Molecular Phylogenetics and Evolution 113 (2017) 67–75

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Cryptic speciation and gene flow in a migratory songbird Species Complex: Insights from the Red-Eyed Vireo (*Vireo olivaceus*)

C.J. Battey [a,*], John Klicka [b]

[a] University of Washington, Department of Biology, Kincaid Hall, Box 351800, Seattle, WA 98195, USA
[b] Burke Museum of Natural History and Culture, University of Washington, Box 353010, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 25 January 2017
Revised 7 April 2017
Accepted 8 May 2017
Available online 12 May 2017

Keywords:
Speciation
Hybridization
Phylogeography
Systematics
Birds
Migration
Genomics
Phylogenetics

ABSTRACT

Migratory species that alternate between sympatry and allopatry over the course of an annual cycle are promising subjects for studies seeking to understand the process of speciation in the absence of strict geographic isolation. Here we sought to identify cryptic species and assess rates of gene flow in a clade of neotropical migrant songbirds in which geography and taxonomy are currently out of sync: the Red-Eyed Vireo (*V. olivaceus*) Species Complex. Phylogenetic, clustering, and statistical species delimitation analyses found that *V. olivaceus* includes two non-sister lineages migrating in opposite directions across the equator. Analyses of gene flow identified low levels of introgression between two species pairs, but none between northern and southern *olivaceus*. We also identified substantial well-supported conflicts between nuclear and mitochondrial topologies. Although the geographic distribution of mito-nuclear discordance is suggestive of hybridization and mitochondrial capture, we found no evidence of introgression in the nuclear genome of populations with discordant mitochondrial gene trees. Our study finds that species boundaries match breeding range and migratory phenology rather than the existing taxonomy in this group, and demonstrates the utility of genomic data in inferring species boundaries in recently diverged clades.

1. Introduction

Geographic isolation is thought to be the dominant factor in initiating speciation in birds (Mayr, 1942; Mayr and O'Hara, 1986; Coyne and Price, 2000; Phillimore et al., 2008), and comparative studies often focus on the role of range gaps and geographic barriers in generating species diversity (Smith et al., 2014; Barber and Klicka, 2010; Zink et al., 2001). For migratory species, the role of allopatry in lineage divergence is unusual. Most individuals are physiologically capable of moving between disjunct ranges even when separated by large distances, but populations are isolated by heritable variation in the direction and timing of migration (Helbig, 1991; Pulido et al., 2001; Delmore and Irwin, 2014). In some cases sister species are allopatric during the breeding season and sym- or parapatric during the nonbreeding season, as occurs in many neotropical migrants breeding in northern temperate regions and wintering in Central and South America. Rates of gene flow across the ranges of these species are dependent on levels of migratory connectivity and distance of natal dispersal, rather than simple inertia in natal location.

The extensive spatial mixing associated with seasonal migration suggests that migratory species are promising subjects for studies seeking to better understand the role of gene flow in speciation. Recent studies in songbirds such as Redpolls (*Acanthis*; Mason and Taylor, 2015), Greenish Warblers (*Phylloscopus trochiloides*; Irwin et al., 2005), and Darwin's Finches (*Geospiza*; Lamichhaney et al., 2015) have found evidence of extensive introgression among morphologically divergent populations. In some cases (e.g. *Geospiza*, *Acanthis*) phenotypic plasticity and divergent selection on a small number of genes is thought to drive morphological variation despite a largely panmictic nuclear genome. One recent whole-genome sequencing study of hybridizing Golden- and Blue-winged warblers (*Vermivora chrysoptera* & *V. cyanoptera*, respectively), found that just 6 genomic regions appear to control phenotypic variation between species despite extensive introgression across the genome (Toews et al., 2016). These studies are part of an increasing focus on the extent and impacts of hybridization on speciation (and, in corollary, the biological meaning of taxonomic designations).

In this study we assess the consequences of divergent migratory behavior on speciation and gene flow in a systematically recalcitrant clade of neotropical songbirds: the Red-Eyed Vireo (*V. olivaceus*) Species Complex. This group includes three widespread migratory

* Corresponding author.
E-mail address: cjbattey@uw.edu (C.J. Battey).

http://dx.doi.org/10.1016/j.ympev.2017.05.006
1055-7903/

species that are sympatric in northern South America during the nonbreeding season and para- or allopatric during the breeding season. Unusually, the most widespread species in the group – *V. olivaceus* – includes populations migrating both north and south of the equator, as well as sedentary populations breeding in northern South America. Due to similar habits, vocalizations, and morphologies, earlier taxonomists typically recognized three (Mayr and Short, 1970) or four (Paynter, 1968; American Ornithologists' Union, 1983) species in the group. The first genetic study of the group (Johnson and Zink, 1985) found the North- and South American *olivaceus* monophyletic relative to the Central American *flavoviridis*, resulting in the current five-species taxonomy (Johnson and Zink, 1985; American Ornithologists' Union, 1998). Surprisingly, a recent study of mitochondrial DNA suggested the presence of at least six geographically and genetically distinct lineages (Slager et al., 2014), though basal nodes in the group were poorly supported.

Here we sequenced a genome-wide sampling of SNP's from museum specimens caught across the western hemisphere in order to assess the strength and distribution of species boundaries across the group. Specifically, we ask: (1) What is the optimal species delimitation scheme for partitioning diversity in the Red-Eyed Vireo Complex? (2) Do migrant species breeding in parapatry show evidence of gene flow after lineage divergence? And (3) Does introgression lead to erosion of neutral genetic differentiation among putative species?

2. Methods

2.1. Sampling and library preparation

We obtained frozen tissue samples from vouchered natural history museum specimens representing 40 individuals and 6 species of Vireo. These included four members of the Red-eyed vireo complex (*V. olivaceus, V. flavoviridis, V. altiloquus, V. magister*) and two outgroup taxa (*V. gilvus, V. plumbeus*) (Fig. 1, Supplementary Table 1). No tissue samples of *V. gracilirostris* (endemic to Noronha Island off the coast of Brazil) could be located, so this taxon was not included in our study. Southern and northern *V. olivaceus* were distinguished by subspecific identity in museum records, or by location and time of capture for any specimens lacking subspecific identification.

Whole genomic DNA was extracted using a Qiagen DNEasy spin-column extraction kit, following manufacturer instructions. Sequencing libraries were prepared following the ddRADseq protocol (Peterson et al., 2012). Briefly, we digested 300–500 ng of DNA per sample with NEB high-fidelity sbf1 and msp1 restriction enzymes. Barcoded adapters were ligated onto the resulting fragments and those in the 415–515 bp range were selected with a Blue Pippin Prep. Illumina flowcell annealing primers, PCR primers, and multiplexing indices were ligated to size-selected fragments, and the products amplified over 11 PCR cycles. We used a 1.5× ratio of AmPure XP beads to remove small DNA fragments between all steps. Amplification and size distribution were checked prior to sequencing with a qBit 2.0 fluorometer and a BioAnalyzer run. Samples were sequenced for 50 bp single-end reads on shared lanes (with other Passerine birds) across two runs of an Illumina Hiseq 2000.

