochondrial genomes were annotated with MITOS revision 942 (Bernt et al., 2013).

2.3 | Phylogenetic relationships

As phylogenetic inference requires probabilistic estimates of gene sequences, we called sequences for each of the 84 individuals separately for reads mapping to the mitochondrial and nuclear references using ANGSD v0.921 (Korneliussen et al., 2014). We called the most frequent base at each position with a minimum coverage threshold of 10 \times .

2.3.1 | Mitochondrial DNA

In order to infer the timing of diversification of *Chorthippus* species and their relation to *Pseudochorthippus parallelus*, we estimated a mitochondrial time tree for our 84 individuals plus all the available mitogenomes of Acrididae (Table S2). Single gene alignments were constructed using MUSCLE (Edgar, 2004) as implemented in SEAVIEW v.4.7 (Gouy et al., 2010). Protein-coding genes were translated with the invertebrate mitochondrial genetic code, aligned at the amino acid level, and back-translated to nucleotides. The 13 protein-coding genes and two ribosomal RNA genes were concatenated into a single data set, and alignment columns with >80% missing data were removed with BMGE v.1.12 (Crisuolo & Gribaldo, 2010). We inferred a maximum likelihood tree using IQTREE v.1.6.5 (Nguyen et al., 2015), with BIC-selected best-fit substitution models and partitions (Table S3) as identified using ModelFinder (Kalyaanamoorthy et al., 2017). Branch support was assessed with 1,000 bootstrap replicates. The inferred tree topology was then used to estimate the timing of diversification of nodes representing the most recent

common ancestors of (a) *Chorthippus* and *Pseudochorthippus*; (b) the subspecies *Pppar* and *Ppery*; and (c) of all focal *Chorthippus* species. We used the uncorrelated relaxed clock model implemented in BEAST v.1.10.4 (Suchard et al., 2018), with a normally distributed prior on the substitution rate with mean of 0.0115, and standard deviation of 0.33 substitutions per million year (Brower, 1994), which we believe is appropriate given the remarkably conserved mitochondrial rates of insects (e.g., 0.0115 in butterflies and 0.0133 in beetles; Brower, 1994; Papadopoulou et al., 2010, respectively). The tree was fixed to the previously inferred maximum likelihood topology, and we used a Yule speciation tree prior. To facilitate convergence, tree topology, clock models and substitution parameters (GTR + Γ) were linked across genes. Two independent MCMC chains were run for 100 million generations, sampling every 10,000. Convergence was checked a posteriori with TRACER v.1.7.1 (Rambaut et al., 2018), and all values had ESS values > 300. The final divergence times were obtained from one of the chains using TreeAnnotator, after excluding the initial 10 million as burn-in.

2.3.2 | Nuclear DNA

To test if nominal species based on behavioural isolation form monophyletic clades, and to infer their evolutionary relationships, we inferred nuclear species trees under the multispecies coalescent model, which accommodates incomplete lineage sorting expected in fast radiations. Maximum likelihood trees were inferred for all single copy genes (UTRs were excluded due to the prevalence of assembly errors in these regions) that included at least three individuals per sampling locality, and called genotypes for >50% of the sequence length. We used BIC-selected models in IQTREE (Nguyen et al., 2015), 1,000 replicates of ultrafast bootstrapping (UFBoot), and SH-like approximate likelihood ratio tests (SH-aLRT) (Guindon et al., 2010; Hoang et al., 2018). To account for the uncertainty in gene tree estimation, we collapsed branches with <50% bootstrap support, and used these trees for downstream analyses. In addition, we performed likelihood mapping in IQTREE to assess the information content of each gene tree (Strimmer & Haeseler, 1997). The consensus gene trees were fed into ASTRAL II (Mirarab & Warnow, 2015). This method estimates the species tree that maximizes the agreement among independent gene trees. We calculated the quartet support for each internal node, which reflects the fraction of quartets supporting the main, the first and second alternative topologies, and thus the amount of gene tree discordance at that node that is caused by incomplete lineage sorting or by gene flow. From this analysis, we output the main species tree topology, posterior probabilities for this main topology, branch lengths in coalescent units (T/N_e , Degnan & Rosenberg, 2009), and quartet supports. We first built a species tree considering each individual as a terminal branch, then a second by treating individuals of the same sampling locality as populations (as in Table S1).

We also used this set of nuclear genes to estimate relative divergence times between all pairs of populations and test if divergence

patterns are consistent with those observed in the mitochondrial phylogeny. For this, we calculated Nei's measure of absolute divergence (D_{xy}) that, contrary to the branch length of the species tree, is independent of effective population size, assuming no selection. We calculated D_{xy} per gene from the individual gene alignments used to estimate nuclear gene trees using the R package PopGenome (Pfeifer et al., 2014; R Core Team, 2017). We plotted the distribution of average per gene D_{xy} for each pairwise comparison and calculated the median value for each population comparison.

2.4 | Population genetic analyses

Recent studies have shown that genotype likelihoods are more reliable than standard genotype calls to estimate demographic parameters (da Fonseca et al., 2016), especially when coverage is variable, such as is expected for RNAseq data. We thus estimated genotype likelihoods using ANGSD (Korneliussen et al., 2014). We used both trimmed and collapsed reads, excluding reads with the following criteria: mapping quality lower than 15 after adjustment, a base read quality lower than 20, and containing multiple hits. Additionally, we limited the analyses to genomic positions (sites) that were present in at least 80% of the individuals, with a depth greater than twice the total number of individuals, in single copy genes, and excluding the first and second codon positions that are generally involved in nonsynonymous substitutions. All the specific commands were deposited in a public repository (https://github.com/zjnolen/chorthippus_radiation). Together, this filtering approach restricts transcriptome data to biallelic sites that in principle are not directly affected by selection, conforming to the assumptions of population genetic methods (Gayral et al., 2013).

2.4.1 | Population structure

To assess population structure, we performed a principal component analysis (PCA) with the program ngsPopgen (Fumagalli et al., 2014). We excluded nonvariable sites by limiting the analysis to sites with a SNP p -value $< 1.0 \times 10^{-2}$. We have determined this value to be sufficient to remove the majority of nonvariable sites, while not distorting the site frequency spectrum based on all sites (Figure S1).

To test if reproductive isolation between sympatric species of *Chorthippus* is strong, or if they show frequent hybridization, we used NGSadmix (Skotte et al., 2013). This method estimates admixture proportions in each individual to each of the K presumed ancestral clusters, maximizing Hardy-Weinberg and linkage equilibria within clusters. We performed this analysis restricting to sites with a SNP p -value $< 1.0 \times 10^{-2}$, considering K from two to the number of sampling localities (nine), with 50 replicates, and selecting the replicate with the highest likelihood for each K . Under a mutation-drift equilibrium, we expect each species or sampling locality to form a cluster, with genetic admixture only in sympatric localities if hybridization is common (Lawson et al., 2018).

2.4.2 | Genetic variability

To estimate the relative effective population size of each sampling locality and whether it changes over time, we have calculated Watterson's Theta (Watterson, 1975) and Tajima's D (Tajima, 1989), respectively. For each of these summary statistics, we first calculated the site frequency spectrum (SFS) for each population, and then calculated Theta and Tajima's D per site in ANGSD (Korneliusson et al., 2014), combined values across sites within each gene, and plotted their distributions for the genes sampled across all populations.

To determine how differentiated populations of *Chorthippus* and *Pseudochorthippus* are from each other, we estimated pairwise F_{ST} (Reynolds et al., 1983) between all population pairs. We calculated a two-dimensional SFS for each population pair, calculated F_{ST} per site in ANGSD, per gene F_{ST} , and plotted their distributions for genes showing at least 10 SNPs across all population pairs.

2.4.3 | Genetic introgression

To test for the presence of introgression between *Chorthippus* species, and whether it is more common in sympatric relative to allopatric populations, we used Patterson's *D*-statistics (Durand et al., 2011). This test weights the number of biallelic sites that have a different topology from a species tree, with an outgroup and three ingroup taxa (i.e., ABBA and BABA sites). We used the AbbaBaba2 method implemented in ANGSD, which extends this analysis to comparisons among populations (Soraggi et al., 2018). We used all individuals of *Pppar* and *Ppery* as the outgroup, and all possible combinations of *Chorthippus* populations that conformed to our estimated species tree, restricting to sites with a SNP *p*-value $< 1.0 \times 10^{-6}$. We inferred significance using a *p*-value calculated from 10 windows of ~150,000 relevant sites. A similar proportion of ABBA and BABA sites are expected under the null hypothesis of incomplete lineage sorting driving discordance, while significantly different proportions must be explained by gene flow. When different populations of the same species are considered, finding gene flow in sympatric but not in allopatric comparisons suggests that introgression has occurred only recently, while finding gene flow across both comparisons suggests also ancestral introgression.

