Dense geographic and genomic sampling reveals paraphyly and a cryptic lineage in a classic sibling species complex

Ethan Linck[1]*, Kevin Epperly[1], Paul van Els[2,3], Garth M. Spellman[4], Robert W. Bryson Jr.[1], John E. McCormack[5], Ricardo Canales-del-Castillo[6], John Klicka[1]

[1]*Burke Museum of Natural History & Department of Biology, U. of Washington, Seattle, WA USA*
[2]*Groningen Institute for Evolutionary Life Sciences, U. of Groningen, Groningen, NL*
[3]*Museum of Natural Science, Louisiana State U., Baton Rouge, LA USA*
[4]*Department of Zoology, Denver Museum of Nature & Science, Denver, CO USA*
[5]*Moore Laboratory of Zoology, Occidental College, Los Angeles, CA USA*
[6]*Facultad de Ciencias Biológicas, U. Autónoma de Nuevo León, San Nicolás de los Garza, NL MX*

*Corresponding author: elinck@uw.edu

Abstract:

Incomplete or geographically biased sampling poses significant problems for research in phylogeography, population genetics, phylogenetics, and species delimitation. Despite the power of using genome-wide genetic markers in systematics and related fields, approaches such as the multispecies coalescent remain unable to easily account for unsampled lineages. The *Empidonax difficilis / E. occidentalis* complex of small tyrannid flycatchers (Aves: Tyrannidae) is a classic example of a widely distributed species with limited phenotypic geographic variation that was broken into two largely cryptic (or "sibling") lineages following extensive study. Though the group is well-characterized north of the U.S. Mexico border, the evolutionary distinctiveness and phylogenetic relationships of southern populations remain obscure. In this paper, we use dense genomic and geographic sampling across the majority of the range of the *E. difficilis / E. occidentalis* complex to assess whether current taxonomy and species limits reflect underlying evolutionary patterns, or whether they are an artifact of historically biased or incomplete sampling. We find that additional samples from Mexico render the widely recognized species-level lineage *E. occidentalis* paraphyletic, though it retains support in the best-fit species delimitation model from clustering analyses. We further identify a highly divergent unrecognized lineage in a previously unsampled portion of the group's range, which a cline analysis suggests is more reproductively isolated than the currently recognized species *E. difficilis* and *E. occidentalis.* Our phylogeny supports a southern origin of these taxa. Our results highlight the pervasive impacts of biased geographic sampling, even in well-studied vertebrate groups like birds, and illustrate what is a common problem when attempting to define species in the face of recent divergence and reticulate evolution.

Keywords: ultraconserved elements, RADseq, phylogenomics, Tyrannidae

Introduction:

The effects of incomplete or biased taxon sampling are a perennial concern in evolutionary biology (Hillis 1998, Heath et al. 2008; Rosenberg et al. 2005; Schwartz and McKelvey 2009; Puechmaille 2016). In phylogenetics, missing taxa can pose significant problems for tree estimation (Felsenstein 1978; Hillis 1998; Reddy et al. 2017), with cascading consequences for macroevolutionary analysis, biogeography and studies of trait evolution (e.g. FitzJohn et al. 2009, Ahrens et al. 2016). In

population genetics, sampling scheme can drive spurious clustering and influence the value of summary statistics (Rosenberg et al. 2005; Schwartz and McKelvery 2009; Chikhi et al. 2010), in turn biasing inference of selection (Excoffier et al. 2009) and genome-wide association studies ('GWAS'; Astle and Balding 2009).

For systematists studying nonmodel organisms, this source of error is often compounded by geographic biases in biological knowledge. Both range size and body size can be correlated with the probability of a given species being described (Blackburn and Gaston 1995; Collen et al. 2004), but are also variables with strong relationships to latitude in many taxa (e.g. Stevens 1989; Freckleton et al. 2003; Reimer et al. 2018). Taxonomic practices and species concepts also vary across space: species are often more narrowly defined (or "split") at higher latitudes (Jones et al. 2012, Barrowclough et al. 2016), in part reflecting the higher density of scientific institutions in Europe and North America.

As species are widely cited as the fundamental unit of biodiversity (Hull 1977, Coyne 1994, Magurran 2013, Hohenegger 2014; de Queiroz 2005), the implications of misdiagnosis or unevenly applied diagnostic criteria are significant. When logistical or political constraints prevent wider geographic sampling, taxonomic decisions are forced to rely on samples from an incomplete or unrepresentative portion of a given species' range (Avendaño et al. 2017). Given that lower latitudes are likely to contain greater genetic diversity and a greater number of distinct lineages (Harvey and Brumfield 2015, Smith et al. 2014, Barrowclough et al. 2016, Smith et al. 2017), species with large ranges that were originally delimited from sampling biased towards higher latitudes may eventually prove paraphyletic (McKay and Zink 2010; Zucker et al. 2016), requiring revised taxonomic treatments that affect biodiversity estimates and conservation priorities along with the numerous disciplines dependent on accurate species lists (Jones et al. 2012, Barrowclough 2016).

These concerns are especially salient for delimitation of cryptic and sibling species, where the absence of easily diagnosable morphological variation across ranges limits the ability to make informed choices about the potential significance of sample exclusion on inference. Molecular phylogenetics has revolutionized the study of cryptic and sibling lineages (e.g. Knowlton 1993; de Leon and Nadler 2010; Adams et al. 2014; Papakostas et al. 2016) and substantially altered our understanding of the relationship between reproductive isolation and phenotype, geography, and genetic variation (Bickford et al. 2007; Fišer et al. 2018). The recent emergence of model-based species delimitation with genome-wide genetic markers aids in reducing bias due to arbitrary divergence thresholds and the stochasticity of a small sample of genes (Leaché et al. 2014; Hoatling et al. 2016). However, these approaches remain unable to correct for phylogenetic relationships with unsampled lineages, underscoring the continued importance of geography to systematic biology.

Tyrannid flycatchers (Aves: Tyrannidae) are the largest bird family in the New World, and have long posed problems for taxonomists and systematists due to limited morphological variation in many groups, conflicting characters, and evidence of convergent evolution (Traylor 1977; Birdsley 2002; Chaves et al. 2008). Of the tyrannids, the *Empidonax difficilis / E. occidentalis* complex provides a representative and classic example of widely-distributed species with limited geographic variation, broken into separate, largely cryptic, lineages only after extensive study (Johnson 1980, Zink and Johnson 1984, Johnson and Martin 1988). Based on quantitative analysis of museum specimens and vocalizations across its full range, in addition to putative assortative mating at a contact zone, the Western Flycatcher *E. difficilis* was split into interior western North American (Cordilleran Flycatcher, *E. occidentalis*) and coastal western North American (Pacific Flycatcher, *E. difficilis*) species (Johnson 1980), themselves sister to the allopatric species *E. flavescens* (Yellowish Flycatcher) which occurs throughout Central American highlands, largely east of the Isthmus of

2

Tehuántepec. Subsequent population genetic analysis using allozyme data supported the reciprocal monophyly of these groups, though sampling was restricted to populations in the United States (Johnson and Marten 1988). Both an analysis of a contact zone in interior southwest Canada (Rush et al. 2009) and unpublished mtDNA and nuDNA phylogeography studies (Klicka et al. unpublished; Rush et al. in prep) suggest locally extensive hybridization between forms, but to date, the affinities of resident and migratory populations in Mexico remain unexamined with genetic data.

Here, we explore the influence of biased sampling on determining species limits in this complex of *Empidonax* flycatchers. We combine dense population sampling with RADseq, UCE, and mtDNA sequence data from across the majority of the latitudinal range of the currently recognized species *E. difficilis* and *E. occidentalis.* Using these data, we ask whether comprehensive geographic and genomic sampling changes estimated phylogenetic relationships and species limits in a classic sibling species complex, shedding light on our ability to infer fundamental evolutionary units and processes in morphologically conserved, geographically widespread taxa.

Materials and Methods:

Sampling, Library Preparation, and Sequencing: We collected 308 vouchered tissues from four *Empidonax species*, with localities in Canada, the United States, Mexico, Guatemala, Honduras, and Panama (**Figure 1**; **Table S1**). Of these, 295 belonged to our ingroup, the *E. difficilis / E. occidentalis* complex. We represented the entirety of the group's geographic range with the exception of *E. difficilis* populations in northern British Columbia and Alaska and *E. occidentalis* populations in southeastern British Columbia and Alberta, a region previously examined in detail by Rush et al. 2009. The remaining 13 samples in our study belonged to outgroup taxa: *E. flaviventris* (n = 4), a long-distance migrant that breeds in North American boreal habitats, and *E. flavescens* (n = 9), a Central American endemic.

We extracted total genomic DNA from all samples using a Qiagen DNeasy tissue extraction kit, following the manufacturer's recommended protocol. For all samples, we sequenced 1041 bp of the mitochondrial gene NADH dehydrogenase subunit 2 (ND2), using primers L5215 (Hackett 1996) and TrC (Miller et al. 2007) and 12.5 μL PCR reactions on a T100 thermal cycler (Bio-Rad, Hercules, CA) as follows: 94°C for 2.5 min, 35 cycles of (94°C for 30 s, 54°C for 30 s, 72°C for 1 min), 10 min at 72°C, 10°C hold. We sent PCR products to the High-Throughput Genomics Unit at the University of Washington for cleanup and sequencing.