2.2. Sequence assembly

Raw sequencing reads were processed with pyRAD v. 2.17 (Eaton, 2014). PyRAD demultiplexes and quality-filters reads, then uses the Usearch algorithm (Edgar, 2010) to cluster reads into loci within samples and loci into stacks between samples. Stacks of putatively orthologous loci are then aligned with Muscle (Edgar, 2004) and output in a variety of sequence alignment formats for downstream analysis. We set a minimum read depth of 6 for calling consensus sequences within samples and used a clustering threshold of 0.90 for all Usearch runs. We filtered out all loci sharing heterozygotic sites across more than 8 individuals to avoid including paralogs in our sequencing alignment.

Because sample coverage – the proportion of sampled individuals sequenced for any given locus – is known to have a large effect on the total number of loci recovered (Leaché et al., 2014; Huang and Knowles, 2014; DaCosta and Sorenson, 2016), we prepared two sequence assemblies for downstream analyses by varying this parameter in pyRAD. The "MD10" alignment includes all loci sequenced for at least 90% of samples, while the "MD50" alignment includes loci sequenced for at least 50% of samples. To explore the source of missing data in our alignments we also used a Mantel Test (Mantel, 1967) to test for a correlation between pairwise genetic distance and the number of shared loci among samples.

Sequence alignments for concatenated phylogenetic analyses and structure input files were prepared with PyRAD's standard output options. STRUCTURE and Adegenet analyses used the MD50 alignment. SNAPP and RAxML were run on both MD50 and MD10 alignments, though only MD10 alignments were used for Bayes Factor Delimitation (BFD*). Both STRUCTURE and SNAPP input files include one random SNP per locus. For SNAPP, pyRAD's "unlinked_snps" output was converted to a three-state numeric input file (two alleles plus heterozygotes) suitable for analysis in SNAPP using custom R scripts written by Dr. Barbara Banbury and available online on the shinyphry webserver (https://rstudio.stat.washington.edu/shiny/phrynomics/). We developed additional custom R scripts for visualizing patterns of missing data, writing D-test input files, summarizing D-test output, and plotting STRUCTURE results with consistent color matching (github.com/cjbattey/radplots/). All analyses were conducted in the R programming language (R Development Core Team, 2014).

2.3. Phylogenetics

We ran two phylogenetic analyses on both the MD50 and MD10 alignments. First, we inferred a phylogenetic tree of concatenated full sequences (i.e. full RAD loci stripped of barcode, adapter, and restriction overhang sequences) in a maximum likelihood framework with the program RAxML v8 (Stamatakis, 2014). We used the GTR + Gamma model of sequence evolution and ran 1000 bootstrap iterations to assess support.

Next, we inferred a species tree in the program SNAPP v. 1.3 (Bryant et al., 2012), an add-on to the BEAST 2.2.0 package (Bouckaert et al., 2014). SNAPP requires the user to input the number of putative taxa and assign individual samples to these taxa. We initially assigned individuals to taxa based on nodes with at least 70% bootstrap support in the RAxML analysis (5 populations), and later re-ran the SNAPP analysis for the best-fitting BFD* model (6 populations, see Results). Because SNAPP models gene/species tree conflicts as the result of incomplete lineage sorting (ILS) rather than horizontal gene flow, we eliminated any individuals with more than 5% admixture in the best-performing STRUCTURE analyses from our SNAPP runs, resulting in the exclusion of two samples of *V. olivaceus* from Trinidad and Tobago. We used default values for forward and back mutation rates (3.33 and 5.08, respectively), and sampled the coalescence rate from a uniform infinite distribution. SNAPP analyses were run for one million generations, sampling every 10,000. Convergence was assessed through estimated ESS values and trace plots in Tracer v1.6 (Rambaut and Drummond, 2003).