We additionally estimated the most likely demographic scenario of speciation for all pairs of *Chorthippus* species and of *Pseudochorthippus* subspecies, in order to understand if behavioural isolation evolved under continuous gene flow or during periods of geographic isolation. First, we considered four relatively simple demographic models that assumed constant population size after divergence: model 1: "No Migration" (divergence without gene flow, three parameters: time since divergence, N_e -pop1, N_e -pop2), model 2: "Continuous Migration" (divergence with continuous gene flow, fourth parameter: symmetric gene flow), model 3: "Ancestral Migration" (divergence with ancestral gene flow fifth parameter: time until cessation of gene flow), and model 4: "Secondary Contact" (divergence with gene flow after secondary contact, fifth parameter:

time without gene flow between split and secondary contact). We constructed the two-dimensional SFS using ANGSD for each of the 16,969 single-copy genes, and summed them into a single complete SFS per population pair. We folded each complete SFS and fit it to all four demographic models using the diffusion approximation methods implemented in *δaδi* v2.0.5 (Gutenkunst et al., 2009). To ensure a proper fitting of each model to the observed SFS, we used the *δaδi*-pipeline developed by Portik et al. (2017), which performs several rounds of replicated optimizations using the highest likelihood parameters of the previous rounds as the initial parameters in subsequent rounds. For these four models each pipeline run involved four rounds with [10, 10, 5, 5] replicates, [3, 2, 1, 1]-fold perturbed starting parameters, and [5, 20, 125, 200] max iterations per round respectively. We performed three replicates of each pipeline run to ensure model fittings converged on a single set of parameter values. This procedure was repeated for all pairs of populations. Second, as Momigliano et al. (2020) have demonstrated that the exclusion of population size changes in demographic models can bias model selection towards a model of secondary contact (model 4 above) relative to a model with continuous gene flow (model 2 above), we estimated the likelihood of these two models including two extra parameters for population growth ("Continuous Migration with Growth" and "Secondary Contact with Growth", respectively; models adapted from Rougeux et al., 2017). We altered the pipeline run for these more complex models in order to reach convergence between three independent runs, first changing the optimizer from the default *optimizer_log_fmin* to *optimizer_log* which had better performance under these models. We then performed each run with five rounds of [20, 20, 30, 20, 5] replicates, [3, 2, 1, 1, 0.5]-fold perturbed starting parameters, and [1, 5, 20, 50, None] max iterations per round respectively, where "None" results in the chain continuing until convergence.

In order to account for the physical linkage between SNPs within the same gene, we built 100 nonparametric bootstrap SFS per taxa pair. To do this, we randomly resampled the individual gene SFS with replacement, summing them into a bootstrap SFS containing the same number of gene SFS as the complete data SFS (Gutenkunst, 2015). We used these bootstraps to (a) select the best demographic model using the adjusted likelihood ratio test; and (b) estimate parameter uncertainties using the Godambe Information Matrix (Coffman et al., 2016; Godambe, 1960). If speciation occurred in sympatry, we expect higher support for models containing continuous gene flow or with reductions of gene flow over time. Conversely, if speciation was facilitated by an initial period of geographic isolation, we expect higher support for the models without gene flow or with gene flow after secondary contact.

Because there is no mutation rate (μ) available for transcriptomes in grasshoppers, the demographic parameters were estimated in reference to the constant μ , as the number of mutations per site per generation. Nevertheless, these estimates reflect comparable, relative measures of isolation and gene flow, as μ is probably the same for such closely related taxa. As the relationship between μ is inverse for N_e and m and cancels out when calculating $2Nm$, we calculated absolute measures of $2Nm$ using the estimated parameters.

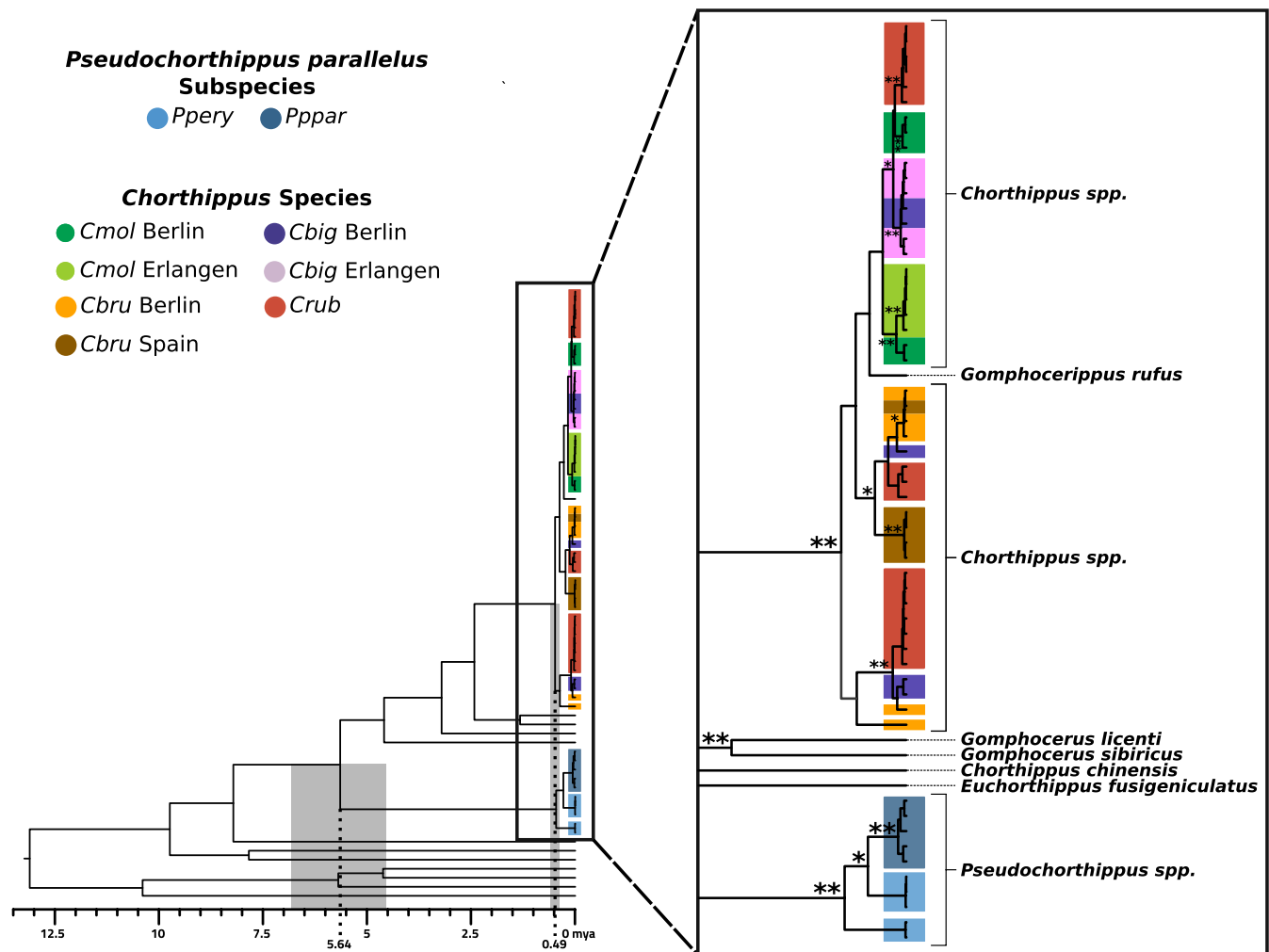


FIGURE 2 Time-calibrated tree based on full mitochondrial data shows that the radiation of *Chorthippus* species is relatively recent. The most recent common ancestor (MRCA) of *Chorthippus* haplotypes occurred in the mid-Pleistocene, at the same time of the MRCA of *Pseudochorthippus* haplotypes, leading to incomplete lineage sorting among species and subspecies (coloured squares). The MRCA of haplotypes of both genera predates the Pleistocene. Dotted lines and grey boxes mark the estimated times and confidence intervals. Topology support is illustrated at the nodes, **100% BP, *95% BP. All nodes outside the outset tree have 100% BP [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS

3.1 | Transcriptome assembly and mapping

The individual assemblies had similar levels of quality across the 10 high-coverage *Pppar* and *Ppery* individuals. Contig sizes were relatively large across individuals (Figure S2), with contigs showing an N50 around 2,200 bp for genes explaining 90% of the expression. Around 4,300 genes contain the complete gene length found in *Locusta migratoria*, and ~11,000 are at least 10% complete (Figure S3). Based on expectations for metazoans, between 78%–82% of the BUSCO genes were complete, 6.6%–9.7% were fragmented, and 10%–12% were missing (Table S4). After merging and filtering individual assemblies, our reference transcriptome contained: 12,735 single-copy genes with identified ORF; 4,235 single-copy genes without ORFs; 4,263 multicopy genes; 18,623 singleton genes; and the 13 mitochondrial genes plus two rRNAs. Within the single-copy genes with identified ORFs, there are 20,604,833 synonymous positions.

The number of filtered nucleotides and read length was similar across *Chorthippus* spp. and higher in *Pseudochorthippus*, reflecting a shorter library insert size and larger sequencing effort for the reference transcriptome. Accordingly, coverage for the single copy nuclear genes averaged 70x for the outgroup, and 25x for the ingroup species (Figure S4). The reference mitogenomes lacked any stop codons and showed an individual coverage above 8,000x due to their high expression.

3.2 | Phylogenetic relationships

3.2.1 | Mitochondrial DNA

The mitogenomes produced a well-supported topology, due to a relatively large number of substitutions (Figure 2). *Pseudochorthippus parallelus* is highly divergent from *Chorthippus* spp. (bootstrap proportion, BP = 100%), with the genera *Euchorthippus* and

Gomphocerus being more closely related to *Chorthippus* than to *Pseudochorthippus* (BP = 100%). Interestingly, *Chorthippus chinensis* nested between these two genera, suggesting that they do not belong to the same genus as the European *Chorthippus* (BP = 100%). The mitogenomes of the focal *Chorthippus* spp. formed a monophyletic group (BP = 100%), yet none of the four species nor the seven sampling localities formed monophyletic clades. Interestingly, the mitogenome of the *Gomphocerippus rufus* individual nested within our focal species of *Chorthippus*.

The time-calibrated trees show that the diversification of Acrididae is relatively old, dating to some 13.6 Ma (95% HPD 16.2–11.1), and that *Pseudochorthippus* and *Chorthippus* diverged around 5.9 Ma (7.1–4.7), before the Pleistocene. We estimate that *Pseudochorthippus* subspecies and *Chorthippus* species diversified around the same time, between 476 to 506 ka respectively (see Table S5 for HPD).

3.2.2 | Nuclear DNA

As expected for transcriptome data, the completeness of nuclear genes was generally consistent across species, reflecting general expression values in grasshoppers (Figure S5). After filtering, we recovered 3,963 genes that were used to estimate nuclear gene trees.

The individual consensus trees often showed polytomies, reflecting the uncertainty on gene tree estimation (Figure S6). The likelihood mapping analysis showed that most gene trees have

enough phylogenetic signal to resolve 60%–100% of the quartets (Figure S7). No gene tree contained more than 20% partially resolved quartets.