Based on fragment size and DNA concentration, we then chose a 144-sample subset for reduced representation genome sequencing with the double digest restriction-associated DNA sequencing library preparation method (Peterson et al. 2012). This subset included 134 ingroup samples (46 *E. difficilis*, 88 *E. occidentalis*) and 10 outgroup samples (two *E. flaviventris*, eight *E. flavescens*) representing all four currently recognized species mentioned above. We used digestion enzymes `Sbf1` and `Msp1` and a Pippin Prep size selection device to retain fragments between 415 and 515 bp. Our libraries were then sequenced for 100 bp single-end reads on an Illumina HiSeq 2500 at University of California Berkeley's QB3 Vincent J. Coates Genomics Sequencing Laboratory. After sequencing, we excluded three ingroup samples and one outgroup sample (*E. flaviventris*) due to poor data quality.

Finally, we chose a second, nonexclusive subset of 37 individuals for the reduced representation sequencing approach of targeted capture for ultraconserved elements and their flanking regions (UCEs; Smith et al. 2014). This subset focused on the southwestern US, Mexico, and Central
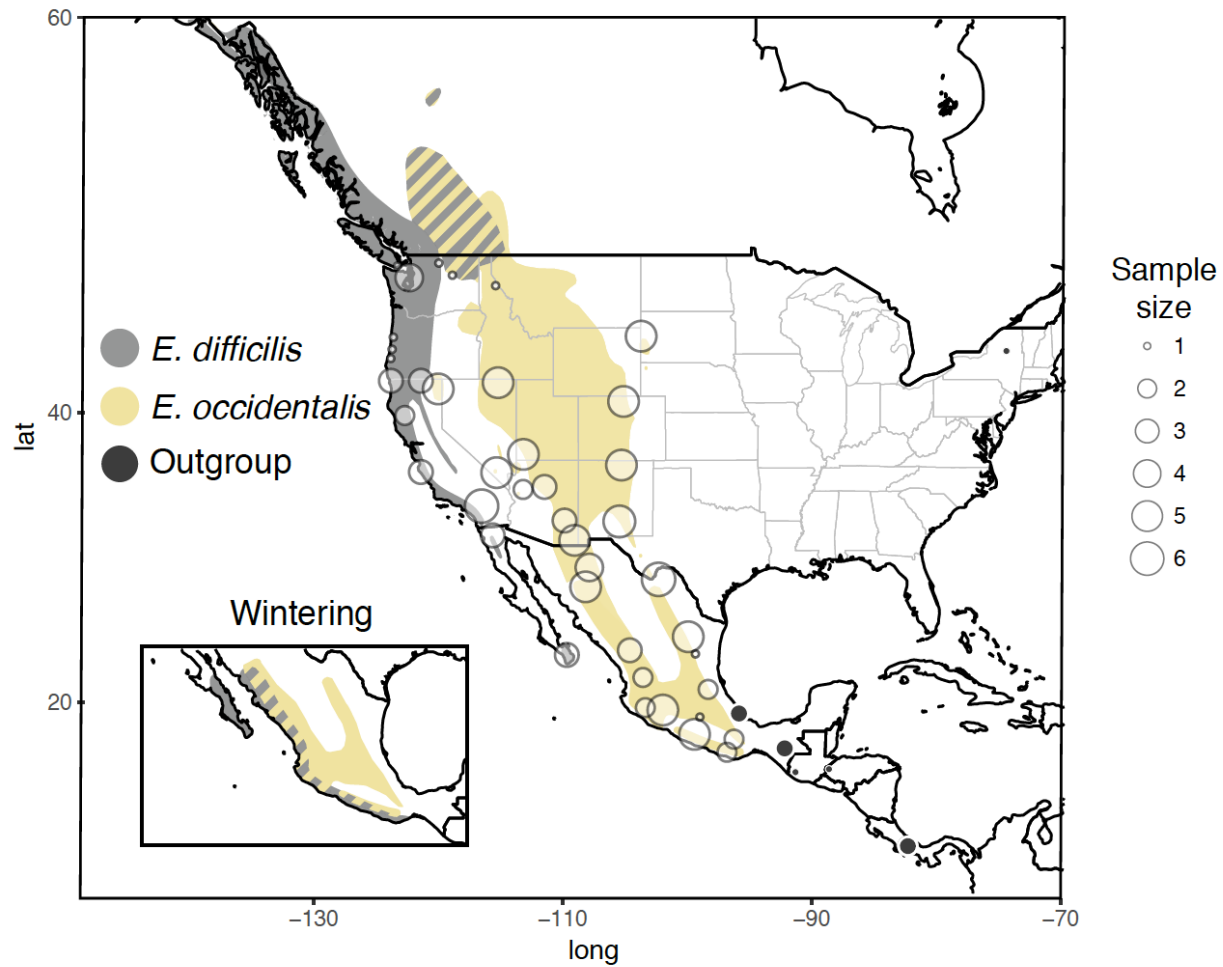
3

Figure 1: Localities and sample size for the *Empidonax difficilis / E. occidentalis* complex, with outgroups *E. flaviventris* and *E. flavescens*.

America. It also included all four species, with 34 ingroup samples (1 *E. difficilis*, 33 *E. occidentalis*) and four outgroup samples (1 *E. flaviventris*, 3 *E. flavescens*). We sent raw extractions to RAPiD Genomics (Gainesville, FL USA), where they enriched each pool using a set of 5,472 custom-designed probes (MYbaits; MYcroarray, Inc., Ann Arbor, MI, USA) targeting 5,060 UCE loci (Faircloth et al. 2012; a full protocol can be found on http://ultraconserved.org.) Pooled libraries were then sent to the University of Florida ICBR Facility for 150 bp read paired-end sequencing on an Illumina NextSeq500.

Sequence Assembly: We edited and aligned forward and reverse chromatograms from ND2 sequences using `Sequencher` v. 4.9 and the `MUSCLE` alignment algorithm implemented in `MEGA` v. 7.0 (Kumar, Stecher, and Tamura 2015). We performed *de novo* sequence assembly and variant detection from ddRADseq reads using `ipyrad` v. 0.7.8 (Eaton 2014), implementing a clustering threshold of 85% and a minimum read depth of 6. To test the effect of increased phylogenetic distance on SNP number and alignment completeness, we repeated all steps in the `ipyrad` pipeline with outgroup individuals excluded, ultimately performing all phylogenetic analyses with the full set of taxa, and population assignment analyses using only the ingroup. Our final sequence alignment only included loci where no more than 50% of individuals were missing, and only included individuals with more than 50% of all loci sequenced, resulting in the loss of three samples. To exclude putative paralogs, we dropped loci with more than 5 heterozygous sites, and included no loci sharing a heterozygous site across more than 50% of individuals. We assembled UCE reads using a modified version of the `PHYLUCE` pipeline (Faircloth 2015), described in detail elsewhere (Bryson et al. 2017), and called SNPs using the custom pipeline described in Zarza et al. (2018), with our most completely assembled sample (UWBM 118082; *E. occidentalis* from Nuevo León, MX) serving as the reference.

Phylogenetic Inference: For each of our three datasets, we inferred phylogenetic relationships among taxa with two methods. We first inferred maximum-likelihood phylogenies in `RAxML` v 8.2.4 (Stamatakis 2006) via the `CIPRES` portal (Miller et al. 2010). After determining the best-fit model of sequence evolution for our concatenated data matrices (`GTRGAMMA`), we ran the program for 1000 bootstrap replicates to estimate node support.

We next used `SVDquartets` (Chifman and Kubatko 2014), implemented in `PAUP` v 4.0 (Swofford 2002), to estimate both a phylogeny of individual lineages and a species tree under a coalescent model robust to the influence of gene flow between sister lineages (Long and Kubatko 2018). `SVDquartets` randomly samples quartets (sets of four taxa, or tips) from the data matrix and calculates the singular value decomposition score to evaluate their optimal relationship. The program then combines these quartets algebraically to infer the overall species tree, and estimates branch support with 100 nonparametric bootstrap replicates. For our ddRADseq data, we additionally inferred a species tree using only individuals determined to represent "pure" populations, which we defined as $\leq 5\%$ admixed, based on our `structure` analysis (see below).

Population Structure Inference: To identify genetically differentiated populations (or species) and identify putative regions of admixture among them, we performed multivariate ordination and Bayesian coalescent clustering on ddRADseq and UCE SNP datasets independently. We conducted a principal-components analysis (PCA; Pearson 1901) using the covariance matrix of allele frequencies in each sample. We then identified putative genetic clusters with k-means clustering using the R package `adegenet` v2.0.1 (Jombart 2008), setting the maximum number of clusters to 10 and retaining all principal components. Next, we conducted Bayesian clustering under a coalescent model allowing for admixture using the program `structure` (Pritchard et al. 2000). Because including singletons can strongly bias inference of population structure (Linck and Battey 2018), we used a custom script to restrict our data matrix to only parsimony-informative SNPs

(`https://github.com/slager/structure_parsimony_informative`). We employed default priors, correlated allele frequencies, and stipulated a chain length of 1,000,000 with the first 100,000 steps discarded as burn-in. We replicated structure analyses 10 times for each value of *k* from 1–8, assessed change in marginal likelihood across values of *k* using `STRUCTURE HARVESTER` (Earl 2012), and used `CLUMPP` (Jakobsson and Rosenberg 2007) to average permuted matrices across replicates while controlling for label switching. We plotted individual ancestry proportions using `structurePlotter` v 1.0 (Battey 2017).