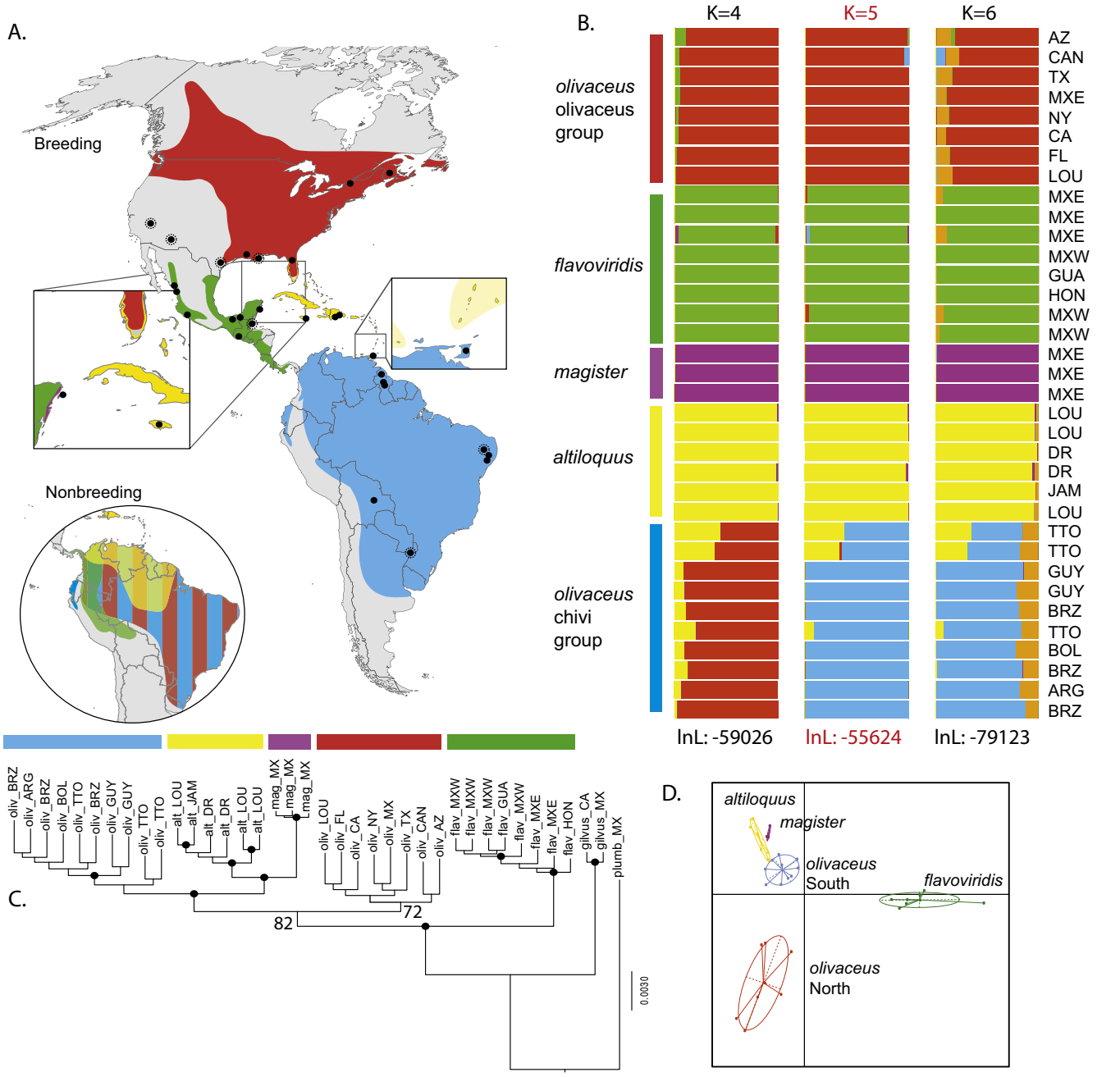

Fig. 1. A: range map for breeding (top) and nonbreeding (bottom, circled) seasons, adapted from Birdlife International & Natureserve, 2013. Circled points represent specimens captured during the migration season. Magnified cutouts are provided for two regions of parapatry. B: STRUCTURE output for highest marginal-likelihood runs at $k = 4$–6. C: RAxML MD50 concatenated phylogeny. Numbers at nodes indicate bootstrap support, and circles represent bootstrap support > 90. D: PCA sample and population coordinates on the first two PC axes, using population assignments from k-means clustering.

2.4. Clustering and ordination

We used two unsupervised clustering algorithms to assign individuals to genetic clusters: STRUCTURE (Pritchard et al., 2000; Falush et al., 2003) and k-means clustering in the R package Adegenet (Jombart, 2008). For STRUCTURE analyses we adjusted the prior mean on F (parameter "FPRIORMEAN" in mainparams) to 0.1 (from the default value of 0.01) in order to reflect the higher expected divergence among named species. Analyses were run using the admixture model, correlated allele frequencies, and with a starting state based on population assignments from museum specimen records. We conducted 10 runs for each value of k between 2 and 8 and used the StructureHarvester web server to identify the value of k associated with the largest second-order change in marginal likelihood (Evanno et al., 2005; Earl, 2012).

For Adegenet analyses we first conducted a Principal Components Analysis (PCA; Pearson, 1901) on allele counts from the STRUCTURE input file, and then used k-means clustering to estimate the number of clusters and assign individuals to genotype clusters. We kept all PC axes for k-means clustering and inspected both population assignments and change in Bayesian Information Criterion (BIC) scores across values of k to select an optimal partitioning scheme for the group. PCA results were plotted using the first two PC axes and population assignments from k-means clustering (Fig. 1).

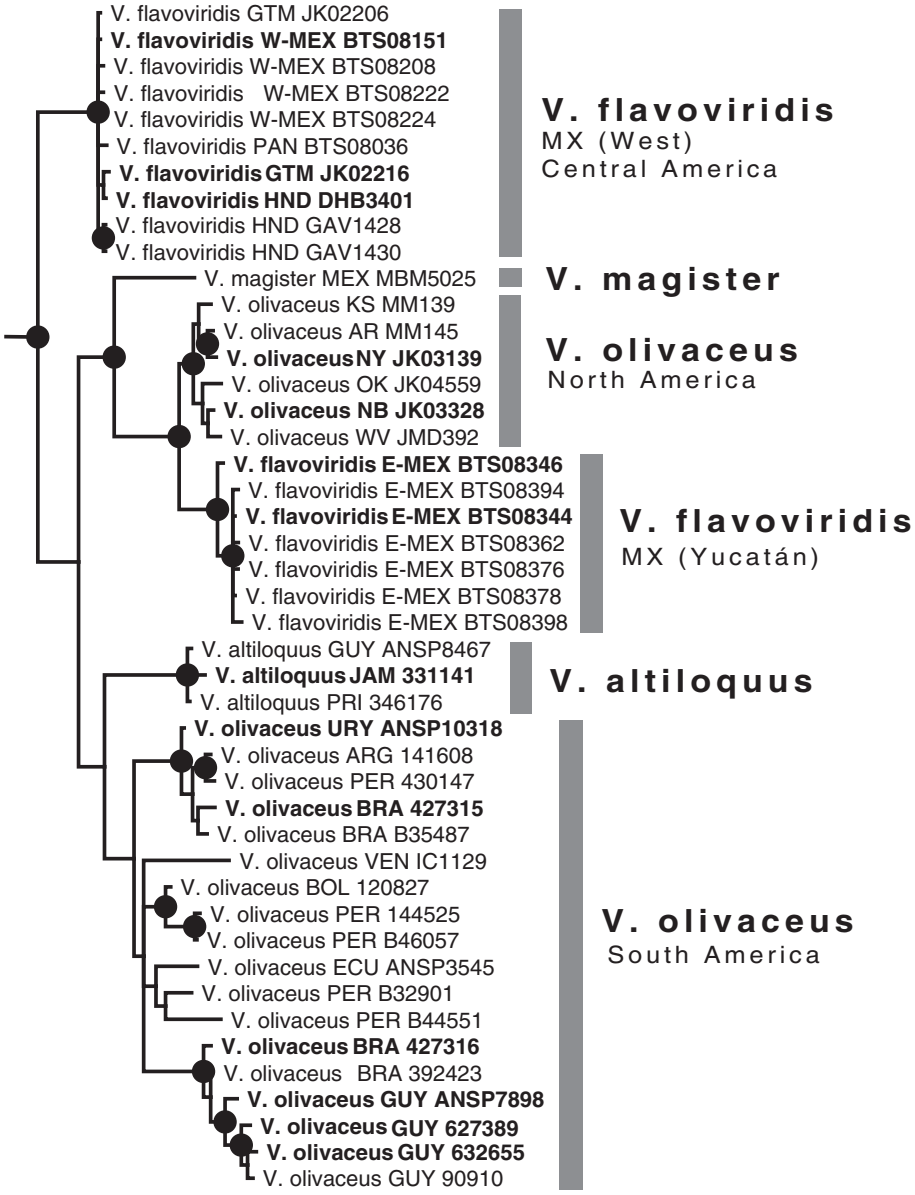

Fig. 2. ND2 gene tree from Slager et al. (2014). Circles represent supported nodes. Bolded samples were also sequenced for nuclear SNP's. Bayesian posterior and ML bootstrap values are given at unsupported nodes.

2.5. Admixture analysis

To test for introgression after the point of lineage divergence we calculated the D statistic ("ABBA/BABA"; Green et al., 2010) for all species pairs. D tests measure the imbalance in loci that conflict with the species tree, under the assumption that two types of discordant gene trees should occur with equal frequency if ILS is the sole source of discordance. An imbalance in the distribution of discordant alleles is taken as evidence of introgression, and significance assessed by a Z score describing departure from a null mean derived from bootstrap replicates of the original data.

As in recent studies of plants (Eaton and Ree, 2013; Eaton et al., 2015) and birds (Lamichhaney et al., 2015), we define a "test" in this context as asking "do populations X and Y show significant signs of introgression?" We calculated D for all unique combinations of individuals on a pectinate four-tip tree consistent with the SNAPP species tree for each test and applied a Holm-Bonferroni correction (Holm, 1979) to each *p* value to correct for multiple tests. Any combination of individuals returning less than 100 total discordant loci (including heterozygous sites) was removed from the test. We considered a population pair to show signs of introgression if any combination of individuals within the test gave a corrected two-tailed p value of less than 0.05.