The individual coalescent tree (Figure S8) showed that individuals sampled in each locality generally form well-supported monophyletic clades (posterior probability, PP > 0.9). The exceptions are *Cbig* from Berlin, which includes the monophyletic clade from Erlangen, and *Cbru* from Berlin, which includes a monophyletic clade from Spain. All species and genera are reciprocally monophyletic (PP = 1.0).

In agreement, the population coalescent tree shows that populations of every species form monophyletic clades (PP = 1.0; Figure 3a). *Cmol* has the oldest split within the *Chorthippus* radiation (PP = 1.0). The relationship between *Crub*, *Cbru* and *Cbig* is less clear, but most likely *Crub* is a sister species of *Cbru* (PP = 0.67). Coalescent branch lengths are shorter between *Chorthippus* species compared to those between the *Pseudochorthippus* populations, and the coalescent times within populations strongly vary, being longest in *Cmol* from Erlangen.

The quartet support for the common ancestor of *Chorthippus* and *Pseudochorthippus* shows a complete prevalence of the main topology, reflecting no gene tree discordance at those nodes. In contrast, for all nodes within the radiation of *Chorthippus* the fraction of the main topology is only slightly larger (34%–41%) than the fractions of alternative topologies (29%–33%). The ancestor of *Crub* and *Cbru* has the highest amount of gene tree discordance, consistent with the lower support for this relationship.

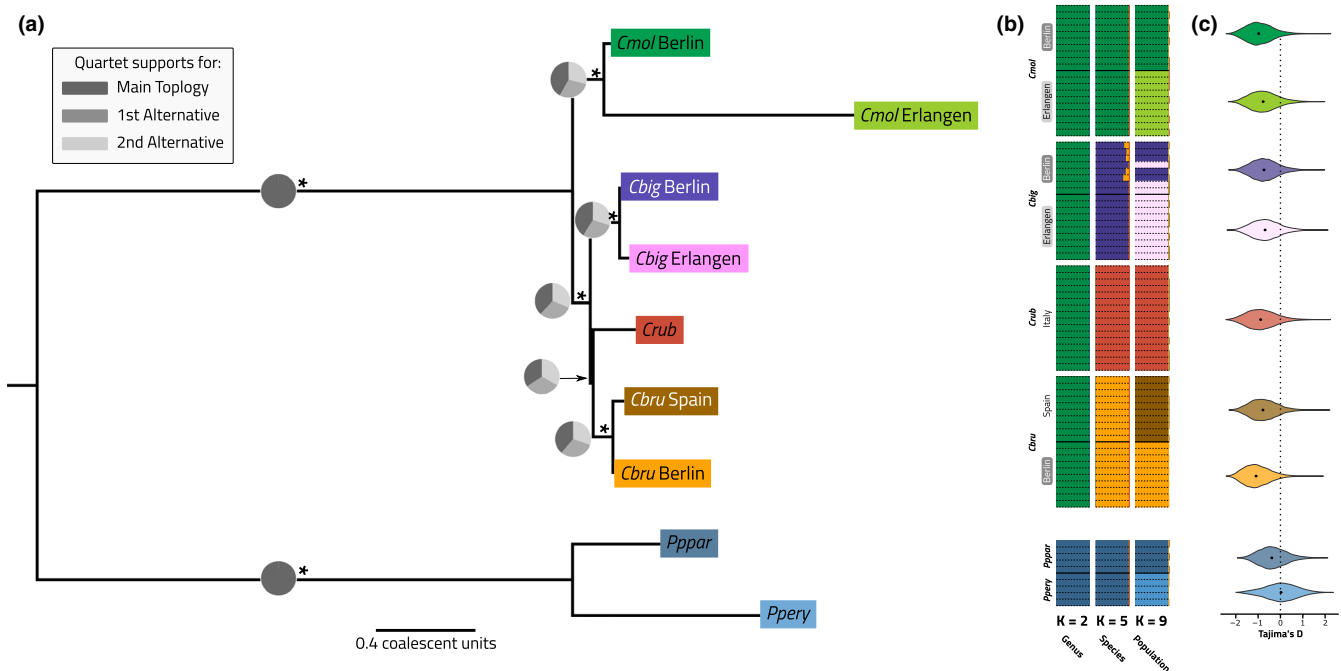


FIGURE 3 The radiation of *Chorthippus* leads to strong reproductive isolation in sympatry. (a) Multispecies coalescent using 3,963 independent nuclear gene trees shows that monophyletic species of *Chorthippus* rapidly radiated from a recent common ancestor. Pie charts show the quartet support at each node and reflect the amount of discordant gene trees caused by incomplete lineage sorting and/or gene flow. Branch lengths reflect coalescent times of T/N_e ; *mark branch support of 1.0 PP. (b) Colour of the bar plots show the cluster assignment of each individual based on 554,849 SNPs, at three K relevant levels of genetic structure (see Figure S10 for all K clusters). (c) Distribution of Tajima's D for 9,143 genes [Colour figure can be viewed at wileyonlinelibrary.com]

In agreement with what was observed in the mitochondrial genome, absolute divergence (D_{xy} , Figure 4a) was highest in all comparisons between *Chorthippus* and *Pseudochorthippus* populations (medians = 0.00471–0.00476). Yet, divergence between *Chorthippus* species (medians = 0.00155 – 0.00170) was similar to that between *Pseudochorthippus* subspecies (median = 0.00142).

3.3 | Population genetic analyses

3.3.1 | Population structure

PC1 explains a large fraction of the genetic variance (18.6%) and separates *Pseudochorthippus* from *Chorthippus* individuals. PC2, 3, and 4 explain relatively similar fractions of variance (2.91, 2.55, and 2.11%, respectively), with PC2 separating the two *P. parallelus* subspecies. Only PC3 and PC4 separate *Chorthippus* species, reflecting a genetic similarity between most species, with the exception of *Cmol* (Figure S9).

The admixture analysis showed that likelihood increases with each K cluster (Table S6). We observe a marked difference between *Pseudochorthippus* and *Chorthippus* at K = 2, without genetic admixture (Figure 3b). From K = 3 to 5, all species became assigned to their own distinct clusters, with admixture only observed between

the sympatric *Cbig* and *Cbru* in Berlin (Figure S10). At K = 9, each locality formed its own cluster, without noticeable genetic admixture between species; the only exception are two populations of *Cbig* sampled in close geographic proximity.

3.3.2 | Genetic variability

The number of genes used to calculate Theta and Tajima's D ranged between 9,675 and 12,521, with 9,143 genes sampled across all populations for both parameters. The per-gene Theta values (Figure S11) were generally higher for *Chorthippus* populations (medians between 0.011 and 0.013), intermediate for *Pppar* (median = 0.007), and lowest for *Ppery* (median = 0.005). Mean Tajima's D is generally negative in all populations (medians between -1.093 and -0.039), with the exception of *Ppery* (median = 0.026), which conforms to the neutral expectation (Figure 3c).

The number of genes for the F_{ST} analysis varied from 9,228 to 12,390, and 8,935 genes were common across population comparisons. Per gene F_{ST} showed a high variance across all comparisons, but distributions were skewed to three modes (Figure 4b). All comparisons between *Pseudochorthippus* and *Chorthippus* populations showed the highest differentiation (medians = 0.678–0.727), the subspecies of *Pseudochorthippus* showed intermediate differentiation

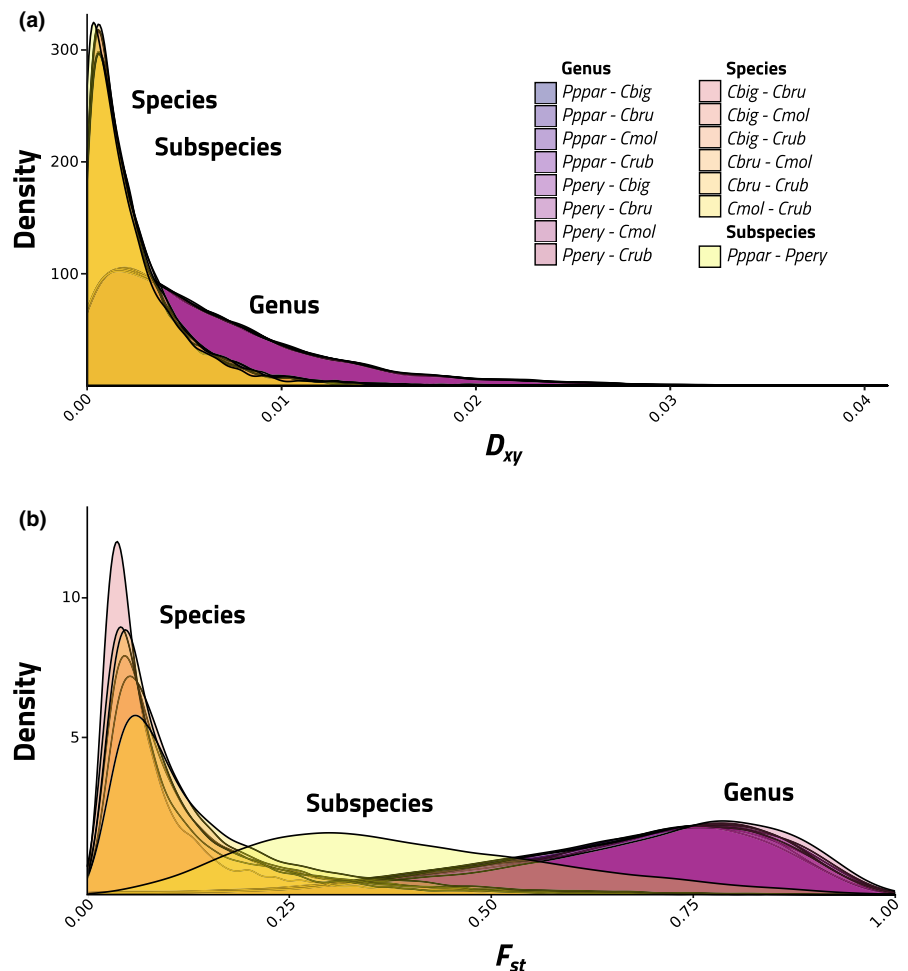


FIGURE 4 Taxa that diverged at similar times show contrasting patterns of genetic differentiation. (a) Absolute divergence (D_{xy}) across 3,963 common genes. (b) Genetic differentiation (F_{ST}) across 8,935 common genes [Colour figure can be viewed at wileyonlinelibrary.com]

(median = 0.371), and comparisons between *Chorthippus* species showed the lowest genomic differentiation of all comparisons (medians = 0.057–0.105).