Because `structure` may not reliably infer population structure when population sample sizes are uneven (Puechmaille 2016), we also performed all population structure analyses with replicated subsampling. We first assigned samples to putative populations based on their assigned *k*-means cluster, using the subdivision scenario where Bayesian information criterion (BIC) showed the greatest change in value (*k*=4 for both datasets). From each of these populations, we next randomly selected the number individuals equal to sampling the smallest population (n=7 for ddRADseq; n=2 for UCEs), while retaining admixed individuals. For the RADseq dataset, we repeated this process until all individuals were included in at least two of the subsets resulting in a total of 18 subsets. These data matrices were then used for PCA / *k*-means clustering and `structure` analyses with identical parameters to those described above, with the final RADseq `structure` results plotted from the averaged *Q*-value for each individual from across all subset runs.

Genomic Cline Analysis: To assess the relative level of reproductive isolation between pairs of taxa with adjacent geographic ranges, we analyzed the width and relative position of genomic clines using the R package `HZAR` (Derryberry et al. 2013). We selected three transects across regions identified as hybrid zones in our population structure analyses, representing contact points between USE and USW, USW and MXO, and MXO and MXS (**Figure S6**; see **Results** section for definitions of regional abbreviations).

We first calculated total transect length and the distance between each transect locality using the R package `geosphere` (Hijmans 2017). Using averaged *Q*-value from structure analyses using our 18 individual subsampled matrix, we selected scores for each species as follows: for Transect 1 (between USW and USE), we used *Q*-values from USE; for Transect 2 (between USE and MXO), we used *Q*-values from MXO; and for Transect 3 (between MXO and MXS), we used *Q*-values from MXS. Setting maximum cline width to +/-100 km of geographic distance sampled to increase the efficiency of the algorithm, we ran 10 models. These consisted of the presence of exponential decay on both ends of the cline in isolation, on both ends of the cline together, or the absence of exponential decay on either end of the cline, repeated with a fixed and with a free scaling parameter.

Checking to ensure the convergence of each chain, we ran 3 MCMC chains for 1,000,000 generations with 10,000 generations of burn-in for each model. We then concatenated results from each independent chain for each model, and selected among our competing models using corrected Akaike Information Criterion (AICc) scores. We extracted maximum likelihood cline center and width values from the best model, expressing confidence intervals of each parameter as 2 log-likelihoods. We modeled maximum likelihood clines using 95% of the credible cline region.

Results

Sequencing and Assembly: We successfully amplified and sequenced ND2 for all 308 samples. Illumina sequencing of our ddRADseq libraries returned an average of 1,241,201 quality filtered reads per sample. Clustering within samples while including outgroup taxa resulted in a total of 594,004
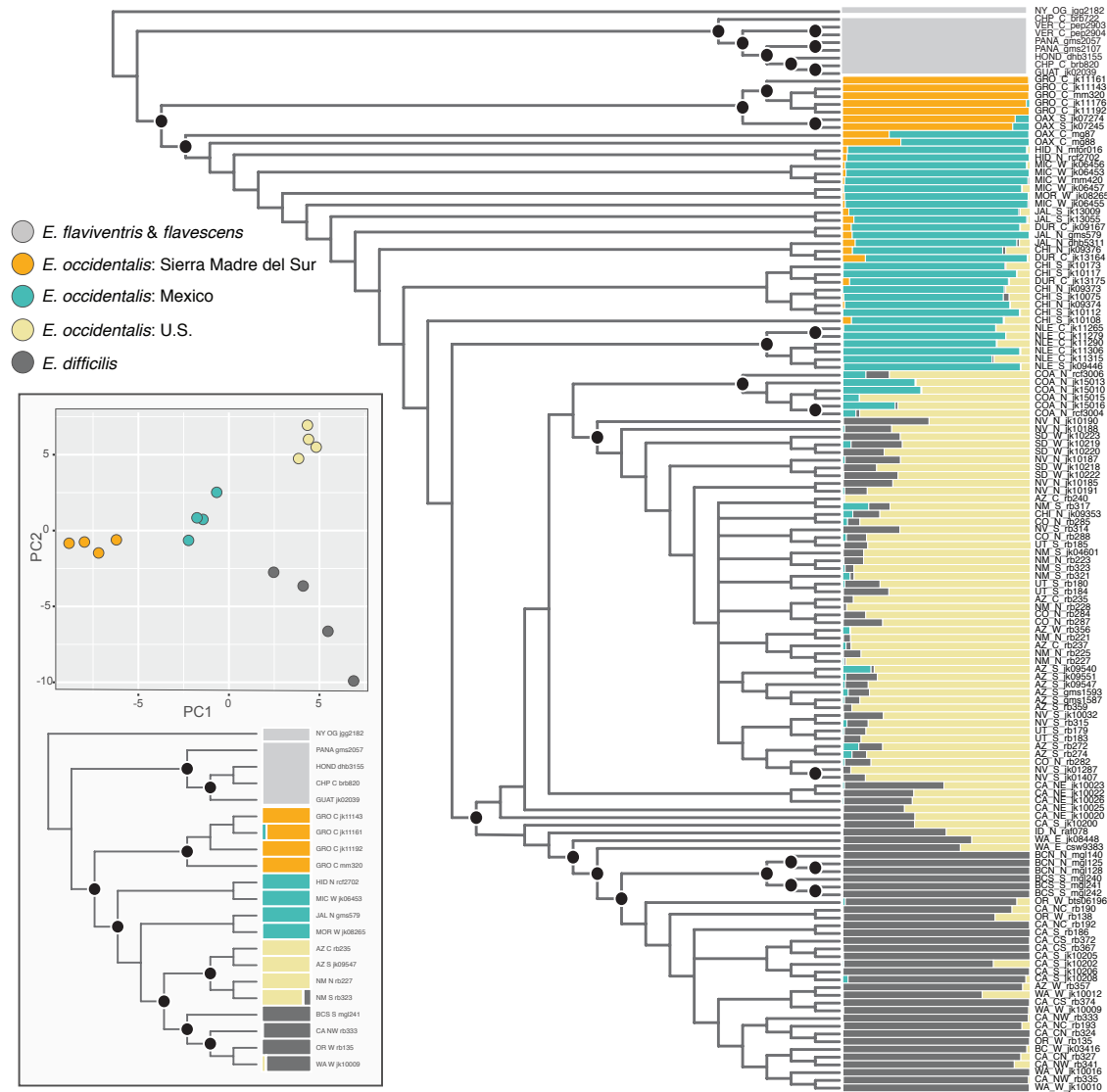
Figure 2: Population limits and phylogenetic relationships estimated using ddRADseq SNP data. (L): SVDquartets species tree of "pure" populations and PCA. (R): SVD quartets tree including all individuals and paired with structure ancestry proportions at k=4.

putative loci, with an average depth of coverage of 26.7. After clustering across individuals and applying paralog and depth-of-coverage filters, we retained an average of 6,880 loci per sample, which were 631,934 bp long when concatenated. Variant detection resulted in a total of 50,210 SNPs, which we reduced to 6,808 unlinked SNPs. After applying a filter to select only parsimony-informative sites, this number was further reduced to 3,427 SNPs. Excluding outgroup taxa, we identified a total of 55,410 putative loci with an average depth of coverage of 27.2. Clustering across individuals and filtering reduced this to 7,319 loci and 670,994 bp. These loci contained 44,757 total SNPs, 7,194 unlinked SNPs, and 4,048 unlinked parsimony informative SNPs. Our four evenly subsampled alignments ranged from 1,864 to 1,910 parsimony informative SNPs.

UCE libraries returned an average of 1,404,000 quality filtered reads per sample. After assembly and alignment to probe sequences, our data matrix included 4,203 UCE loci from all samples for a total concatenated length of 2,066,792 bp. Mean locus length was 470 bp, with a range of 162–1,626. There were 21,307 parsimony-informative SNPs across all loci. Reducing our data matrix to evenly sampled populations, we retained 8,970 parsimony informative SNPs. We deposited aligned ddRADseq and UCE matrices on Dryad (doi: pending) and raw reads on NCBI's Sequence Read Archive (SRR: pending).

Phylogenetic Inference: Phylogenetic relationships were broadly concordant across different inference methods (**Figures 2, S8, S9**). Our maximum likelihood phylogeny, SVDquartets individual lineage tree, and SVDquartets species tree from ddRADseq SNP data supported the monophyly of the *E. difficilis / E. occidentalis* complex relative to *E. flavescens*. Both also recovered a fully-supported sister relationship between a clade containing all unadmixed individuals from from the Sierra Madre del Sur of Guerrero and Oaxaca in southern Mexico (hereafter "MXS") and a clade containing individuals in the Transvolcanic Belt, Sierra Madre Occidental, and Sierra Madre Oriental of Mexico ("MXO"), individuals in the US Rocky Mountains and Great Basin ("USE") and individuals in the US Sierra Nevada, Cascades, and Pacific Coast Ranges ("USW").