2.6. Species delimitation with Bayes factors

We used Bayes Factor Delimitation in SNAPP (BFD*; Leaché et al., 2014; Grummer et al., 2014) to rank species delimitation models in a multispecies coalescent framework. Briefly, BFD* consists of running SNAPP analyses on models with different numbers of species and assignments of individuals to species, estimating the marginal likelihood of each model, and ranking model fit among runs by comparing Bayes factors. In theory, BFD* provides a quantitative method for applying a flexible version of the genealogical species concept (Baum and Shaw, 1995) across multiple unlinked loci, in which the operational criteria for delimiting "species" is fit to a coalescent model with no gene flow.

For BFD* analyses we ran SNAPP on five species delimitation models under default parameters at a chain length of 50,000 (discarding the first 5000 as burn-in) for 48 steps. In this context "steps" refers to the different levels of power-posterior used in path sampling, and "chain length" refers to the length of the MCMC used to optimize parameters at each step. As in our species tree inference approach, we excluded samples with greater than 5% admixture in our best fitting STRUCTURE model from BFD* analyses, because SNAPP does not account for horizontal gene flow. BFD* analyses were run on the MD10 dataset only.

3. Results

3.1. RAD loci and missing data

After demultiplexing, trimming adapter and barcode sequences, and quality filtering in pyRAD, we recovered an average of 1.4e6 39-bp reads per individual. Usearch clustering at similarity 0.90 within individuals returned an average of 13,323 loci per individual. Two samples sequenced from relatively degraded DNA returned high levels of missing data (over two standard deviations above mean missing data percentage in the initial MD50 dataset; Supplementary Fig. 1) and were removed from the analysis, resulting in a final dataset of 38 individuals.

Sample coverage had a large impact on the total number of loci included in the alignment after sequence assembly. Our final "MD10" dataset consists of 938 loci, and the "MD50" dataset of 7799 loci. Total concatenated matrix lengths for MD10 and MD50 alignments are 36,582 and 304,161 base pairs, respectively. Mantel tests found that pairwise missing data were significantly correlated with genetic distance when all samples were analyzed ($p = 0.012$) but not when outgroups were removed ($p = 0.57$; Supplementary Fig. 2). This agrees with previous studies that failed to find significant correlations between missing loci and genetic distance at shallow (<10 my) timescales (Eaton and Ree, 2013; Eaton et al., 2015), suggesting that allelic dropout via mutation in restriction sites is not the primary source of missing data in ddRAD studies sampling at the population level in recently diverged groups.

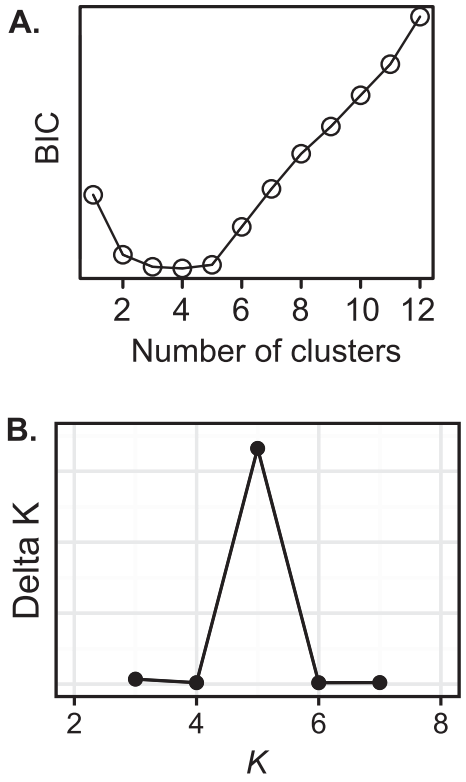

Fig. 3. Summary of clustering results. A: BIC scores from *k*-means clustering in Adegenet. B: Evanno method output from STRUCTURE HARVESTER.

3.2. Phylogenetics

Concatenation and species-tree analyses produced very similar topologies (Figs. 1 and 4) for both MD50 and MD10 alignments, though MD10 analyses had lower support values (Supplementary Figs. 3 and 4). The species tree is fully pectinate, with the Central American species *flavoviridis* representing the earliest-diverging extant lineage. Northern and southern *olivaceus* are paraphyletic, with South American breeders more closely related to the Caribbean taxa *altiloquus* and *magister* than their northern-hemisphere conspecifics.

SNAPP fully supported all nodes in the species tree using the MD50 alignment on both 5- and 6-population models (Fig. 4, Supplementary Fig. 5). For the MD50 alignment, RAxML bootstraps strongly supported the monophyly of a clade including southern *olivaceus*, *altiloquus*, and *magister*, but provided only moderate support for the monophyly of northern *olivaceus* and the sister relationship of *flavoviridis* to other taxa. In particular, some bootstrap replicates returned topologies in which northern *olivaceus* forms a paraphyletic grade sister to all ingroup taxa other than *flavoviridis*. This topology was also observed using the MD10 alignment in RAxML (Supplementary Fig. 4).

The nuclear topology conflicted with that found in an earlier study of mitochondrial DNA in the clade (Slager et al., 2014; Fig. 2); with the mtDNA topology finding *flavoviridis* polyphyletic in addition to *olivaceus*, and placing *magister* as sister to a combined clade of northern *olivaceus* and eastern *flavoviridis*.

3.3. Clustering

STRUCTURE analyses strongly favored a five-population model that split northern and southern *olivaceus* and assigned individuals

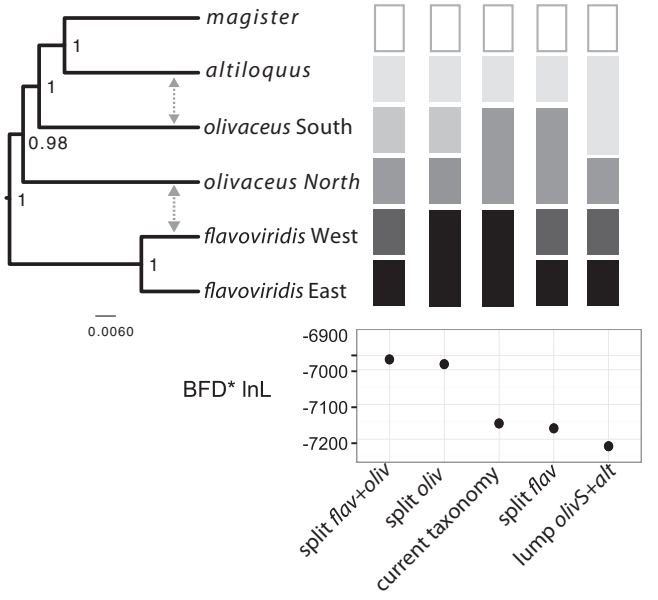

Fig. 4. SNAPP species tree (MD50 alignment) and BFD* summary model summaries. Shaded bars depict species delimitation models tested in BFD*, with path-sampling marginal likelihoods provided below. Grey arrows indicate introgression events inferred by D tests.

to the population cluster suggested by museum specimen records and concatenated phylogenetic analyses (Fig. 3; Table 1). The two samples of southern *olivaceus* placed outside the supported monophyletic clade in RAxML analyses were identified as hybrids between *olivaceus* and *altiloquus* in STRUCTURE output, a result that matches their geography: Trinidad and Tobago is the nearest point of contact between the ranges of *olivaceus* and *altiloquus* during the breeding season. Aside from these samples, STRUCTURE found very low levels of admixture between species in the complex at $k = 5$, with no other individuals showing more than 5% admixture from a different putative species.