3.3.3 | Genetic introgression

D-statistics showed significant deviations from the null expectation of no gene flow in 27 out of 35 relevant combinations (Figure S12, Table S7). Gene flow involved all species pairs within *Chorthippus*, except between *Cbig* and *Crub*. Notably, we found evidence of gene flow between the same species pairs, irrespective of whether we used allopatric or sympatric populations (Figure 5).

Considering the simpler demographic models, we observed that all technical replicates converged to similar likelihoods and parameter estimations. The No Migration model exhibits the lowest likelihood across the 19 taxa pairs, followed by the Ancestral Migration model (Figure 6a). The most likely models are Secondary Contact (five parameters), followed by the simpler Continuous Migration model (four parameters) (delta-likelihoods between 241 and 8,203). When using an adjusted likelihood ratio test (LRT) to compare these two models, Secondary Contact is always significantly more likely (p -value $< 5.1 \times 10^{-6}$; Table S8). When considering the two

more complex demographic models that include population growth (Continuous Migration with Growth and Secondary Contact with Growth), we still find the model incorporating secondary contact to be the most likely for 17 taxa pairs (delta-likelihoods between seven and 746). The only exceptions are *Cmol* Berlin versus *Crub*, and *Pppar* versus *Ppery*, for which both models are equally likely. Out of the 17 comparisons showing that the model of Secondary Contact with Growth is the most likely, 10 show that it is significantly better fitting than Continuous Migration with Growth, favouring a demographic history involving a period of genetic isolation (Table S8). Some of the taxa pairs for which the two models are not significantly different are currently allopatric (*Crub* versus *Cmol*, and *Crub* versus *Cbig*), or have established recent secondary contact in a narrow hybrid zone (*Cppar* versus *Cpery*; Hewitt, 1993), suggesting that our data shows limited power in estimating such complex demographic models with seven parameters. Therefore, for parameter estimation, we focus on the simpler Secondary Contact demographic model (five parameters) that showed stronger convergence and clearer likelihood peaks.

Inferred divergence times for all population pairs were similar, ranging from $0.00235/\mu$ – $0.00302/\mu$ (where μ is the number of mutations per site per generation for the transcriptome, Figure 6c, Table S9). Times of secondary contact were more variable, ranging from $0.00035/\mu$ – $0.00107/\mu$ (or 13%–35% of the total divergence

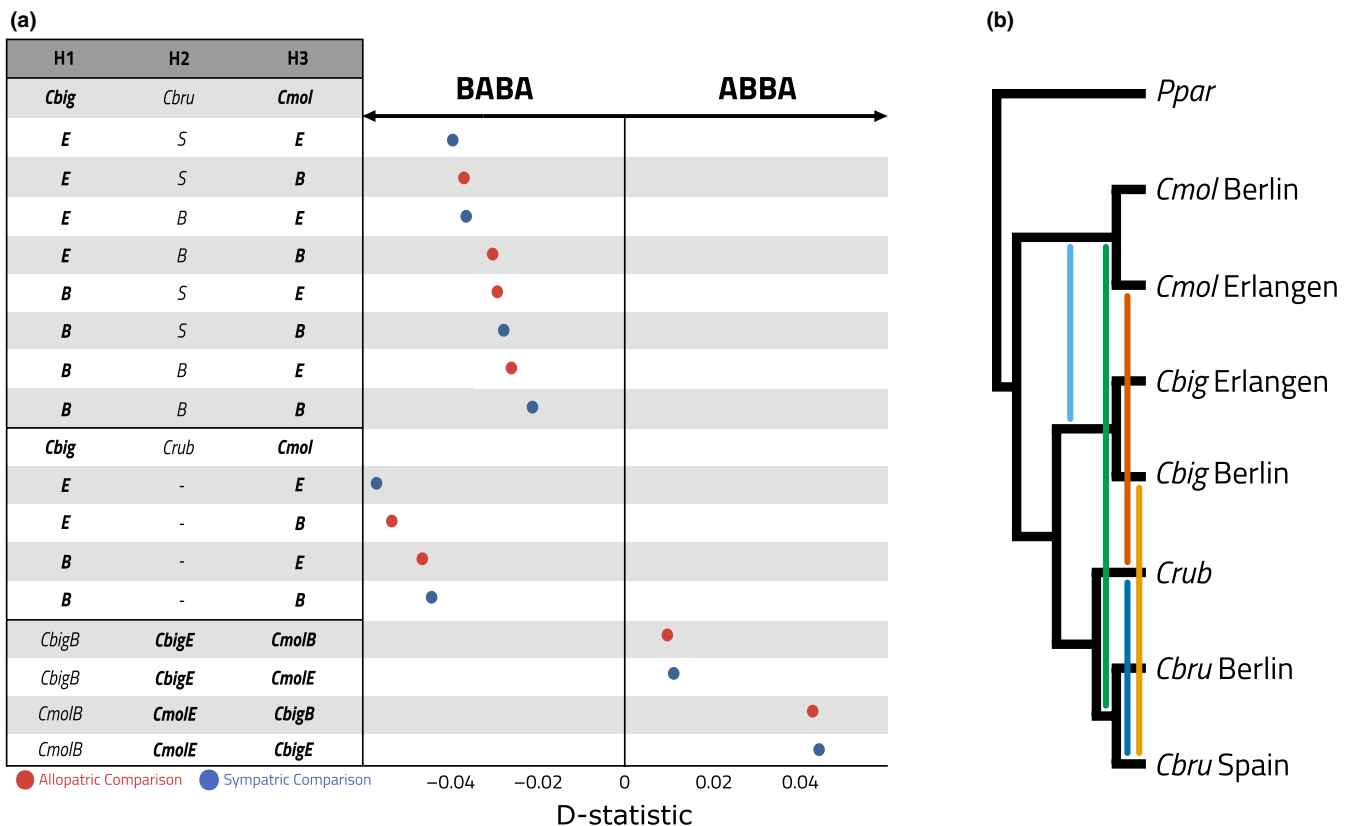


FIGURE 5 Extensive introgression during the radiation of *Chorthippus*. (a) D-statistics are significant irrespective of taxa being sampled in allopatry (orange) or sympatry (blue), consistent with ancestral gene flow; see Figure S12 and Table S7 for all pairwise comparisons. (b) Interspecific gene flow for all species pairs, according to the D-statistic. When both populations of the same species are involved with gene flow with a second species, it is most parsimonious that gene flow has occurred in the ancestor of each species, rather than between terminal branches [Colour figure can be viewed at wileyonlinelibrary.com]

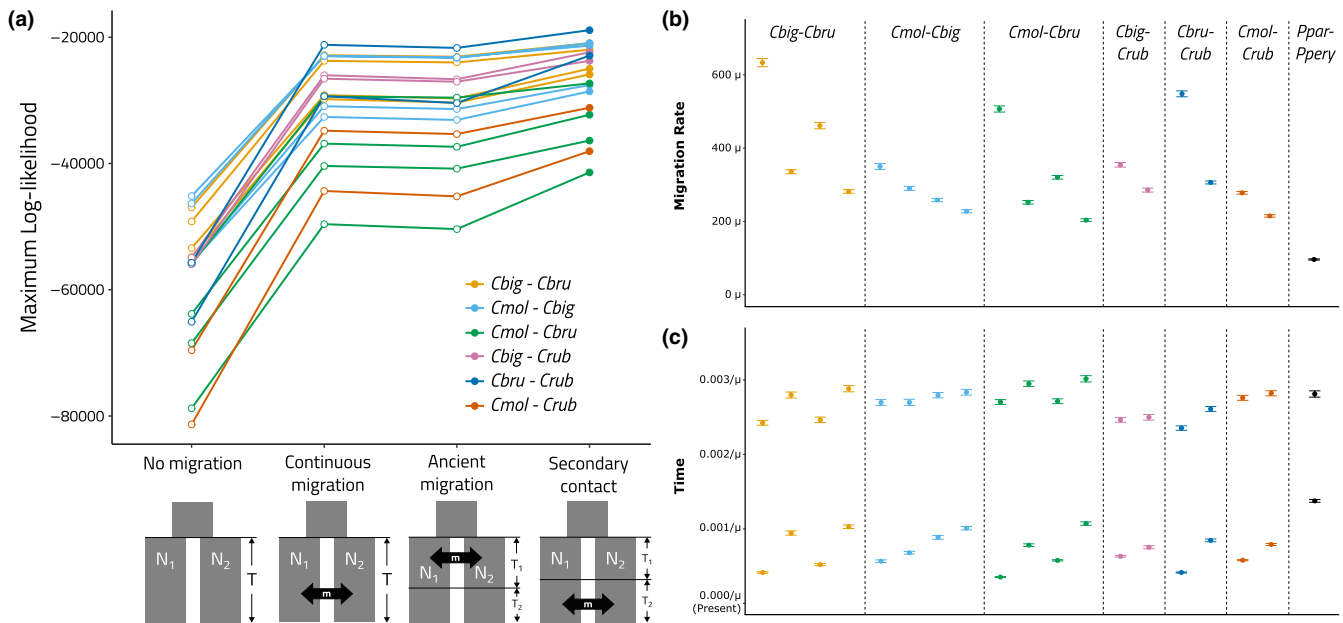


FIGURE 6 Demographic modelling is most consistent with an initial period of genetic isolation followed by introgression. (a) Maximum likelihood values (top) from demographic modelling of polymorphisms considering four simplified models of speciation (bottom). Full circles show the most likely model after accounting for physical linkage using an adjusted likelihood ratio test. (b) Maximum likelihood parameter estimates of migration rates between each taxa pair with 95% confidence intervals. (c) Maximum likelihood parameter estimates of time of population split (upper point) and secondary contact (lower point) with 95% confidence intervals. (b) and (c) are both represented in units scaled to mutation rate. See Tables S8 and S9 for results on model selection and parameter estimation. [Colour figure can be viewed at wileyonlinelibrary.com]