Within the *E. difficilis / E. occidentalis* clade, our species tree recovered fully supported subclades corresponding to USW and USE, nested within a grade of MXO individuals. Maximum likelihood inference recovered fully supported clades coinciding with the majority of USW, USE, and MXO individuals, but nested within a grade of individuals of putative hybrid ancestry (see below). Discordant nodes between maximum-likelihood and species trees lacked bootstrap support in all instances. A species tree inferred with our UCE SNP dataset recovered relationships within USE and MXO individuals concordant with those in our ddRADseq trees (**Figure S8**). Maximum likelihood inference with the mitochondrial gene ND2 differed only from SNP datasets by placing MXS individuals sister to a clade comprised of *E. flavescens* and USW / USE (**Figure S9**), rendering the latter paraphyletic. This placement, however, was not well supported.

Table 1: Likelihood scores for different population assignment / species limit partitioning schemes (values of k in `structure`).

k	Replicates	Mean	Standard Deviation
1	10	-190860.17	2.23
2	10	-183007.93	683.88
3	10	-176774.65	34.33
4	10	-176606.12	363.79
5	10	-237514.34	111236.38
6	10	-269824.40	129220.10

8

Population Structure Inference: Assessing clustering across multiple values of *k* with Bayesian Information Criteria revealed a single shift in slope at *k*=4. This result was corroborated by mean estimated likelihood values for assignment schemes in `structure`, which narrowly supported a four-population model over a three-population model (**Table 1**). Individual population assignments were broadly concordant between *k*-means and Bayesian clustering approaches. The primary difference was driven by `structure's` ability to model "admixture" (e.g., to minimize assignment uncertainty by invoking proportional ancestry contributions from multiple populations), which identified a number of individuals of putative admixed ancestry. These admixed individuals were predominately found in northern Coahuila, Mexico, and northeastern California.

Genomic Cline Analysis: For Transect 1, the best model included fixed scaling and exponential decay in the right end tail of the cline (AICc=-24.368). The best model for Transect 2 included fixed scaling and no exponential decay in either tail (AICc=-62.644). The best model for Transect 3 also included fixed scaling and no decay (AICc=-49.615). Transect 3 was narrowest, with a mean width of 73.789 km, followed by Transect 1, with a mean width of 210.078 km, and then Transect 2, with a mean width of 574.840 km (**Figure 4**). Full confidence intervals are provided in **Table S2**.

Discussion:

Geographic Bias in Sampling and Species Limits: Increased sampling of low-latitude populations revealed that the *E. difficilis* / *E. occidentalis* complex contains four parapatric lineages, all of which interbreed to varying degrees with neighboring populations and / or share a significant amount of ancestral polymorphism (**Figure 2**). The apparent paraphyly of *E. occidentalis* with respect to *E. difficilis* and previously unrecognized distinctiveness of the population of *E. occidentalis* in Mexico's Sierra Madre del Sur provide an important cautionary tale for systematists and phylogeneticists attempting to proscribe species limits with incomplete geographic sampling. Few vertebrate faunas have been as intensively studied as North American birds, yet our results demonstrate that the presence of unsampled populations or sister species in less well-studied regions of a taxon's range may dramatically alter species-delimitation schemes and inferred phylogenetic hypotheses—even when these unsampled lineages appear to be peripheral, small in population size, or otherwise 'unimportant.' The presence of strong latitudinal gradients in species richness (Willig et al. 2003) and intraspecific genetic diversity (Smith et al. 2017) increase the probability that many lower-latitude populations may represent distinct lineages. We encourage biologists working with taxa found across large latitudinal ranges to make efforts to include samples from the full extent of their distribution whenever possible, and consider the biases of missing samples when impossible.

Extremely limited phenotypic variation in tyrannid flycatchers broadly and *Empidonax* in particular has perplexed generations of naturalists, and raised questions about the degree and rate of evolution of reproductive isolation in birds (Zink and Johnson 1984). Our data suggest these studies have at least in part been hampered by a misunderstanding of the true evolutionary relationships and total diversity of these taxa. Would detailed study of other *Empidonax* alter our understanding of evolution across the genus, informing our understanding of trait evolution, reproductive isolation, and phylogenetic relationships? We encourage researchers to prioritize studies of this poorly known group to further resolve questions of species limits.

Species limits, gene flow, and evolutionarily significant units: Our results highlight a conflict between phylogenetic and population genetic (clustering-based) species delimitation schemes that is likely to become increasingly important as genomic datasets reveal previously hidden histories of gene flow
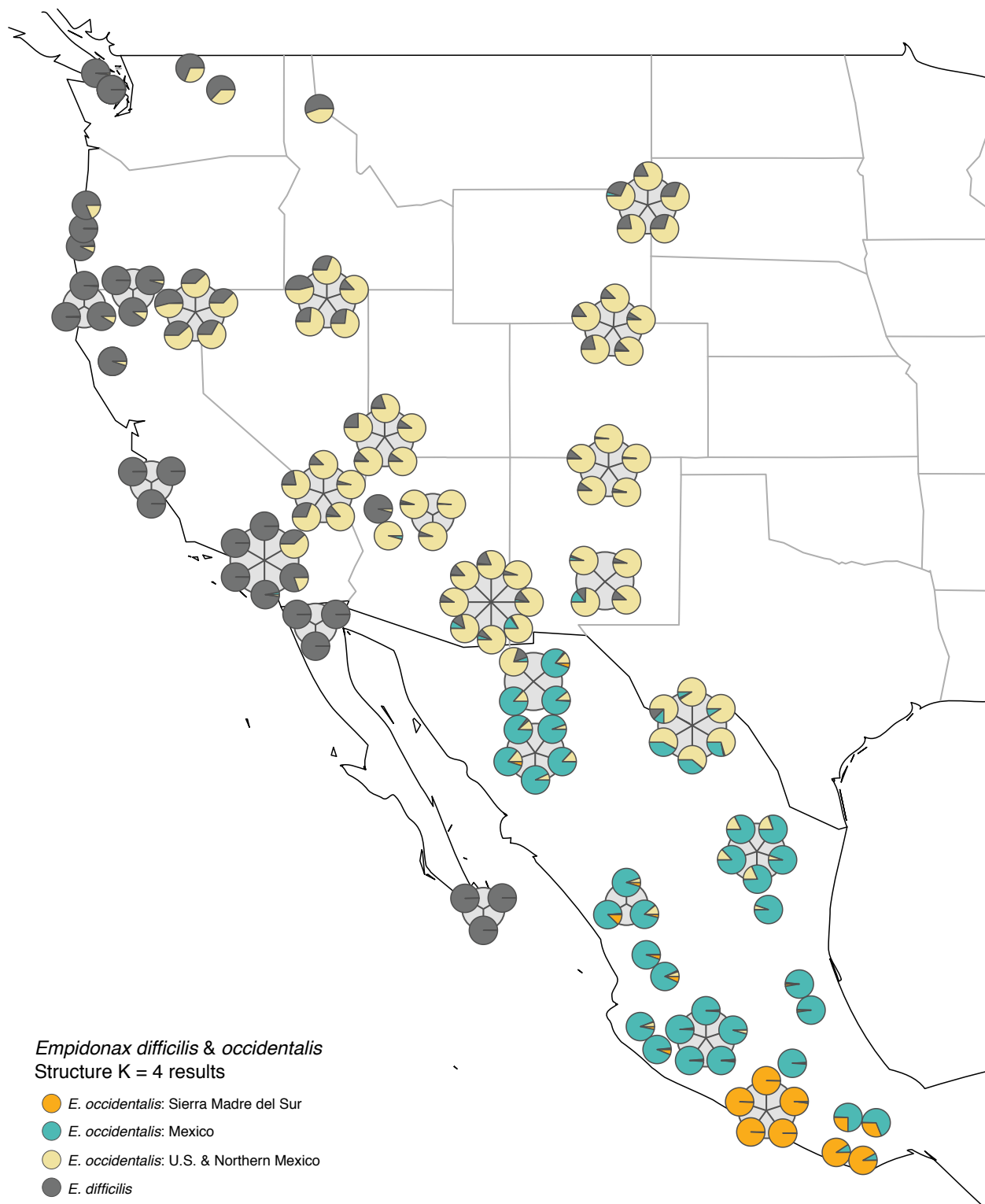
Empidonax difficilis & *occidentalis*
Structure K = 4 results

🟠 *E. occidentalis*: Sierra Madre del Sur

🟢 *E. occidentalis*: Mexico

🟡 *E. occidentalis*: U.S. & Northern Mexico

⚫ *E. difficilis*

Figure 3: Proportional ancestry at $k{=}4$ for *Empidonax difficilis* and *E. occidentalis* individuals. Each circle represents an individual, with clusters of circles representing multiple individuals at a single sampling locality.

and incomplete reproductive isolation during population divergence (Zarza et al. 2016). A strictly phylogenetic species concept encounters two problems as applied to western *Empidonax* flycatchers that limit its utility. First, the species complex includes at least two regions of probable admixture among three geographically and genetically structured groups that appear strong enough to distort phylogenetic inference with both coalescent model-based and maximum likelihood approaches. While admixture proportions in `structure` analyses reflect uncertainty in assignment probability rather than explicitly modeling introgression, and may thus reflect the signature of incomplete lineage sorting rather than secondary contact and introgression in recent divergence events like those in *Empidonax*, we believe that the concentration of admixed individuals at zones of contact between geographically sorted genotypic clusters nonetheless implicates some degree of recent gene flow. Formal tests for introgression (Durand et al. 2011) or phylogenetic network analyses (Wen et al. 2018) could help evaluate the relative contributions of these sources of uncertainty. Second, *E. occidentalis* appears to be a paraphyletic taxon, taxon, composed of a grade in which pure and hybrid *E. difficilis* are nested (though low support for internal nodes may indicate biased inference from incomplete lineage sorting or reticulation; **Figure 2**). Nevertheless, support for a four taxa model from population genetic clustering shows that the division between northern and southern populations of *E. occidentalis* is not arbitrary: although not explicitly modeled in our study, these results suggest a reduction in gene flow between halves of the species range by some combination of habitat limitations, geographic isolation, or incipient reproductive isolation. Regardless of the epistemic reality of species themselves, speciation is clearly a dynamic and continuous process, and evolutionarily significant units do not need to adhere to a strictly bifurcating tree (Turgeon and Bernatchez 2003).