K-means clustering in Adegenet found that models with 3–5 clusters minimized BIC scores (Fig. 3; Table 1). In all *k*-means analyses, five was the highest value of *k* that matched results from existing taxonomy or mitochondrial DNA, with higher values placing individual *olivaceus* specimens in their own clusters. The first two PC axes cumulatively represent 13.2% of variance in the data. Distances among clusters using the first two PC axes were concordant with phylogenetic trees (Fig. 1). Interestingly, both clustering methods showed a tendency to lump northern and southern *olivaceus* when run at $k = 4$, rather than lumping any monophyletic clade identified in phylogenetic analyses. These results occurred in roughly half of STRUCTURE models run at $k = 4$ (others lumped *altiloquus* and southern *olivaceus*, or *altiloquus and magister*) and were consistently produced by *k*-means clustering, though in STRUCTURE all $k = 4$ models returned lower marginal likelihoods than five-population models.

3.4. D-Tests

D statistics identified significant introgression between southern *olivaceus* and *altiloquus* in all tests that included one of the samples from Trinidad and Tobago. We also identified significant introgression between *olivaceus* and western *flavoviridis* in four combinations of individuals (Fig. 4; Table 2). In both cases, gene flow appears to be unidirectional (into *olivaceus*), as reciprocal tests failed to find significant introgression (though limited sampling in some areas may limit resolution of gene flow restricted to narrow hybrid zones). In the case of *flavoviridis*, all significant combinations involved west-Mexican individuals – conflicting with the mitochondrial topology that places east-Mexican populations in a monophyletic clade with northern *olivaceus* and *magister* (Fig. 2). No tests with *magister* as the putative receptor lineage had over 100 discordant loci, reflecting the low genetic variability observed between samples of this micro-endemic species. Notably, for *olivaceus*, no combinations of individuals supported significant introgression between northern and southern populations.

3.5. Coalescent species delimitation

Bayes factor delimitation selected a model of 6 species, reflecting the major clades observed in the ND2 gene tree (Fig. 4; Table 3). This species delimitation scheme splits *olivaceus* between North and South American breeding populations, and *flavoviridis* between eastern Mexico and west Mexico + Central America. The second ranked model split *olivaceus* and retained a monphyletic *flavoviridis*, reflecting the clades in our concatenated RADseq tree. All models that did not split *olivaceus* were at least an order of magnitude worse performing than those that did (in terms of Bayes Factors), and splitting only *olivaceus* resulted in a much larger increase in marginal likelihood than splitting only *flavoviridis*. The mean ESS across all steps and all models was 326.7 (standard deviation = 211.0), suggesting that our chain length was sufficient to reach convergence.

4. Discussion

4.1. Migration, range, and gene flow

Dispersal ability and levels of genetic divergence across biogeographic barriers have generally been found to be inversely corre-

Table 1
Clustering analyses summary.

	Structure		Adegenet
K	Mean LnP(*K*)	Delta *K*	BIC
3	−63155.08	30.75	123.63
4	−61946.92	1.22	123.56
5	−55761.99	627.23	123.72
6	−94794.66	6.32	125.42
7	−72233.84	2.03	127.12

Table 3
BFD* results summary.

Model	lnL	2lnBF	Rank
Split *flavoviridis* + *olivaceus*	−6969.89	NA	1
Split *olivaceus*	−6983.59	27.41	2
Current taxonomy	−7168.87	397.96	4
Split *flavoviridis*	−7154.83	369.88	3
Lump *olivaceus* + *altiloquus*	−7220.66	501.54	5

Table 2
D-Test results summary. Number of loci, ABBA and BABA are mean values for all combinations in a given test. Tests are given as the species assigned to P1 + P2 and P3 tips.

Test	Number of loci	ABBA	BABA	Z range	Significant combinations
alt flav	3387	51	50	(0.03,1.69)	0/8
alt mag	2795	54	52	(0,1.11)	0/16
alt olivN	3148	48	43	(0.07,2.65)	0/30
alt olivS	3015	49	46	(0,2.89)	0/37
flav alt	3634	43	48	(0.24,1.83)	0/7
flav mag	3472	42	40	(0.09,0.68)	0/2
flav olivN	3515	39	44	(0.23,2.55)	0/14
flav olivS	3594	39	47	(0.06,1.86)	0/11
olivN alt	2357	63	65	(0.02,3.02)	0/126
olivN flav	**2211**	**59**	**63**	**(0.03,3.9)**	**4/153**
olivN mag	2218	63	65	(0,2.14)	0/62
olivN olivS	2203	59	59	(0,2.55)	0/205
olivS alt	**2318**	**53**	**65**	**(0.02,5.71)**	**23/172**
olivS flav	2476	57	58	(0.06,2.8)	0/99
olivS mag	2281	59	63	(0.03,2.46)	0/75
olivS olivN	2539	53	57	(0,2.88)	0/148

Bolded rows indicate tests returning significant signal of introgression.

lated in birds, with higher dispersing species experiencing more gene flow and lower levels of divergence among putatively allopatric populations (Smith et al., 2013). Long-distance migration represents the high end of the spectrum of dispersal abilities, suggesting that populations of migrant species are likely to experience high rates of gene flow, even when ranges are separated by substantial biogeographic barriers. Even in the Eurasian Blackcap (*Sylvia atricapilla*), a species for which migratory orientation is known to be heritable (Helbig, 1991), studies of up to 17 microsatellite loci have failed to find significant genetic structure among populations that vary in migratory behavior in the wild (Mettler et al., 2013; Linossier et al., 2016); suggesting that gene flow associated with incomplete philopatry may prevent divergence at selectively neutral loci. Even among well-recognized species of songbird, interspecific hybridization is relatively commonly reported (McCarthy, 2006). However, determining whether these events contribute significantly to the genetic makeup of extant species requires sampling extensively across the genome, which has historically limited the resolution of studies employing Sanger-sequencing approaches.