time) in *Chorthippus* population pairs and $0.00137/\mu$ (or 49% of the total divergence time) in the *Pseudochorthippus* subspecies (Figure 6c, Table S9). Secondary contact times were most recent in each species with a sympatric comparison in the northernmost sympatric pair (Berlin $0.00041/\mu$ – $0.00057/\mu$ or 13%–21% of the total divergence time). Migration rates were high between *Chorthippus* species (203.48μ – 633.33μ) relative to that found between the parapatric subspecies of *Pseudochorthippus* (97.57μ) (Figure 6b, Table S9). Northernmost sympatric pairs exhibited the highest migration rates for their species comparisons (Berlin, 350.12 – 633.33μ). Estimates of effective population sizes were highly convergent across independent datasets (Figure S13), being higher in populations of *Chorthippus* (0.0031 – $0.0080/\mu$) relative to populations of *Pseudochorthippus* (0.0016 – $0.0025/\mu$).

4 | DISCUSSION

4.1 | The rapid radiation of *Chorthippus* species

Genomic studies on ecological radiations primarily driven by divergent natural selection have shown that speciation can be relatively rapid, often leading to large amounts of incomplete lineage sorting that do not lead to monophyletic nominal species (e.g., Lamichhaney et al., 2015; Soria-Carrasco et al., 2014). It is yet unclear how rapid nonecological radiations are when sexual selection seemingly plays a primary role in speciation, and whether such radiations lead to monophyletic species.

Here, we study the radiation of grasshopper species of the genus *Chorthippus* to provide insights into these questions.

Mitochondrial studies in grasshoppers have been challenging due to abundant nuclear copies of mitochondrial genes that become coamplified during PCR (“numts”; Bensasson et al., 2000, 2001). Bypassing PCR using RNA sequencing, we were able to assemble full mitochondrial genomes with high sequencing depth ($>8,000\times$) and estimate a first time-calibrated phylogeny for this group. Our analyses of the Acrididae mitogenomes confirm that *Pseudochorthippus parallelus*, previously recognized as *Chorthippus parallelus* in the speciation literature (e.g., Butlin, 1998; Hewitt, 1993), is in fact a highly divergent lineage with respect to other *Chorthippus* species (BP = 100%). We estimate it to have diverged from *Chorthippus* some 5.9 Ma, during the transition between the Miocene and Pliocene, followed by the split of the genus *Euchorthippus*, of the Asian *Chorthippus chinensis*, and of the genus *Gomphoceris*. The most recent common ancestor (MRCA) of the focal species of European *Chorthippus* dates to around 506 ka, in the mid-Pleistocene. Interestingly, *Gomphocerippus rufus*, a species characterized by a distinctive mating display and clubbed antenna (Elsner & Huber, 1969; Jacobs, 1953), is nested inside the European *Chorthippus* clade, suggesting that this lineage belongs to the same radiation and should instead be classified as *Chorthippus*. None of the focal *Chorthippus* species is monophyletic in respect to the mitochondria. Likewise, subspecies of *Pseudochorthippus parallelus* do not form reciprocally monophyletic clades, and their MRCA dates to around 476 ka, also within the mid-Pleistocene. These time estimates imply that *Chorthippus* species must have survived through

the multiple glacial and interglacial periods that characterize phylogeographic patterns of European species, including *Pseudochorthippus parallelus* (Hewitt, 2000).

Because mitochondrial markers are often discordant from nuclear markers in animals (Toews & Brelsford, 2012), we estimated thousands of independent nuclear genes. *Pseudochorthippus* subspecies usually form two reciprocally monophyletic clades (BP = 100%), and *Chorthippus* species are paraphyletic (Figure S6). Moreover, absolute divergence across nuclear genes (D_{xy}) is similar between *Pseudochorthippus* subspecies (median = 0.00142) and *Chorthippus* species (medians = 0.00155–0.00170), again reflecting that those lineages diversified around a similar time. *Pseudochorthippus* and *Chorthippus* are about 3.3 times more divergent than taxa within these genera (medians = 0.00471–0.00476), confirming that these lineages should be considered different genera. Considering these recent times since divergence for the *Chorthippus* species, and the large effective population sizes typical of invertebrates (Leffler et al., 2012), it is not surprising that we find a large amount of incomplete lineage sorting across independent gene trees (Figure S6), as we also observe in the mitochondrial tree (Figure 2).

By accounting for incomplete lineage sorting through the multispecies coalescent, we estimate a well-supported species tree where every species forms a monophyletic clade, irrespective of whether we consider the 84 individuals separately (Figure S8; PP > 0.9) or grouped by the seven sampling localities (Figure 3a, PP = 1.0). Our species tree shows that *C. mollis* was the first species to split, also in agreement with the differentiation reflected in the Principal Component (Figure S9) and admixture (Figure S10) analyses. Most probably, *C. biguttulus* split next, followed by *C. rubratibialis* and *C. brunneus*. Yet, the phylogenetic relationship between *C. rubratibialis* and the other two species is not strongly supported (PP = 0.67), and it is unlikely to be resolved with more data than the 12,735 genes used here. A previous study based on song (Ragge et al., 1990) hypothesized that *C. rubratibialis* has a hybrid origin between *C. biguttulus* and *C. mollis*, but a different sampling strategy and methodologies are required to test that hypothesis. Based on the short branch lengths estimated in our species tree (in coalescent units of T/N_e), the radiation of *Chorthippus* spp. has probably occurred very rapidly and/or with very large effective population sizes (which may be due to high population size and/or a high level of gene flow that is not accounted for by the tree model). This led to a large fraction of gene discordance reflected by the quartet supports (pie charts in Figure 3a), where only 34%–41% of the quartets share the main topology.

Together, our results suggest that the radiation of *Chorthippus* species fits several key criteria for ecological radiations (Schluter, 2000). First, species form a monophyletic clade (Figure 2), together with *Gomphocerippus rufus* that probably belonged to the same radiation. Second, the radiation is relatively recent (up to 506 ka; Table S5), on the lower range of what has been described for well-studied ecological radiations (e.g., ranging from 200 ka to 10 Ma in cichlid fishes (Irisarri et al., 2018) and from 2 to 18 Ma in stick insects (Riesch et al., 2017)). Third, the radiation led to a burst of monophyletic

species (Figure 3a), suggesting that behavioural isolation evolved once in each nominal species, and that this phylogenetic signal has not been completely eroded by homogenizing gene flow. These findings set the radiation of *Chorthippus* as a valuable system to provide key insights on how sexual selection can drive species formation in the absence, or reduced contribution, of other selective forces.

4.2 | Leaky genomic barriers between *Chorthippus* species, despite behavioural incompatibilities

When speciation is primarily driven by sexual selection, gene flow is hypothesized to be a major impediment for the maintenance of sympatric species, because assortative mating needs to be strong enough to maintain the association between species-specific signaling and preference traits (Felsenstein, 1981; Fry, 2003; Servedio et al., 2011). Here, we assess whether species of *Chorthippus* that presumably diverged by sexual selection frequently hybridize in sympatry, and whether they have experienced significant gene flow during their evolution.

Our population structure analyses revealed well defined genetic clusters at the genus, species, and population level (Figure 3b). When we allow the maximum number of clusters ($K = 9$), we do not observe any substantial admixture between species, but only between populations of the same species (*C. biguttulus*) sampled in close geographic proximity (Berlin and Erlangen). Notably, considering 4–8 clusters, we observe that some *C. biguttulus* individuals from Berlin contain up to 18% of the genome of the sympatric *C. brunneus* (Figure S10), probably because some admixture is shared among most individuals of that population at similar proportions (Lawson et al., 2018). This indicates that ongoing hybridization between *Chorthippus* species is too rare to be detected by our sampling (up to 28 individuals of three species in sympatry), as also suggested by other studies based on song (Baur, 2006; Jacobs, 1963; Perdeck, 1958). Yet, it is important to note that these species closely resemble each other in morphology (Ragge & Reynolds, 1988), and thus sampling biases against intermediate hybrids may lead to an underestimation of real hybridization rates in nature.

Given that most of these species are sympatric across Europe (Figure 1), even rare hybridization may lead to extensive introgression, as observed in well-known examples of ecological radiations (e.g., Brawand et al., 2014; Heliconius Genome Consortium, 2012; Lamichhaney et al., 2015). The D-statistics offer a direct test of interspecific gene flow during the radiation of *Chorthippus*, because an overrepresentation of ABBA or BABA topologies cannot be solely explained by incomplete lineage sorting. Out of the 35 tested comparisons, 27 show significant deviations from the neutral expectation, indicating that gene flow among almost all species pairs has been pervasive during their evolution (Figure S12). Because our sampling often includes the same species pair in allopatry and in sympatry, we used this to explicitly test if gene flow is only observed when species were sampled in sympatry, consistent with a major role of recent gene flow. Our results show significant deviations from the

neutral expectation regardless of the geographic overlap (Figure 5), suggesting that interspecific gene flow has started before the species established their current distributions.