Across *Empidonax* more broadly, cline analysis refutes suggestions of a linear relationship between phenotypic divergence and reproductive isolation (Johnson 1980). MXS populations are essentially indistinguishable from adjacent MXO populations with many standard morphometric measurements (Johnson 1980; Klicka et al. unpublished), but show a steep transition in allele frequencies (**Figure 4**) indicating limited gene flow between groups relative to the more morphologically divergent but less isolated pair of USE and USW. Our study shows the *E. difficilis / E. occidentalis* complex continues to illustrate this "gray area" of contemporary speciation research and species theories by suggesting several confounding processes may be at play: reticulation of lineages, local adaptation, differential rates of trait evolution, and unobserved behavioral differences.

Evolution and Biogeography of Empidonax difficilis *and* E. occidentalis: The range of western North American *Empidonax* flycatchers is typical of an array of roughly codistributed taxa restricted to mountainous Nearctic habitats characterized by forests of pine, oak, and fir. Because much of northern North America was glaciated during the Pleistocene, phylogeographic patterns in these species have been attributed to the expansion out of multiple, isolated, ice-free refugia (Klicka and Zink 1999). Though likely locally distorted by the effects of gene flow, our phylogeny of western *Empidonax* is at least geographically consistent with this model: with Central American *E. flavescens* and *E. flaviventris* as sister taxa, we see sequential splits to the southernmost portion of the range of E. occidentalis in the Sierra Madre del Sur, a split between the Sierra Madre del Sur and the Sierra Madres Occidental / Oriental, the Transvolcanic Belt, and the U.S. Rocky Mountains, followed by a split between these ranges and Pacific Coast Ranges, with continued interbreeding at contact zones of suitable habitat. We believe this sequence most likely reflects dispersal and colonization followed by allopatric divergence, although vicariance and secondary contact through the contraction of appropriate montane habitat as a result of climatic cycling is at least possible. Regardless of the mode of divergence, these data are consistent with previous studies of diversification across the Mexican highlands (Bryson et al. 2010) and with the conclusion that the Sierra Madre del Sur is a

hotspot of phylogenetic endemism (Blancas-Calva et al. 2010; van Els et al. 2014), and we encourage increased attention to this distinctive region.

Are these lineages evolving neutrally, or adapting to the broadly varied climate space across the group's range? Satisfactorily answering this question is beyond the scope of the current paper, but we wish to emphasize the role of topography in isolating *E. difficilis* / *E. occidentalis* populations across their range: each of the four clusters identified in our analysis broadly coincides with a complex of mountain ranges (the Cascades / Sierra Nevada / Coast Ranges, the Rocky Mountains, the Sierra Madre Occidental / Oriental, and the Sierra Madre del Sur) separated by less suitable lowland habitat (the Great Basin, the Sonoran and Chihuahuan Desert, and the Balsas Basin) though linked by discontinuous 'sky islands' in some places. These barriers have been implicated in restricting gene flow in other terrestrial vertebrate taxa (Bryson et al. 2010; Gowen et al. 2014; van Els et al. 2014; Mastretta-Yanes 2015). In light of previous research suggesting ecologic diversification is responsible for divergence of *Empidonax* north of the Mexican border (Johnson and Cicero 2002), however, local adaptation may be playing a role in further reducing gene flow across these barriers.

Taxonomic Recommendations: Our results conflict with contemporary taxonomic treatments, and are consistent with a range of one to four species depending on the invoked species concept. As discussed above, a strict phylogenetic species concept is difficult to apply in the face of probable reticulation, but identifies either a single, widely distributed species (if putative hybrid individuals are included) or two species (if hybrids are excluded), with one lineage inhabiting the Sierra Madre del Sur and the other the remainder of continental North America. A genotypic clustering species definition (Mallet 1995) identifies four distinct species, coinciding with our USW, USE, MXO, and MXS regional definitions. This assignment scheme is further supported by genomic cline analysis: while cline width varies across our sampled transects, either as a consequence of different levels of reproductive isolation or the availability of samples, the underlying sigmoidal model for all three clines is consistent with some degree of selection against hybrids (Endler 1977). We recognize that a pragmatic approach to species delimitation for end users (e.g., field naturalists, conservationists, and ecologists) needs to incorporate diagnostic phenotypic characters, whether in range, ecology, behavior, song or morphology. A detailed analysis of these data will be presented elsewhere.

Conclusions: Re-examining a classic morphologically cryptic sibling species complex with genome-wide data and expanded geographic sampling, we identify a highly distinct unrecognized lineage in a previously unsampled portion of the group's range. We additionally identify the paraphyly of a widely recognized species, although this conflicts with the best-supported species-delimitation model. Our phylogeny supports a model of expansion from southern Mexico. Our results highlight the pervasive impacts of biased geographic sampling, even in well-studied vertebrate groups like birds, and illustrate what is a common problem to phylogenetic species concepts in the face of recent divergence or reticulate evolution.

Data availability and Supplemental Material: Supplemental material, code, and sequence alignments are available via Dryad (pending). Raw sequence data are available from the NCBI SRA (pending).
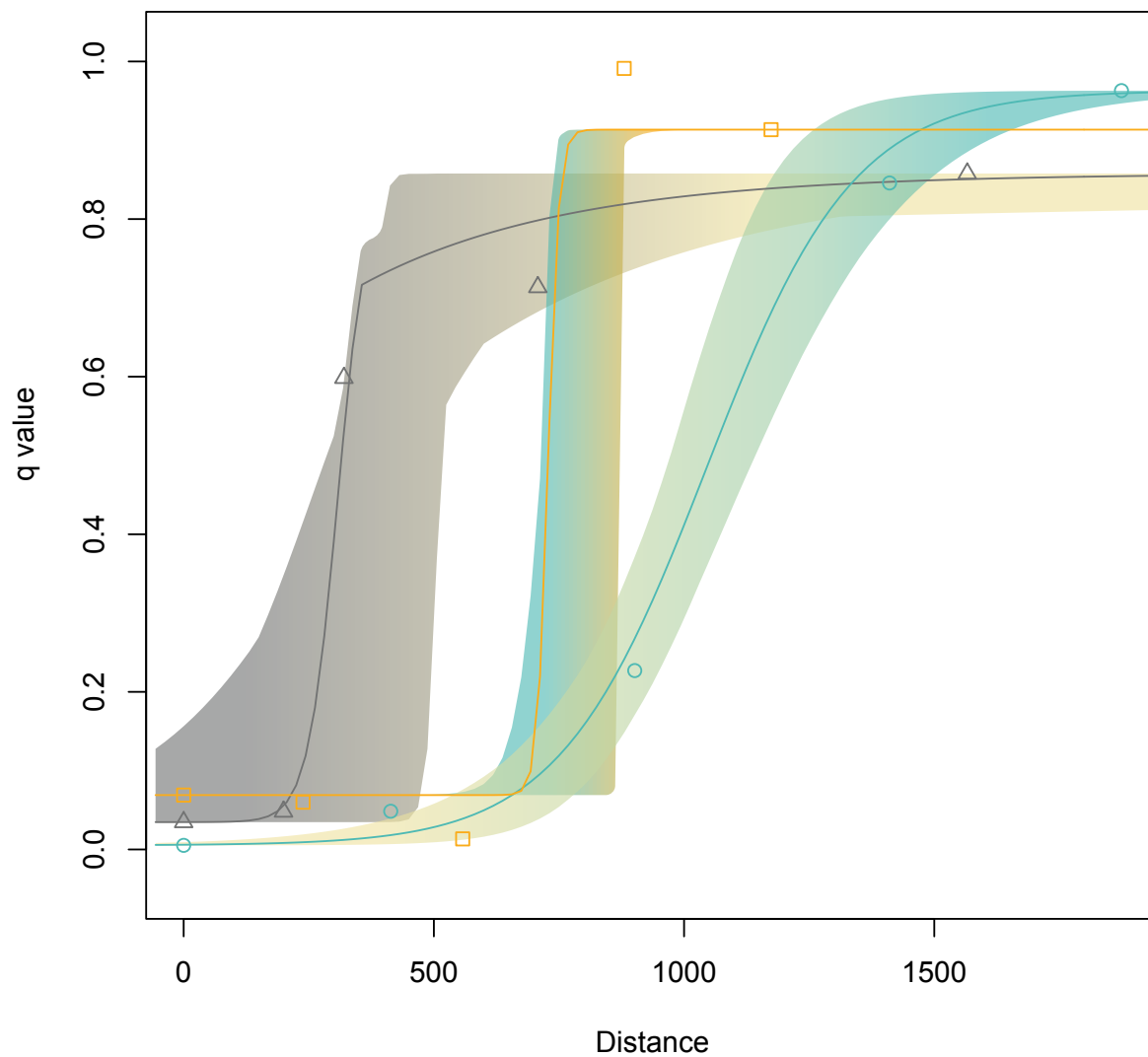
Figure 4: Genomic clines for transects spanning hybrid zones between *Empidonax difficilis* and *E. occidentalis* in NW USA (left, gray triangles; "Transect 1"); between Mexican *E. occidentalis* and resident *E. occidentalis populations* in the Sierra Madre del Sur (center, orange squares; "Transect 3"), and between US *E. occidentalis* and Mexican *E. occidentalis* (right, turquoise squares; "Transect 2"). *Q* value indicates admixture proportion from `structure` analyses.