Using genome-wide SNP's we observed significant introgression in two species pairs in the Red-Eyed Vireo Complex, but found that hybridization contributes little to the overall genomic makeup of most species. Intriguingly, we only observed introgression between species pairs for which the breeding ranges are separated by substantial range gaps. Among species breeding in sym- or parapatry (*olivaceus* and *altiloquus* in Florida; *olivaceus* and *flavoviridis* in the Darien, *magister* and *flavoviridis* on the Yucatán) we observed no introgressed individuals. In the cases where we did observe significant introgression it either occurred at very low levels suggestive of historic rather than contemporary gene flow (e.g. *flavoviridis* and northern *olivaceus*), or was limited to areas near the range boundaries of both species (*altiloquus* and southern *olivaceus*). Unlike the extensive homogenizing gene flow seen in *Acanthis* (Mason and Taylor, 2015), introgression in the Red-Eyed Vireo complex appears to contribute relatively little to the genomic makeup of these species despite occurring at detectable levels.

In addition to this broad view of gene flow in the group, our results provide the basis for several novel inferences of natural history and biogeography. First, putative *olivaceus* populations on the island of Trinidad appear to be hybrids between *olivaceus* and *altiloquus*. Although our sample size is quite limited, we observed significant introgression in all three TTO samples analyzed. The smallest gap between the breeding ranges of *olivaceus* and *altiloquus* in the southern Caribbean is the approximately 80 miles of open ocean between the islands of Tobago and Grenada, and both island populations are reported to be sedentary (Cimprich et al., 2000; Chace et al., 2002). Contemporary hybridization would thus require either long-distance overwater dispersal by putatively sedentary birds, or double-breeding by overwintering *altiloquus* from the northern Caribbean. Although more extensive sampling is required to confidently infer the presence of a hybrid zone in the region, our observation should prompt further investigations of the nature of dispersal and migration in island populations of both species.

Second, the Yucatán vireo appears to represent an instance of speciation via loss of migration, as its ancestors are predominantly migratory and its range is adjacent to a known stopover location for its sister species (*altiloquus*) during fall migration (Birdlife International and NatureServe, 2013). Migratory drop-off has been proposed to form a mechanism of speciation in *Sylvia* warblers (Voelker and Light, 2011) and *Catharus* thrushes (Outlaw et al., 2007). The recent finding that migration evolved primarily via temperate-breeding taxa extending their ranges to the south (Winger et al., 2014), coupled with the high species diversity of the tropics, suggests that loss of migration in tropical lineages is likely a common biogeographic pattern in neotropical birds.

4.2. Species delimitation

We found that the current five-species taxonomy (AOU, 1998) under represents diversity in the Red-Eyed Vireo Complex. Both *olivaceus* and *flavoviridis* include genetically divergent lineages breeding in disjunct ranges (North and South America and Eastern and Western Mexico, respectively). In the case of *olivaceus*, genetics and life history both support splitting the species. Northern and southern hemisphere breeding populations are non-monophyletic, do not exchange genes at a level detectable in our data, and are characterized by variation in a heritable life-history trait that confers effective reproductive isolation between populations (e.g. direction and timing of migration; Helbig, 1991; Pulido et al., 2001). The apparent migratory divide in the species is an artifact of faulty taxonomy, not recent divergence, novel mechanisms of gene flow through resident populations, or the rapid evolution of migratory divides as documented in Barn Swallows (*Hirundo rustica*; Garcia-Perez et al., 2013) and Eurasian Black-Caps (Bearhop et al., 2005; Rolshausen et al., 2009).

V. flavoviridis is also structured between breeding populations in eastern and western Mexico and splitting the species was supported by BFD*, but other analyses suggested that these populations are at an earlier stage in the process of speciation than *olivaceus*. Although mitochondrial haplotypes are polyphyletic and entirely sorted between eastern and western Mexico (demonstrating that females migrating to breeding areas in the Sierra Madre Occidental and Sierra Madre Oriental have high fidelity to their breeding region), the species is monophyletic in nuclear SNP's. Clustering algorithms failed to distinguish eastern and western populations when run on the full dataset, and divergence between these populations in the species tree was much more recent than among any other putative species pair.

V. flavoviridis thus appears to be firmly within the "species/subspecies conundrum" (Huang and Knowles, 2015), in which different species delimitation criteria are likely to disagree on the taxonomic rank granted to hierarchically structured allo- or parapatric populations. In the absence of other data demonstrating reproductive isolation or intrinsic ecological or life-history variation among populations of *flavoviridis* (as seen in *olivaceus*), we take a conservative approach and propose elevating all populations of *olivaceus* breeding in South America to species status under the original name *Vireo chivi* (Vieillot, 1817), while retaining *V. flavoviridis* and its existing eastern and western subspecies (*V. f. flavoviridis* and *V. f. forreri*, respectively).

4.3. Mito-nuclear discordance

Topologies inferred from nuclear SNP's conflicted with mitochondrial DNA in three species that co-occur seasonally in eastern Mexico. Northern *olivaceus*, eastern *flavoviridis*, and *magister* form a monophyletic clade in ND2 sequences but are paraphyletic in nuclear SNP's. Two scenarios may explain this discordance: incomplete lineage sorting (ILS) of ancestral ND2 in the common ancestor of the complex, followed by fixation of a derived allele in eastern *flavoviridis* and *magister*; or introgression and mitochondrial capture among the parapatric species in eastern Mexico. Evidence distinguishing these scenarios is equivocal. In favor of hybridization, the discordant populations have (nearly) parapatric breeding ranges and overlap extensively in migration; while ILS is not expected to lead to any coherent geographic patterns. However, we found no evidence of nuclear introgression among northern *olivaceus,* eastern *flavoviridis*, or *magister*, as would be expected if hybridization was the cause of mtDNA discordance.

Although we cannot exclude either potential cause, we propose that the geographic and topological extent of discordance is more suggestive of past hybridization than ILS. At least one study has

identified a case of complete mitochondrial capture despite statistically insignificant levels of nuclear introgression (Good et al., 2015). We hypothesize that hybridization may have initially transferred northern *olivaceus* mitochondrial haplotypes to *flavoviridis* and *altiloquus*, which subsequently became fixed in these populations through some combination of drift (aided by the reduced effective population size of the mitochondrial genome) and selection. Most existing evidence of interspecific hybridization and introgression has been drawn from studies analyzing a small number of loci, precluding accurate measurement of nuclear introgression (Toews and Brelsford, 2012). Our analysis supports the general findings of Good et al. (2015) that nuclear and mitochondrial introgression can be entirely decoupled, and suggests that this system is a promising target for future investigations of the evolutionary dynamics of mitochondrial introgression.

5. Conclusions

We identified a cryptic species of songbird distinguished from its congeners by divergence in migratory direction and timing, and identified a set of unexpected hybrid individuals from putatively sedentary allopatric populations in the southern Caribbean. Our study demonstrates that post-divergence gene flow has relatively little impact on the overall genetic makeup of species in the Red-Eyed Vireo complex, despite occurring at detectable levels in two species pairs. Previous results from studies of mitochondrial DNA received partial support in our data, but did not reflect the genome-wide species tree of the group. Our study demonstrates the utility of genomic data for identifying cryptic species and generating novel natural history observations in groups with little phenotypic diversity.

Data accessibility

Raw genomic data, sequence assembly parameters, and scripts for running analyses are available at: http://dx.doi.org/10.5061/dryad.9b6p8.