Our findings of similar times of divergence between different species of *Chorthippus* and between the two subspecies of *Pseudochorthippus parallelus*, both in mitochondrial (Figure 2) and nuclear markers (Figures 3a and 4a), would normally suggest similar levels of genomic differentiation. Instead, our results show that *Chorthippus* species that rarely hybridize in sympatry (all medians of pairwise F_{ST} are between 0.057 and 0.105) are at least 3.5 times less differentiated than the *Pseudochorthippus parallelus* subspecies ($F_{ST} = 0.371$) that form a clinal hybrid zone in parapatriy (Hewitt, 1993). Although absolute divergence (D_{xy} , Figure 4a) is mainly affected by time since speciation, differentiation (F_{ST} , Figure 4b) is also affected by demographic parameters such as effective population size (N_e) and migration rates (m) (Nachman & Payseur, 2012). Thus, we hypothesize that the low levels of genetic differentiation can be explained by larger N_e and/or m in *Chorthippus* spp. relative to *Pseudochorthippus* subspecies. Our parameter estimations using demographic modeling (Figure 6b (m), 6c (T), S13 (N_e), Table S9) confirm this hypothesis. Consistent with our observation of similar D_{xy} , our estimates of time since divergence are similar between all species pairs (between $0.0030/\mu$ and $0.0024/\mu$). Yet, effective population sizes are between 1.5 to 6.7 times larger in *Chorthippus* relative to *Pseudochorthippus* (Figure S13, Table S9), and migration rates among taxa are between 2 to 6.5 times higher in *Chorthippus* relative to *Pseudochorthippus* (Figure 6b, Table S9). Such differences in N_e and m lead to higher levels of standing genetic variation in *Chorthippus*, explaining the lower observed levels of genomic differentiation despite apparently higher levels of the reproductive isolation. Such large differences in migration rates may simply reflect differences in geographic opportunities for hybridization and introgression, since *Chorthippus* species are sympatric throughout most of the European continent (Figure 1), while *Pseudochorthippus* subspecies are parapatric in a narrow hybrid zone in the Pyrenees.

Values of population migration rates ($2 Nm$) greater than or equal to one can prevent populations from accumulating divergence (Wright, 1931). Our observed high values of migration rate translate into $2 Nm$ values between 1.6 and nine migrants by population per generation (Table S9), implying that strong selection is required to oppose the homogenizing effect of gene flow (Pinho & Hey, 2010). Such leaky genomic barriers ($F_{ST} \sim 0.08$), despite strong behavioural incompatibilities have also been observed in other systems where sexual selection is thought to play a more prominent role than natural selection, such as in crows (Vijay et al., 2016; $F_{ST} \sim 0.13$) and in warblers (Toews et al., 2016; $F_{ST} \sim 0.005$). In such systems, loci underlying male signaling and female preference are thought to be under strong selection, resisting genome wide gene flow. Likewise, our results suggest that the loci underlying strong behavioural isolation in these grasshopper species must also be resistant to the pervasive gene flow experienced by most of the genome. Follow up studies using admixture mapping in

hybrid zones or quantitative trait loci (QTL) experiments (e.g., Xu & Shaw, 2019a, b) are promising ways to identify genes underlying behavioural isolation.

4.3 | Historical geographic isolation occurred during the evolution of behavioural isolation

The evolution of assortative mating is hard to explain in sympatric species experiencing continuous gene flow because recombination can prevent the initial fixation of male signals and female preferences within each species (Felsenstein, 1981; Fry, 2003; Servedio et al., 2011). Although *Chorthippus* species are currently sympatric across Europe, most of their range has only become available during the last ~12,000 years. It is therefore unclear if reproductive isolation was established in a similar geographic setting to today. Using demographic modeling, we clarify the demographic history that has led to the strong behavioural isolation that currently maintains species boundaries in *Chorthippus*.

When fitting the observed patterns of genomic variability to relatively simple demographic models, we always find that the Secondary Contact model shows the highest likelihood for all taxa pairs (Figure 6a), and that this model is always significantly more likely than the second most likely model showing continuous migration after performing a likelihood ratio test (LRT; Table S8). When testing biologically more realistic models including change in effective population size, we still find the model incorporating secondary contact to be the most likely in 17 of the 19 pairs and equivalent to a model with continuous migration in the remaining two (Table S8). Although this difference is significant in only 10 of the 17 cases, nonsignificant taxa pairs include comparisons that are currently allopatric (i.e., with the Italian endemic species *C. rubratibialis*), or taxa that established secondary contact in a hybrid zone (*Pseudochorthippus*, Hewitt, 1993). Thus, we interpret this as a power limitation in estimating too many parameters with the current data, and refer to the simpler Secondary Contact model for parameter estimation (Figure 6). Our time estimates suggest that *Chorthippus* species have spent most of their time in genetic isolation, with gene flow occurring only from 13%–35% of the total time since divergence (Figure 6c). Given that our mitochondrial time tree dates the MRCA of all *Chorthippus* to around 506 ka (Figure 2), *Chorthippus* species must have experienced multiple episodes of contraction and expansion during the multiple glacial periods that have since affected the European continent (Hewitt, 2000). A demographic scenario of secondary contact is also consistent with our estimates of Tajima's D (Figure 3c), which are negative for all populations sampled in previously glaciated areas. Future studies sampling these species at or near the described refugia in Eurasia may help testing whether *Chorthippus* species diverged in different glacial refugia.

Together, our demographic analyses suggest that the evolution of the strong behavioural barriers observed among currently sympatric *Chorthippus* species (von Helversen & von Helversen, 1994;

Perdeck, 1958) probably evolved during the long periods of geographic isolation, possibly associated with the glacial periods that affected Europe since the mid-Pleistocene. Historical periods of geographic isolation have also been recently described in emblematic ecological radiations primarily maintained by divergent natural selection, such as the radiation of cichlid fishes (Meier et al., 2017) and of sea snails (Stankowski et al., 2020), potentially facilitating the establishment of strong reproductive isolation, and the subsequent coexistence of taxa in sympatry. Likewise, in the context of nonecological radiations where sexual selection seemingly plays a major role in reproductive isolation, historical geographic isolation might provide opportunity for different sexually selected characters to develop and be driven to fixation by Fisherian runaway selection in a mutation-order process (Mendelson et al., 2014), leading to the coevolution of male song and female preference. Once that association is established within a species, behavioural barriers can resist the homogenizing effects of gene flow, resulting in cryptic species that are genetically, morphologically, and ecologically similar, but that otherwise generally behave as good biological species.

ACKNOWLEDGEMENTS

We thank the Spanish and French authorities from the Gobierno de Aragón and the French Parc National des Pyrénées for the permission to sample the individuals used for this study. RJP was funded by the European Union's Horizon 2020 research and innovation programme, under the Marie Skłodowska-Curie grant agreement No. 658706. It was supported by a Juan de la Cierva-Incorporación postdoctoral fellowship (IJCI-2016- 29566) from the Spanish Ministry for Science and Competitiveness (MINECO). Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under Project SNIC 2019/8-67. Photographs in Figure 1 were courtesy of C. Roestri and P. Fontana. We thank the four anonymous reviewers and the Editor R. Gillespie for the valuable feedback in an earlier version of the manuscript. Open access funding enabled and organized by ProjektDEAL.

AUTHOR CONTRIBUTIONS

The project was conceived by R.J.P. The data was produced by D.B.A., F.M., and R.J.P. The analyses were performed and interpreted by Z.J.N., B.Y., I.I., S.L., C.G.C., and R.J.P. Reagents and analytical tools were contributed by M.T.P.G., and R.J.P. The first version of the manuscript was drafted by Z.J.N. and R.J.P., and all remaining coauthors contributed to the final version.

DATA AVAILABILITY STATEMENT

Raw reads were deposited in the NCBI Sequence Read Archive (BioProject PRJNA665162; BioSample Acc. SAMN16264388–SAMN16264413). The Dryad archive (<https://doi.org/10.5061/dryad.pzgmbschj>) contains the de novo assembly of the transcriptome, coordinates of protein coding genes, coordinates of neutral sites, list of F_{ST} , D_{xy} , theta and Tajima's D per gene, the covariance

table for the PCA, the input Beagle file for the admixture analysis, the input xml file for the BEAST analysis, and all 19 two population SFS used for demographic analysis. Bioinformatic scripts are available in a public GitHub repository (https://github.com/zjnolen/chorthippus_radiation). (Nolen, 2020a; 2020b; 2020c; Liu, & Huang, 2008; Vanderbilt University, 2015; Berdan, Mazzoni, Waurick, Roehr, & Mayer, 2015; Shi & Huang, 2009; Gao, & Huang, 2010; Huang, & Liu, 2010; Sun, Zheng, & Huang, 2010; Zhao, Zheng, Huang, & Sun, 2010; Jiang, Ma, & Han, 2012; Zhang, Zhao, Zheng, & Huang, 2013; Song, Zhi, Liu, Yin, & Zhang, 2014; Xiao, Zhou, & Huang, 2008a; 2008b; Ma, Liu, Yang, & Kang, 2008).