References:

Avise, J.C. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc R Soc Lond [Biol] 265:457-463.

Adams M., Raadik T.A., Burridge C.P., Georges A. 2014. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? Syst. Biol. 63:518-533.

Ahrens D., Fujisawa T., Krammer H.J., Eberle J., Fabrizi S., Vogler A.P. 2016. Rarity and incomplete sampling in DNA-based species delimitation. Syst. Biol. 65:478-494.

Astle W., Balding D.J. 2009. Population structure and cryptic relatedness in genetic association studies. Stat. Sci. 24:451-471.

Avendaño J.E., Arbeláez-Cortés E., Cadena C.D. 2017. On the importance of geographic and taxonomic sampling in phylogeography: A reevaluation of diversification and species limits in a Neotropical thrush (Aves, Turdidae). Mol. Phylogenet. Evol. 111:87-97.

Barrowclough G.F., Cracraft J., Klicka J., Zink R.M. 2016. How many kinds of birds are there and why does it matter? PloS One 11:e0166307.

Battey C.J. 2017. cjbattey/structurePlotter: structurePlotter_v1 (Version 1). Zenodo. http://doi.org/10.5281/zenodo.345154

Bickford D., Lohman D.J., Sodhi N.S., Ng P.K., Meier R., Winker K., Ingram K., Das, I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22:148-155.

Blackburn T.M., Gaston K.J. 1995. Special paper: What determines the probability of discovering a species?: A study of South American oscine passerine birds. J. Biogeogr. 22:7-14.

Blancas-Calva E., Navarro-Sigüenza A.G., Morrone J.J. 2010. Patrones biogeográficos de la avifauna de la Sierra Madre del Sur. Rev. Mex. Biodiv. 81:561-568.

Birdsley J.S. 2002. Phylogeny of tyrant flycatchers (Tyrannidae) based on morphology and behavior. Auk 119:715–734.

Bryson R.W., Murphy R.W., Lathrop A., Lazcano-Villareal D. 2010. Evolutionary drivers of phylogeographical diversity in the highlands of Mexico: a case study of the Crotalus triseriatus species group of montane rattlesnakes. J. Biogeogr. 38:697-710.

Bryson R.W., Linkem C.W., Pavón-Vázquez C.J., Nieto-Montes de Oca A., Klicka J., McCormack J.E. 2017. A phylogenomic perspective on the biogeography of skinks in the Plestiodon brevirostris group inferred from target enrichment of ultraconserved elements. J. Biogeogr. 44:2033-2044.

Chaves A.V., Clozato C.L., Lacerda D.R., Sari E.H.R., Santos F.R. Molecular taxonomy of Brazilian tyrant-flycatchers (Passeriformes: Tyrannidae). Mol. Ecol. Resour. 8:1169-1177.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317-3324.

Chikhi L., Sousa V.C., Luisi P., Goossens B., Beaumont M.A. 2010. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. Genetics 186:983-995.

Collen B., Purvis A., Gittleman J.L. 2004. Biological correlates of description date in carnivores and primates. Glob. Ecol. Biogeogr. 13:459-467.

Coyne J.A. 1994. Ernst Mayr and the origin of species. Evolution 48:19-30.

de León G.P.P., Nadler S.A. 2010. What we don't recognize can hurt us: a plea for awareness about cryptic species. J. Parasitol. 96:453-464.

De Queiroz K. 2005. Ernst Mayr and the modern concept of species. PNAS 102:6600-6607.

Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239-2252.

Eaton D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844-1849

Endler J.A. 1977. Geographic Variation, Speciation, and Clines. Princeton Univ Press.

Earl D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359-361.

Excoffier L., Hofer T., Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103: 285-298.

Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn, T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717-726.

Faircloth B.C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786-788.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401-410.

Freckleton R.P., Harvey P.H., Pagel M. 2003. Bergmann's rule and body size in mammals. Am. Nat. 161:821-825.

FitzJohn R.G., Maddison W.P., Otto, S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595-611.

Fišer C., Robinson C.T., Malard F. 2018. Cryptic species as a window into the paradigm shift of the species concept. Mol. Ecol. 27:613-635.

Gowen F.C., Maley J.M., Cicero C., Peterson A.T., Faircloth B.C., Warr T.C., McCormack J.E. 2014. Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. BMC Evol. Biol. 14:135.

Harvey M.G., Brumfield R.T. 2015. Genomic variation in a widespread Neotropical bird (Xenops minutus) reveals divergence, population expansion, and gene flow. Mol. Phylogenet. Evol. 83:305-316.

Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239-257.

Heath T.A., Zwickl D.J., Kim J., Hillis D.M. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. 57:160-166.

Hillis D.M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47:3-8.

Hijmans R. 2017. geosphere: Spherical Trigonometry (version 1.5.7). R package. https://cran.r-project.org/web/packages/geosphere/index.html

Hohenegger J. 2014. Species as the basic units in evolution and biodiversity: Recognition of species in the Recent and geological past as exemplified by larger foraminifera. Gondwana Res. 25:707-728.

Hotaling S., Foley M.E., Lawrence N.M., Bocanegra J., Blanco M.B., Rasoloarison R, Kappeler P.M., Barrett M.A., Yoder A.D., Weisrock D.W. 2016. Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. Mol. Ecol. 25:2029-2045.

Hull D.L. 1977. The ontological status of species as evolutionary units. In: Foundational Problems in the special sciences. Springer, Dordrecht. p. 91-102

Jakobsson M., Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Johnson N.K. 1980. Character Variation and Evolution of Sibling Species in the Empidonax difficilis-flavescens Complex (Aves: Tyrannidae). Univ of California Press 112:1-151.

Johnson N.K., Marten J.A. 1988. Evolutionary genetics of flycatchers. II. Differentiation in the Empidonax difficilis complex. Auk 105:177-191.

Johnson N.K., Cicero C. 2002. The role of ecologic diversification in sibling speciation of Empidonax flycatchers (Tyrannidae): multigene evidence from mtDNA. Mol. Ecol. 11:2065-2081.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jones O.R., Purvis A., Quicke D.L. 2012. Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. Ecography 35:333-340.

Klicka J., Zink R.M. 1999. Pleistocene effects on North American songbird evolution. Proc R Soc Lond [Biol], 266:695-700.

Knowlton N. 1993. Sibling species in the sea. Annu. Rev. Ecol. Evol. Syst. 24:189-216.

Linck E., Battey C.J. 2018. Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. bioRxiv doi:10.1101/188623

Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R. 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63:534-542.

Long C., Kubatko L. 2018. The effect of gene flow on coalescent-based species-tree inference. Syst. Biol. 67:770-785.

Magurran A.E. 2013. Measuring biological diversity. Hoboken: John Wiley & Sons.

Mallet J.M. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294-299.

Mastretta-Yanes A., Moreno-Letelier A., Pinero D., Jorgensen T.H., Emerson B.C. 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J. Biogeogr. 42:1586-1600.

16

Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE), 2010 (pp. 1-8). Ieee.

McKay B.D., Zink R.M. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. Mol. Phylogenet. Evol. 53:647-650.

Papakostas S., Michaloudi E., Proios K., Brehm M., Verhage L., Rota J., Peña C., Stamou G., Pritchard V.L., Fontaneto D., Declerck S.A. 2016. Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a rotifer cryptic species complex. Syst. Biol. 65:508-524.

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., Hoekstra H.E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One 7:e37135.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Puechmaille S.J. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol. Ecol. Resour. 16:608-627.

Pearson K. 1901. LIII. On lines and planes of closest fit to systems of points in space. Philos. Mag. 2:559-572.

Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.L., Harshman J., Huddleston C.J., Kingston S., Marks B.D, Miglia K.J., Moore W.S., Sheldon F.H., Witt C.C., Yuri T., Braun E.L. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. Syst. Biol. 66:857-879.

Riemer K., Guralnick R.P., White E.P. 2018. No general relationship between mass and temperature in endothermic species. eLife 7:e27166

Rosenberg M.S., Kumar S. 2003. Taxon sampling, bioinformatics, and phylogenomics. Sys. Biol. 52: 119-124.