Author contributions

CJB and JK conceived the study and designed the sampling scheme. CJB conducted labwork, ran analyses, and wrote the manuscript.

Acknowledgements

The authors thank Dr. Robert Bryson and Kevin Epperly for assistance with labwork and analyses. This manuscript was greatly improved by suggestions from two anonymous reviewers, and by comments from David Slager, Dr. Adam Leaché, Cooper French, and Ethan Linck. Dave Slager and Dr. Leaché also provided valuable assistance with implementing species delimitation analyses. Specimens analyzed in this paper represent the combined efforts of generations of collectors and curators without whom this work would be impossible. We thank the following institutions for providing specimen tissue samples: The Louisiana State Museum of Natural Science, American Museum of Natural History, Field Museum of Natural History, National Museum of Natural History, and the Academy of Natural Sciences at Drexel University.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.05.006.

References

American Ornithologists' Union, 1983. Check-List of North American Birds. Am. Ornithol. Union, Washington, D.C..
American Ornithologists' Union, 1998. Check-List of North American Birds. Am. Ornithol. Union, Washington, D.C..
Barber, B.R., Klicka, J., 2010. Two pulses of diversification across the Isthmus of Tehuantepec in a Montane Mexican bird Fauna. Proc. R. Soc. Lond. B: Biol. Sci. 277 (1694), 2675–2681. http://dx.doi.org/10.1098/rspb.2010.0343.
Baum, D.A., Shaw, K.L., 1995. Genealogical perspectives on the species problem. In: Hoch, P.C., Stephenson, A.G. (Eds.), Experimental and molecular approaches to plant biosystematics, Monographs in systematic botany from the Missouri Botanical Garden, vol. 53, pp. 289–303.
Bearhop, Stuart, Fiedler, Wolfgang, Furness, Robert W., Votier, Stephen C., Waldron, Susan., Newton, Jason, Bowen, Gabriel J., Berthold, Peter, Farnsworth, Keith, 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. Science 310 (5747), 502–504.
BirdLife International and NatureServe, 2013. Bird species distribution maps of the world.
Bouckaert, Remco, Heled, Joseph, Kühnert, Denise, Vaughan, Tim, Chieh-Hsi, Wu, Xie, Dong, Suchard, Marc A., Rambaut, Andrew, Drummond, Alexei J, 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4), e1003537.
Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29 (8), 1917–1932.
Chace, Jameson F., Woodworth, Bethany L., Cruz, Alexander., 2002. Chace, Jameson F., Bethany L. Woodworth and Alexander Cruz. 2002. Black-Whiskered Vireo (Vireo Altiloquus), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/607. In: Poole, A. (Ed.), The Birds of North America Online. http://bna.birds.cornell.edu/bna/species/607.
Cimprich, David A, Moore, Frank R., Guilfoyle, Michael P., 2000. Red-Eyed Vireo (Vireo Olivaceus). In: The Birds of North America Online, Cornell Lab of Ornithology, Ithaca, NY. http://bna.birds.cornell.edu.offcampus.lib.washington.edu/bna/species/527.
Coyne, Jerry A., Price, Trevor D., 2000. Little evidence for sympatric speciation in island birds. Evolution 54 (6), 2166–2171. http://dx.doi.org/10.1111/j.0014-3820.2000.tb01260.x.
DaCosta, Jeffrey.M., Sorenson, Michael.D., 2016. DdRAD-Seq phylogenetics based on nucleotide, indel, and presence–absence polymorphisms: analyses of two Avian Genera with contrasting histories. Mol. Phylogenet. Evol. 94, 122–135.
Delmore, Kira E., Irwin, Darren E., 2014. Hybrid songbirds employ intermediate routes in a migratory divide. Ecol. Lett. 17 (10), 1211–1218. http://dx.doi.org/10.1111/ele.12326.
Earl, Dent.A., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4 (2), 359–361.
Eaton, D.A., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics, p.btu121.
Eaton, Deren AR, Ree, Richard H., 2013. Inferring phylogeny and introgression using RADseq Data: an example from flowering plants (Pedicularis: Orobanchaceae), Systematic Biology, syt032.
Eaton, Deren A.R., Hipp, Andrew L., González-Rodríguez, Antonio, Cavender-Bares, Jeannine, 2015. Historical introgression among the american live oaks and the comparative nature of tests for introgression. Evolution 69 (10), 2587–2601.
Edgar, Robert C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32 (5), 1792–1797.
Edgar, Robert C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26 (19), 2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461.
Evanno, Guillaume, Regnaut, Sebastien, Goudet, Jérôme, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14 (8), 2611–2620.
Falush, Daniel, Stephens, Matthew, Pritchard, Jonathan K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164 (4), 1567–1587.
Garcia-Perez, Belen, Hobson, Keith A., Powell, Rebecca L., Still, Christopher J., Huber, Gernot H., 2013. Switching hemispheres: a new migration strategy for the disjunct Argentinean breeding population of Barn Swallow (*Hirundo rustica*). PLoS ONE 8 (1), e55654. http://dx.doi.org/10.1371/journal.pone.0055654.
Good, Jeffrey M., Vanderpool, Dan, Keeble, Sara, Bi, Ke, 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of Chipmunks. Evolution 69 (8), 1961–1972.
Green, Richard E., Krause, Johannes, Briggs, Adrian W., Maricic, Tomislav, Stenzel, Udo, Kircher, Martin, Patterson, Nick, et al., 2010. A draft sequence of the