ORCID

Zachary J. Nolen  <https://orcid.org/0000-0001-8146-8016>

Iker Irisarri  <https://orcid.org/0000-0002-3628-1137>

Clara Groot Crego  <https://orcid.org/0000-0002-4547-3608>

Ricardo J. Pereira  <https://orcid.org/0000-0002-8076-4822>

REFERENCES

- Baur, B. (2006). *Die Heuschrecken der Schweiz*. Bern Naturhistorisches Museum.
- Berdan, E. L., Mazzoni, C. J., Waurick, I., Roehr, J. T., & Mayer, F. (2015). Data from: Three species of *Chorthippus* grasshoppers. *C. biguttulus*, *C. brunneus*, and *C. mollis* raw sequence reads. NCBI SRA; PRJNA284873
- Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology & Evolution*, 16(6), 314–321. [https://doi.org/10.1016/S0169-5347\(01\)02151-6](https://doi.org/10.1016/S0169-5347(01)02151-6)
- Bensasson, D., Zhang, D.-X., & Hewitt, G. M. (2000). Frequent assimilation of mitochondrial DANN by grasshopper nuclear genomes. *Molecular Biology and Evolution*, 17(3), 406–415. <https://doi.org/10.1093/oxfordjournals.molbev.a026320>
- Berdan, E. L., Mazzoni, C. J., Waurick, I., Roehr, J. T., & Mayer, F. (2015). A population genomic scan in *Chorthippus* grasshoppers unveils previously unknown phenotypic divergence. *Molecular Ecology*, 24(15), 3918–3930. <https://doi.org/10.1111/mec.13276>
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., ... Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), 375–381. <https://doi.org/10.1038/nature13726>
- Brower, A. V. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 91(14), 6491–6495. <https://doi.org/10.1073/pnas.91.14.6491>
- Butlin, R. (1998). What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation. In Daniel J. Howard & Stewart H. Berlocher (eds.), *Endless forms: Species and speciation* (pp. 367–378). Oxford University Press.
- Butlin, R. K., & Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *American Naturalist*, 191(2), 155–172. <https://doi.org/10.1086/695136>

- Carleton, K. L., & Kocher, T. D. (2001). Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Molecular Biology and Evolution*, 18(8), 1540–1550. <https://doi.org/10.1093/oxfordjournals.molbev.a003940>
- Chevreaux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E. G., Wetter, T., & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research*, 14(6), 1147–1159. <https://doi.org/10.1101/gr.1917404>
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2020). *Orthoptera species file. Version 5.0/5.0*. Retrieved May 11, 2020, from <http://Orthoptera.SpeciesFile.org>
- Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33(2), 591–593. <https://doi.org/10.1093/molbev/msv255>
- Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10(1), 210. <https://doi.org/10.1186/1471-2148-10-210>
- da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrugal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, 30, 3–13. <https://doi.org/10.1016/j.margen.2016.04.012>
- Defaut, B. (2012). Implications taxonomiques et nomenclaturales de publications récentes en phylogénie moléculaire: 1. Les Gomphocerinae de France (Orthoptera, Acrididae). *Matériaux Orthoptériques Et Entomocénétiques*, 17, 15–20.
- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6), 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Elsner, N., & Huber, F. (1969). Die Organisation des Werbegesanges der Heuschrecke *Gomphocerippus rufus* L. in Abhängigkeit von zentralen und peripheren Bedingungen. *Zeitschrift Für Vergleichende Physiologie*, 65(4), 389–423.
- Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, 35(1), 124–138. <https://doi.org/10.1111/j.1558-5646.1981.tb04864.x>
- Finck, J., Kuntze, J., & Ronacher, B. (2016). Chemical cues from females trigger male courtship behaviour in grasshoppers. *Journal of Comparative Physiology A*, 202(5), 337–345. <https://doi.org/10.1007/s00359-016-1081-4>
- Finck, J., & Ronacher, B. (2017). Components of reproductive isolation between the closely related grasshopper species *Chorthippus biguttulus* and *C. mollis*. *Behavioral Ecology and Sociobiology*, 71(4), 70. <https://doi.org/10.1007/s00265-017-2295-3>
- Fry, J. D. (2003). Multilocus Models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution*, 57(8), 1735–1746. <https://doi.org/10.1111/j.0014-3820.2003.tb00582.x>
- Fumagalli, M., Vieira, F. G., Linderroth, T., & Nielsen, R. (2014). ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30(10), 1486–1487. <https://doi.org/10.1093/bioinformatics/btu041>
- Gao, J., & Huang, Y. (2010). Data from: *Gomphocerus licenti* mitochondrion, complete genome; NCBI GenBank; NC_013847.1
- Gayral, P., Melo-Ferreira, J., Glémin, S., Bierre, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., & Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLOS Genetics*, 9(4), e1003457. <https://doi.org/10.1371/journal.pgen.1003457>
- Gillespie, R. G., Bennett, G. M., De Meester, L., Feder, J. L., Fleischer, R. C., Harmon, L. J., Hendry, A. P., Knope, M. L., Mallet, J., Martin, C., Parent, C. E., Patton, A. H., Pfennig, K. S., Rubinoff, D., Schluter, D., Seehausen, O., Shaw, K. L., Stacy, E., Stervander, M., ... Wogan, G. O. U. (2020). Comparing adaptive radiations across space, time, and taxa. *Journal of Heredity*, 111(1), 1–20. <https://doi.org/10.1093/jhered/esz064>
- Gillespie, R., Howarth, F., & Roderick, G. (2001). Adaptive radiation. In S. A. Levin (Ed.), *Encyclopedia of biodiversity*. Academic Press.
- Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31(4), 1208–1211. <https://doi.org/10.1214/aoms/1177705693>
- Gosalvez, J., López-Fernandez, C., & Esponda, P. (1980). Variability of the DNA content in five Orthopteran species. *Caryologia*, 33(2), 275–281. <https://doi.org/10.1080/00087114.1980.10796840>
- Gottsberger, B., & Mayer, F. (2007). Behavioral sterility of hybrid males in acoustically communicating grasshoppers (Acrididae, Gomphocerinae). *Journal of Comparative Physiology A*, 193(7), 703–714. <https://doi.org/10.1007/s00359-007-0225-y>
- Gottsberger, B., & Mayer, F. (2019). Dominance effects strengthen premating hybridization barriers between sympatric species of grasshoppers (Acrididae, Orthoptera). *Journal of Evolutionary Biology*, 32(9), 921–930. <https://doi.org/10.1111/jeb.13490>
- Gouy, M., Guindon, S., & Gascuel, O. (2010). SEAVIEW Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2), 221–224. <https://doi.org/10.1093/molbev/msp259>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Grant, P. R., & Grant, B. R. (2011). *How and why species multiply: The radiation of Darwin's Finches*. Princeton University Press.
- Gray, D. A., & Cade, W. H. (2000). Sexual selection and speciation in field crickets. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26), 14449–14454. <https://doi.org/10.1073/pnas.97.26.14449>
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Gutenkunst, R. N. (2015). *Re. Nonparametric bootstrapping [Discussion post]*. dadi-user Google Group. <https://groups.google.com/g/dadi-user/c/kvzhF4XSyng/m/K2tqmXmjRP8J>
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Haas, B., & Papanicolaou, A. (2012). TRANSDCODER. Retrieved from Github Website: <https://github.com/TransDecoder/TransDecoder>
- Hahn, C., Bachmann, L., & Chevreaux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13), e129. <https://doi.org/10.1093/nar/gkt371>

- Harz, K. (1975). *The Orthoptera of Europe. II. Series entomologica, Vol. 11.* : Dr. W. Junk Publ.
- Hawiltschek, O., Morinière, J., Lehmann, G. U. C., Lehmann, A. W., Kropf, M., Dunz, A., Glaw, F., Detcharoen, M., Schmidt, S., Hausmann, A., Szucsich, N. U., Caetano-Wyler, S. A., & Haszprunar, G. (2017). DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany and Switzerland. *Molecular Ecology Resources*, 17(5), 1037–1053. <https://doi.org/10.1111/1755-0998.12638>
- Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), 94–98. <https://doi.org/10.1038/nature11041>
- Hewitt, G. M. (1993). After the ice: *Parallelus* meets *erythropus* in the Pyrenees. R. G. Harrison (ed.) *Hybrid zones and the evolutionary process* (pp. 140–164). Oxford University Press.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913. <https://doi.org/10.1038/35016000>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Huang, Y., & Liu, N. (2010). Data from: *Arcyptera coreana* mitochondrion, complete genome; NCBI GenBank; NC_013805.1
- Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., Fischer, C., Lemmon, A. R., Lemmon, E. M., Thallinger, G. G., Sturmbauer, C., & Meyer, A. (2018). Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications*, 9(1), 1–12. <https://doi.org/10.1038/s41467-018-05479-9>
- Jacobs, W. (1953). Verhaltensbiologische Studien an Feldheuschrecken. *Zeitschrift für Tierpsychologie*, Beiheft I, 1–228.
- Jacobs, W. (1963). Über das Singen der Feldheuschrecke *Chorthippus biguttulus* (L.) in verschiedenen Bereichen Mittel- und Westeuropas. *Zeitschrift für Tierpsychologie*, 20(4), 446–460. <https://doi.org/10.1111/j.1439-0310.1963.tb01164.x>
- Jiang, G. F., Ma, J. D., & Han, M. (2012). Data from: *Ceracris kiangsu* isolate H6012 mitochondrion, complete genome; NCBI GenBank; NC_019994.1
- Jiggins, C. D. (2017). *The ecology and evolution of Heliconius butterflies*. Oxford University Press.
- Kalyanamorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15(1), 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Lamichanay, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martínez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375. <https://doi.org/10.1038/nature14181>
- Langerhans, R. B., & Riesch, R. (2013). Speciation by selection: A framework for understanding ecology's role in speciation. *Current Zoology*, 59(1), 31–52. <https://doi.org/10.1093/czoolo/59.1.31>
- Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 3258. <https://doi.org/10.1038/s41467-018-05257-7>
- Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLOS Biology*, 10(9), e1001388. <https://doi.org/10.1371/journal.pbio.1001388>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for Eukaryotic genomes. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503>
- Liu, Y., & Huang, Y. (2008). Data from: *Chorthippus chinensis* mitochondrion, complete genome; NCBI GenBank; NC_011095.1
- Lindgreen, S. (2012). AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Research Notes*, 5(1), 337. <https://doi.org/10.1186/1756-0500-5-337>
- Liu, Y., & Huang, Y. (2008). Sequencing and analysis of complete mitochondrial genome of *Chorthippus chinensis* Tarb. *Chinese Journal of Biochemistry and Molecular Biology*, 4, 329–335.
- Ma, C., Liu, C., Yang, P., & Kang, L. (2008). Data from: *Oedaleus decorus asiaticus* mitochondrion, complete genome; NCBI GenBank; NC_011115.1
- Mason, D. J., Butlin, R. K., & Gacesa, P. (1995). An unusual mitochondrial DNA polymorphism in the *Chorthippus biguttulus* species group (Orthoptera: Acrididae). *Molecular Ecology*, 4(1), 121–126. <https://doi.org/10.1111/j.1365-294X.1995.tb00199.x>
- Mayer, F., Berger, D., Gottsberger, B., & Schulze, W. (2010). Non-ecological radiations in acoustically communicating grasshoppers? In M. Glaubrecht (Ed.), *Evolution in action: Case studies in adaptive radiation, speciation and the origin of biodiversity* (pp. 451–464). Prestel.
- Mayr, E. (1963). *Animal species and evolution*. Harvard University Press.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MAPREDUCE framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Meier, J. I., Sousa, V. C., Marques, D. A., Selz, O. M., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology*, 26(1), 123–141. <https://doi.org/10.1111/mec.13838>
- Mendelson, T. C., Martin, M. D., & Flaxman, S. M. (2014). Mutation-order divergence by sexual selection: Diversification of sexual signals in similar environments as a first step in speciation. *Ecology Letters*, 17(9), 1053–1066. <https://doi.org/10.1111/ele.12313>
- Mendelson, T. C., & Shaw, K. L. (2005). Rapid speciation in an arthropod. *Nature*, 433(7024), 375–376. <https://doi.org/10.1038/433375a>
- Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12), i44–i52. <https://doi.org/10.1093/bioinformatics/btv234>
- Momigliano, P., Florin, A.-B., & Merilä, J. (2020). Biases in demographic modelling affect our understanding of the process of speciation. *BioRxiv*. <https://doi.org/10.1101/2020.06.03.128298>
- Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 409–421. <https://doi.org/10.1098/rstb.2011.0249>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nolen, Z. J., Yildirim, B., Irisarri, I., Liu, S., Groot Crego, C., Amby, D. B., Mayer, F., Gilbert, M. T. P., & Pereira, R. J. (2020a). Data from: Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus grasshoppers*. *Dryad*. <https://doi.org/10.5061/dryad.pzgmbschj>