Rosenberg N.A., Mahajan S., Ramachandran S., Zhao C., Pritchard J.K., Feldman M.W. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 1:e70.

Rush A.C., Cannings R. J., Irwin D.E. 2009. Analysis of multilocus DNA reveals hybridization in a contact zone between Empidonax flycatchers. J. Avian Biol. 40:614-624.

Schwartz M.K., McKelvey K.S. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv. Genet. 10:441-452.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. 2014. Target Capture and Massively Parallel Sequencing of Ultraconserved Elements (UCEs) for Comparative Studies at Shallow Evolutionary Time Scales. Syst. Biol. 63:83-95.

Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D., Pérez-Emán J., Burney C.W., Xie X., Harvey M.H., Faircloth B.C., Glenn T.C., Derryberry E.P., Prejean J., Fields S., Brumfield R.T. 2014. The drivers of tropical speciation. Nature 515: 406-409.

Smith B.T., Seeholzer G.F., Harvey M.G., Cuervo A.M., Brumfield R.T. 2017. A latitudinal phylogeographic diversity gradient in birds. PLoS Biol. 15: e2001073

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stevens G.C. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. Am. Nat. 133:240-256.

Swofford D.L. 2002. PAUP. *Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates: Sunderland, Massachusetts.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2015. MEGA6: Molecular Evolutionary Geneteics Analysis Version 6.0. 2013. Mol. Biol. Evol. 30:2725-2729.

Traylor M.A. 1977. A classification of the tyrant flycatchers (Tyrannidae). Bull. Mus. Comp. Zool. 148:129-184.

van Els P., Spellman G.M., Smith B.T., Klicka J. 2014. Extensive gene flow characterizes the phylogeography of a North American migrant bird: Black-headed Grosbeak (Pheucticus melanocephalus). Mol. Phylogenet. Evol. 78:148-159.

Willig M.R., Kaufman D.M., Stevens R.D. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annu. Rev. Ecol. Evol. Syst. 34:273-309.

Wen D., Yu Y., Zhu J., Nakhleh L. 2018. Inferring Phylogenetic Networks Using PhyloNet. Syst. Biol. 67:735-740.

Zarza E., Faircloth B.C., Tsai W.L., Bryson R.W., Klicka J., McCormack J.E. 2016. Hidden histories of gene flow in highland birds revealed with genomic markers. Mol. Ecol. 25:5144-5157.

Zarza E., Connors E.M., Maley J.M., Tsai W.L.E., Heimes P., Kaplan M., McCormack J.E. 2018. Combining next-generation sequencing and mtDNA data to uncover cryptic lineages of Mexican highland frogs. bioRxiv 10.1101/153601.

Zink R.M., Johnson N.K. 1984. Evolutionary genetics of flycatchers. I. Sibling species in the genera Empidonax and Contopus. Syst. Biol. 33:205-216.

Zucker M.R., Harvey M.G., Oswald J.A., Cuervo A., Derryberry E., Brumfield R.T. 2016. The Mouse-colored Tyrannulet (Phaeomyias murina) is a species complex that includes the Cocos Flycatcher (Nesotriccus ridgwayi), an island form that underwent a population bottleneck. Mol. Phylogenet. Evol. 101:294-302.

Supplemental Material

Tables S1 and **S2** available via Dryad (link pending)
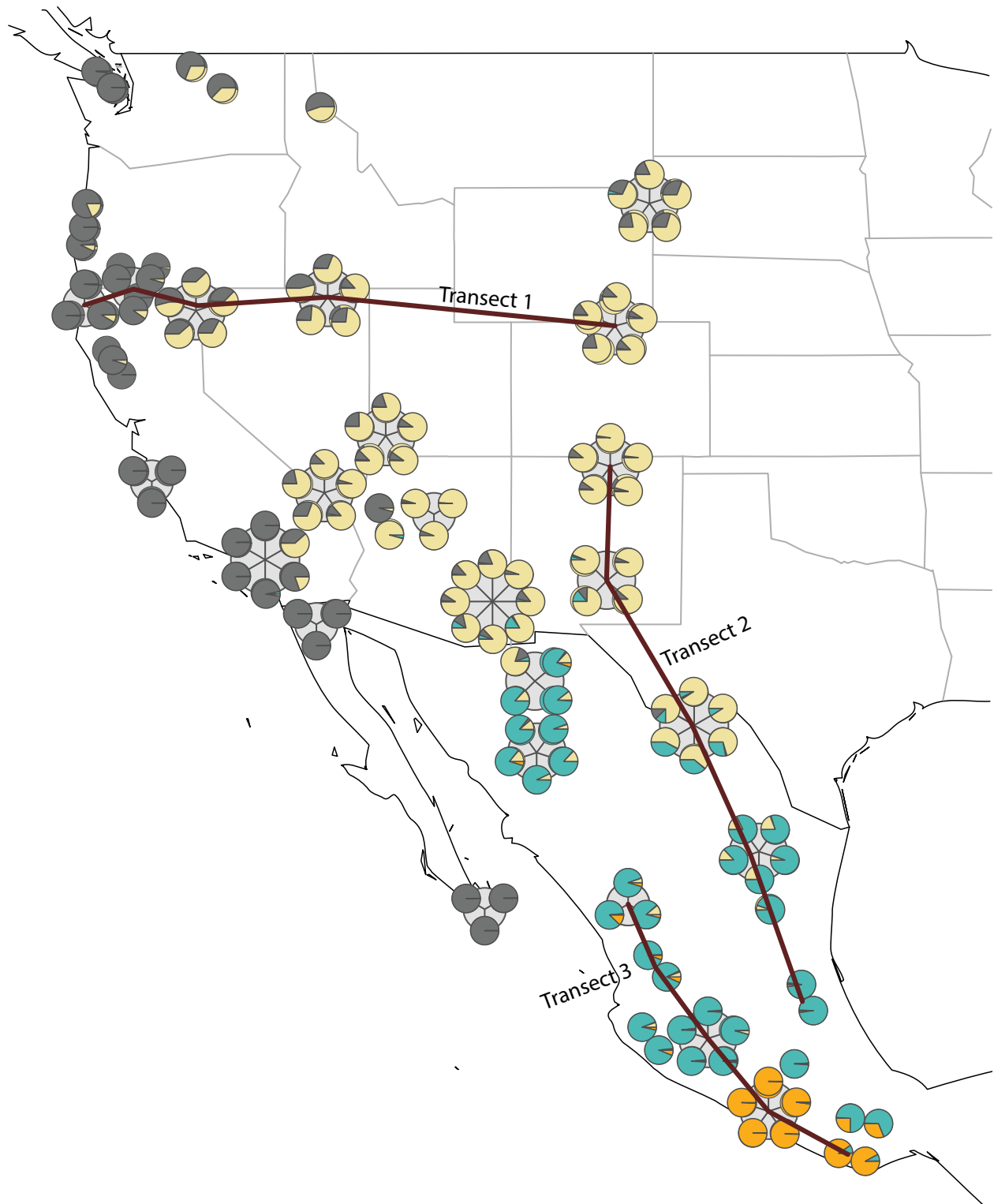
Figure 5: Transects for genomic cline analysis.