Neandertal genome. Science 328 (5979), 710–722. http://dx.doi.org/10.1126/science.1188021.
Grummer, Jared A., Bryson Jr., Robert W., Reeder, Tod W., 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst. Biol. 63 (2), 119–133. http://dx.doi.org/10.1093/sysbio/syt069.
Helbig, Andreas J., 1991. Inheritance of migratory direction in a bird species: a cross-breeding experiment with SE- and SW-migrating Blackcaps (Sylvia Atricapilla). Behav. Ecol. Sociobiol. 28 (1), 9–12. http://dx.doi.org/10.1007/BF00172133.
Holm, Sture, 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6 (2), 65–70.
Huang, Huateng, Knowles, L. Lacey., 2014. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences, Systematic Biology, July. http://sysbio.oxfordjournals.org/content/early/2014/07/27/sysbio.syu046.abstract.
Huang, Jen-Pan, Knowles, L. Lacey., 2015. The Species versus Subspecies Conundrum: Quantitative Delimitation from Integrating Multiple Data Types within a Single Bayesian Approach in Hercules Beetles, Systematic Biology, December. http://sysbio.oxfordjournals.org/content/early/2015/12/16/sysbio.syv119.abstract.
Irwin, Darren E., Bensch, Staffan, Irwin, Jessica H., Price, Trevor D., 2005. Speciation by distance in a ring species. Science 307 (5708), 414–416. http://dx.doi.org/10.1126/science.1105201.
Johnson, Ned K., Zink, Robert M., 1985. Genetic Evidence for Relationships among the Red-Eyed, Yellow-Green, and Chivi Vireos. Wilson Bull. 97 (4), 421–435.
Jombart, Thibaut, 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24 (11), 1403–1405.
Lamichhaney, Sangeet, Berglund, Jonas, Almen, Markus Sallman, Maqbool, Khurram, Grabherr, Manfred, Martinez-Barrio, Alvaro, Promerova, Marta, et al., 2015. Evolution of Darwin/'s Finches and their beaks revealed by genome sequencing. Nature 518 (7539), 371–375.
Leaché, Adam D., Fujita, Matthew K., Minin, Vladimir N., Bouckaert, Remco R., 2014. Species delimitation using genome-wide SNP Data, Systematic Biology, March, syu018. http://dx.doi.org/10.1093/sysbio/syu018.
Linossier, Juliette, Zsebok, Sándor, Baudry, Emmanuelle, Aubin, Thierry, 2016. Acoustic but no genetic divergence in migratory and sedentary populations of blackcaps, *Sylvia atricapilla*. Biol. J. Linn. Soc. 119 (1), 68–79. http://dx.doi.org/10.1111/bij.12799.
Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Can. Res. 27, 209–220.
Mason, Nicholas A., Taylor, Scott A., 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a holarctic songbird. Mol. Ecol. 24 (12), 3009–3025. http://dx.doi.org/10.1111/mec.13140.
Mayr, Ernst, 1942. Systematics and the Origin of Species. Columbia University Press, New York, NY, USA.
Mayr, Ernst, O'Hara, Robert J., 1986. The biogeographic evidence supporting the pleistocene forest refuge hypothesis. Evolution 40 (1), 55–67. http://dx.doi.org/10.2307/2408603.
Mayr, E., Short, L.L., 1970. Species taxa of North American birds: a contribution to comparative systematics.
McCarthy, Eugene.M., 2006. Handbook of Avian Hybrids of the World. Oxford University Press, New York, NY.
Mettler, R., Schaefer, H.M., Chernetsov, N., Fiedler, W., Hobson, K.A., et al., 2013. Contrasting patterns of genetic differentiation among Blackcaps (*Sylvia atricapilla*) with divergent migratory orientations in Europe. PLoS ONE 8 (11), e81365. http://dx.doi.org/10.1371/journal.pone.0081365.
Outlaw, R.K., Voelker, G., Outlaw, D.C., 2007. Molecular systematics and historical biogeography of the Rock-Thrushes (Muscicapidae: Monticola). Auk 124. http://dx.doi.org/10.1642/0004-8038(2007) 124[561:MSAHBO]2.0.CO;2.
Paynter Jr., R.A., 1968. Check-List of Birds of the World, vol. 14. Museum of Comparative Zoology, Cambridge, MA.
Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. Philos. Mag. Ser. 6 2 (11), 559–572. http://dx.doi.org/10.1080/14786440109462720.
Peterson, Brant K., Weber, Jesse N., Kay, Emily H., Fisher, Heidi S., Hoekstra, Hopi E., 2012. Double digest RADseq: an inexpensive method for De Novo SNP discovery and genotyping in model and non-model species. Edited by Ludovic Orlando. PLoS ONE 7 (5), e37135. http://dx.doi.org/10.1371/journal.pone.0037135.
Phillimore, Albert B., Orme, C. David L., Thomas, Gavin H., Blackburn Tim M., Bennett Peter M., Gaston Kevin J., Owens, Ian P. F., Associate Editor: Heard, Stephen B., and Editor: Whitlock, Michael C., 2008. Sympatric speciation in birds is rare: insights from range data and simulations. Am. Nat. 171 (5): 646–657. http://dx.doi.org/10.1086/587074.
Pritchard, Jonathan K., Stephens, Matthew, Donnelly, Peter, 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2), 945–959.
Pulido, F., Berthold, P., Mohr, G., Querner, U., 2001. Heritability of the timing of autumn migration in a natural bird population. Proc. R. Soc. Lond. B: Biol. Sci. 268 (1470), 953–959. http://dx.doi.org/10.1098/rspb.2001.1602.
Rambaut, A., Drummond, A., 2003. Tracer: A Program for Analysing Results From Bayesian MCMC Programs Such as BEAST & MrBayes. University of Edinburgh, UK.
Rolshausen, Gregor, Segelbacher, Gernot, Hobson, Keith A., Martin Schaefer, H., 2009. Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. Curr. Biol. 19 (24), 2097–2101. http://dx.doi.org/10.1016/j.cub.2009.10.061.
Slager, David L., Battey, C.J., Bryson Jr., Robert W., Gary, Voelker, John, Klicka, 2014. A multilocus phylogeny of a major new world avian radiation: the Vireonidae. Mol. Phylogenet. Evol. 80, 95–104. http://dx.doi.org/10.1016/j.ympev.2014.07.021.
Smith, Brian Tilston, Harvey, Michael G., Faircloth, Brant C., Glenn, Travis C., Brumfield, Robb T., 2013. Target Capture and Massively Parallel Sequencing of Ultraconserved Elements (UCEs) for Comparative Studies at Shallow Evolutionary Time Scales". Systematic Biology, syt061.
Smith, Brian Tilston, McCormack, John E., Cuervo, Andres M., Michael, J., Hickerson, Alexandre Aleixo, Cadena, Carlos Daniel, Perez-Eman, Jorge, et al., 2014. The drivers of tropical speciation. Nature 515 (7527), 406–409.
Stamatakis, Alexandros, 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, January, btu033. http://dx.doi.org/10.1093/bioinformatics/btu033.
Team, R.C., 2014. R: A language and environment for statistical computing (Version 3.0. 2). R Foundation for Statistical Computing, Vienna, Austria.
Toews, D.P., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21 (16), 3907–3930.
Toews, David P.L., Taylor, Scott A., Vallender, Rachel, Brelsford, Alan, Butcher, Bornwyn G., Messer, Philipp W., Lovette, Irby J., 2016. Plumage genes and little else distinguish the genomes of hybridizing Warblers. Curr. Biol. 26 (17), 2313–2318. http://dx.doi.org/10.1016/j.cub.2016.06.034.
Vieillot, L.J.P., 1817. Nouv. Dict. Hist. Nat. 11, 174.
Voelker, Gary, Light, Jessica E., 2011. Palaeoclimatic events, dispersal and migratory losses along the Afro-European Axis as drivers of biogeographic distribution in Sylvia Warblers. BMC Evol. Biol. 11 (1), 1–13. http://dx.doi.org/10.1186/1471-2148-11-163.
Winger, Benjamin M., Keith Barker, F., Ree, Richard H., 2014. Temperate origins of long-distance seasonal migration in new world songbirds. Proc. Natl. Acad. Sci. 111 (33), 12115–12120. http://dx.doi.org/10.1073/pnas.1405000111.
Zink, Robert M., Kessen, Ann E., Line, Theresa V., Blackwell-Rago, Rachelle C., 2001. Comparative phylogeography of some aridland bird species. The Condor 103 (1), 1–10. http://dx.doi.org/10.1650/0010-5422(2001) 103[0001:CPOSAB]2.0.CO;2.