- Nolen, Z. J., Yildirim, B., Irisarri, I., Liu, S., Groot Crego, C., Amby, D. B., Mayer, F., Gilbert, M. T. P., & Pereira, R. J. (2020b). Data from: Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus grasshoppers*; NCBI SRA; PRJNA665162
- Nolen, Z. J., Yildirim, B., Irisarri, I., Liu, S., Groot Crego, C., Amby, D. B., Mayer, F., Gilbert, M. T. P., & Pereira, R. J. (2020c). Data from: Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus grasshoppers*. *GitHub*. <https://doi.org/10.5281/zenodo.4238318>
- Nosil, P. (2012). *Ecological speciation*. Oxford University Press.
- Otte, D. (1994). *The crickets of Hawaii: Origin. Systematics and evolution*. : The Orthopterists Society.
- Panhuis, T. M., Butlin, R., Zuk, M., & Tregenza, T. (2001). Sexual selection and speciation. *Trends in Ecology & Evolution*, 16(7), 364–371. [https://doi.org/10.1016/S0169-5347\(01\)02160-7](https://doi.org/10.1016/S0169-5347(01)02160-7)
- Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: The Mid-Aegean Trench calibration. *Molecular Biology and Evolution*, 27(7), 1659–1672. <https://doi.org/10.1093/molbev/msq051>
- Perdeck, A. C. (1958). The isolating value of specific song patterns in two sibling species of grasshoppers (*Chorthippus brunneus* Thunb. and *C. biguttulus* L.). *Behaviour*, 12(1/2), 1–75. <https://doi.org/10.1163/156853957X00074>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). POPGENOME: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Pinho, C., & Hey, J. (2010). Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 215–230. <https://doi.org/10.1146/annurev-ecolsys-102209-144644>
- Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M.-O., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26(19), 5245–5263. <https://doi.org/10.1111/mec.14266>
- Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11(3), 175–180. <https://doi.org/10.1038/nrg2718>
- R Core Team (2017). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>
- Ragge, D. R., & Reynolds, W. J. (1988). The songs and taxonomy of the grasshoppers of the *Chorthippus biguttulus* group in the Iberian Peninsula (Orthoptera: Acrididae). *Journal of Natural History*, 22(4), 897–929. <https://doi.org/10.1080/00222938800770611>
- Ragge, D., Reynolds, W., & Willemse, F. (1990). The songs of the European grasshoppers of the *Chorthippus biguttulus* group in relation to their taxonomy, speciation and biogeography (Orthoptera, Acrididae). *Boletín de Sanidad Vegetal. Plagas (Spain)*.(Spec. No. 20).
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, 105(3), 767–779.
- Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., de Carvalho, C. F., Safran, R. J., Sandoval, C. P., Feder, J., Gries, R., Crespi, B. J., Gries, G., Gompert, Z., & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, 1(4), 1–13. <https://doi.org/10.1038/s41559-017-0082>
- Ritchie, M. G. (2007). Sexual selection and speciation. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 79–102. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095733>
- Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of Lake Whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057–2074. <https://doi.org/10.1093/gbe/evx150>
- Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation, and nonecological speciation. *Trends in Ecology & Evolution*, 24(7), 394–399. <https://doi.org/10.1016/j.tree.2009.02.007>
- Schluter, D. (2000). *The ecology of adaptive radiation*. : OUP.
- Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, 9(5), 1056–1082. <https://doi.org/10.1038/nprot.2014.063>
- Servedio, M. R., & Boughman, J. W. (2017). The role of sexual selection in local adaptation and speciation. *Annual Review of Ecology, Evolution, and Systematics*, 48(1), 85–109. <https://doi.org/10.1146/annurev-ecolsys-110316-022905>
- Servedio, M. R., van Doorn, G. S., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: 'Magic' but not rare? *Trends in Ecology & Evolution*, 26(8), 389–397. <https://doi.org/10.1016/j.tree.2011.04.005>
- Shi, H., & Huang, Y. (2009). Data from: *Phlaeoba albonema* mitochondrion, complete genome; NCBI GenBank; NC_011827.1
- Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693–702. <https://doi.org/10.1534/genetics.113.154138>
- Song, W., Zhi, Y., Liu, G., Yin, H., & Zhang, D. (2014). Data from: *Orinhippus tibetanus* mitochondrion, complete genome; NCBI GenBank; NC_023467.1
- Soraggi, S., Wiuf, C., & Albrechtsen, A. (2018). Powerful inference with the D-statistic on low-coverage whole-genome data. *G3: Genes, Genomes, Genetics*, 8(2), 551–566. <https://doi.org/10.1534/g3.117.300192>
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J., & Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738–742. <https://doi.org/10.1126/science.1252136>
- Stankowski, S., Westram, A. M., Zagrodzka, Z. B., Eyres, I., Broquet, T., Johannesson, K., & Butlin, R. K. (2020). The evolution of strong reproductive isolation between sympatric intertidal snails. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1806), 20190545. <https://doi.org/10.1098/rstb.2019.0545>
- Strimmer, K., & von Haeseler, A. (1997). Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*, 94(13), 6815–6819. <https://doi.org/10.1073/pnas.94.13.6815>
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1), vey016. <https://doi.org/10.1093/ve/vey016>
- Sun, H., Zheng, Z., & Huang, Y. (2010) Data from: *Gomphocerippus rufus* mitochondrion, complete genome; NCBI GenBank; NC_014349.1
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>
- Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, 26(17), 2313–2318. <https://doi.org/10.1016/j.cub.2016.06.034>
- Vanderbilt University (2015). Data from: *Chorthippus parallelus* parallelus raw sequence reads; NCBI SRA; PRJNA280771
- Vedenina, V., & Mugue, N. (2011). Speciation in gomphocerine grasshoppers: Molecular phylogeny versus bioacoustics and courtship behavior. *Journal of Orthoptera Research*, 20(1), 109–125. <https://doi.org/10.1665/034.020.0111>
- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, 7(1), 1–10. <https://doi.org/10.1038/ncomms13195>
- von Helversen, O., & von Helversen, D. (1994). Forces driving coevolution of song and song recognition in grasshoppers. *Fortschritte Der Zoologie*, 253.
- Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Wang, X., Zhang, Y. I., Yang, M., ... Kang, L. E. (2014). The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, 5(1), 1–9. <https://doi.org/10.1038/ncomms3957>
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3), 543–548. <https://doi.org/10.1093/molbev/msx319>
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2), 256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97–159.
- Xiao, L. L., Zhou, Z. J., & Huang, Y. (2008) Data from: *Locusta migratoria migratoria* mitochondrion, complete genome; NCBI GenBank; NC_011114.1
- Xiao, L. L., Zhou, Z. J., & Huang, Y. (2008) Data from: *Locusta migratoria migratoria* mitochondrion, complete genome; NCBI GenBank; NC_011119.1
- Xu, M., & Shaw, K. L. (2019a). Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proceedings of the Royal Society B: Biological Sciences*, 286(1913), 20191607. <https://doi.org/10.1098/rspb.2019.1607>
- Xu, M., & Shaw, K. L. (2019b). The genetics of mating song evolution underlying rapid speciation: Linking quantitative variation to candidate genes for behavioral isolation. *Genetics*, 211(3), 1089–1104. <https://doi.org/10.1534/genetics.118.301706>
- Yang, Y., Servadio, M. R., & Richards-Zawacki, C. L. (2019). Imprinting sets the stage for speciation. *Nature*, 574(7776), 99–102. <https://doi.org/10.1038/s41586-019-1599-z>
- Zhang, H. L., Zhao, L., Zheng, Z. M., & Huang, Y. (2013). Data from: *Gomphocerus sibiricus* mitochondrion, complete genome; NCBI GenBank; NC_021103.1
- Zhao, L., Zheng, Z. M., Huang, Y., & Sun, H. M. (2010). Data from: *Euchorthippus fusigeniculatus* mitochondrion, complete genome; NCBI GenBank; NC_014449.1

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Nolen ZJ, Yildirim B, Irisarri I, et al. Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus* grasshoppers. *Mol Ecol*. 2020;29:4985–5002. <https://doi.org/10.1111/mec.15695>