Figure 6: `RAxML` phylogeny from concatenated RAD loci and `structure` admixture proportions from RADseq SNP data.
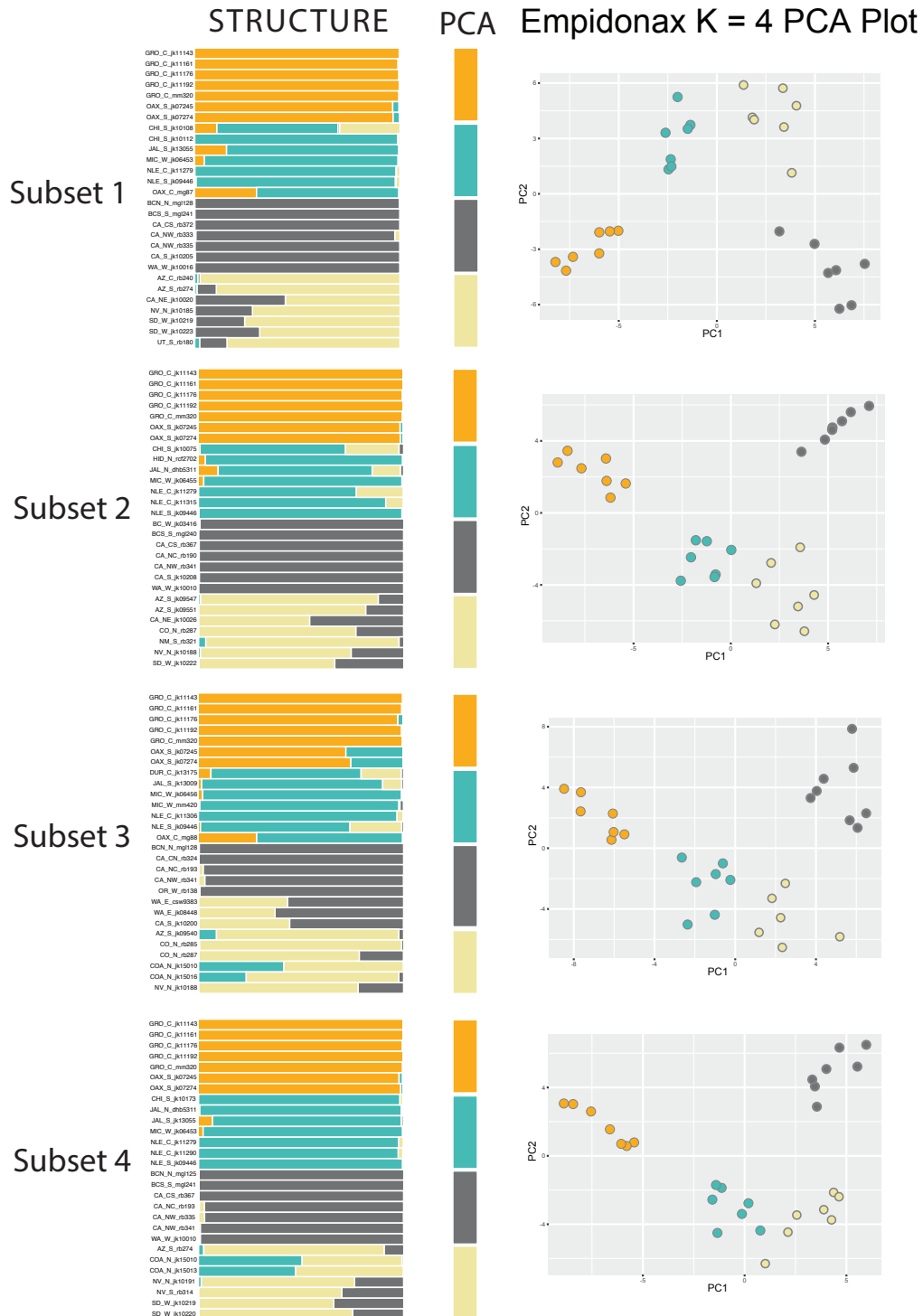
Figure 7: Population structure inference results from equal sample sized subsets, RADseq SNP data.

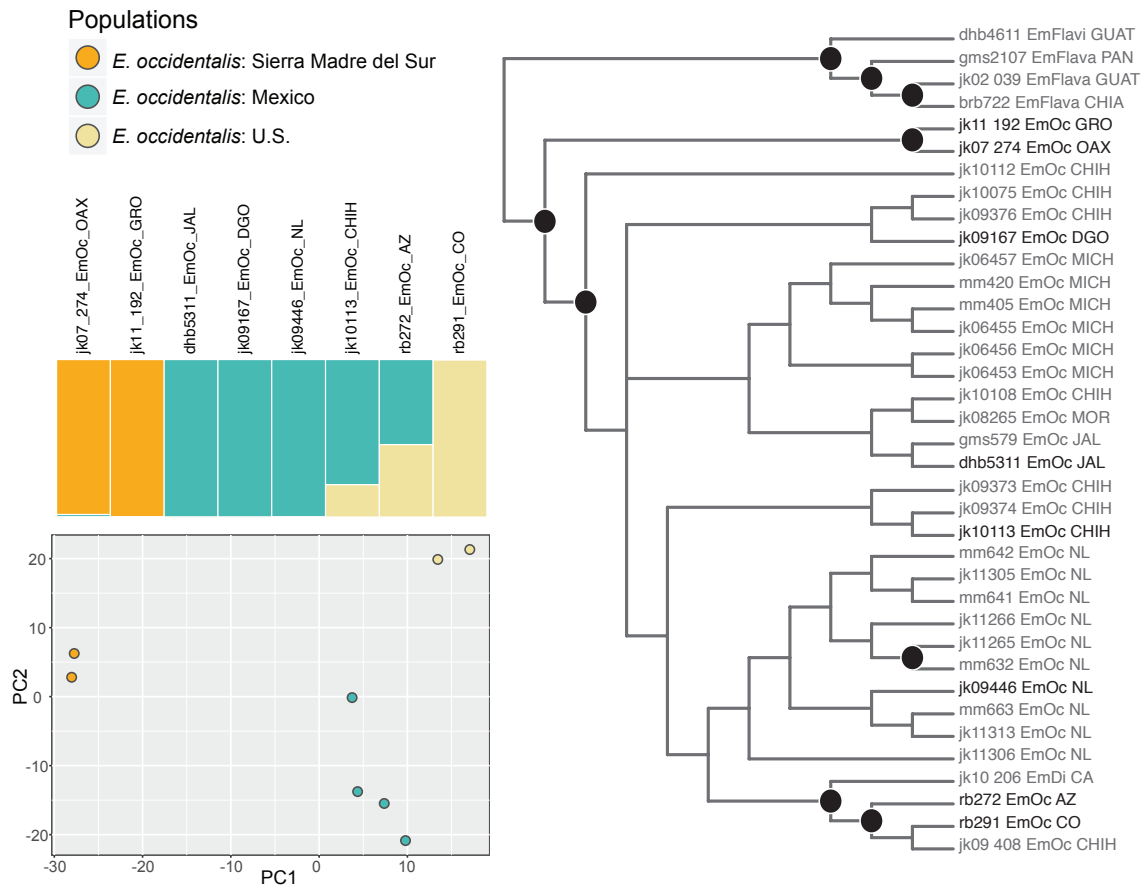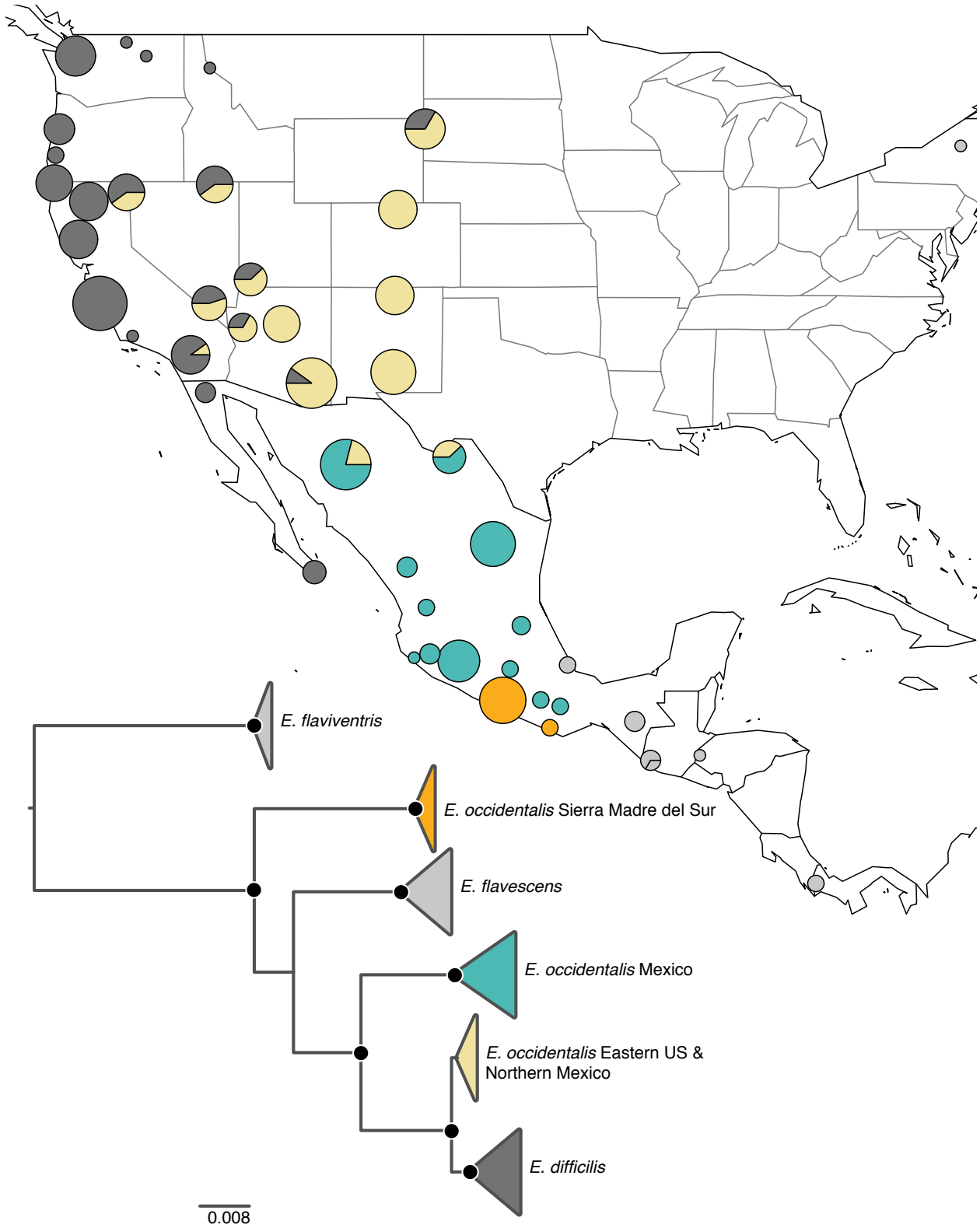Figure 8: UCE-based `SVDquartets` phylogeny and PCA

Figure 9: ND2-based RAxML phylogeny