**Inferring introgression using RADseq and *D*FOIL: power and pitfalls revealed in a case study of spiny lizards (*Sceloporus*)**

Running Title: *Inferring introgression with RADseq and DFOIL*

Shea M. Lambert1\*, Jeffrey W. Streicher1,2, M. Caitlin Fisher-Reid3, Fausto R. Méndez de la Cruz4, Norberto Martínez-Méndez5, Uri Omar García Vázquez6, Adrián Nieto Montes de Oca7, and John J. Wiens1

1. Department of Ecology and Evolutionary Biology, University of Arizona. Tucson, AZ 85721, USA.

2. Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK.

3. Department of Biological Sciences, Bridgewater State University. Bridgewater, MA 02324, USA.

4. Laboratorio de Herpetología, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México 04510, D.F., México.

5. Laboratorio de Bioconservación y Manejo, Departamento de Zoología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional, Ciudad de México 11340, D.F., México.

6. Unidad Multidisciplinaria de Investigación, Facultad de Estudios Superiores Zaragoza, Universidad Nacional Autónoma de México, Batalla 5 de Mayo s/n, Ejercito de Oriente, Iztapalapa, Ciudad de México 09230, D.F., México.

7. Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México 04510, D.F., México.

\*Corresponding author, email: [slambert1@email.arizona.edu](mailto:slambert1@email.arizona.edu)

**Abstract**

Introgression is now commonly reported in studies across the Tree of Life, aided by recent advancements in data collection and analysis. Nevertheless, researchers working with non-model species lacking reference genomes may be stymied by a mismatch between available resources and methodological demands. In this study, we demonstrate a fast and simple approach for inferring introgression using RADseq data, and apply it to a case study involving spiny lizards (*Sceloporus*) from northeastern México. First, we find evidence for recurrent mtDNA introgression between the two focal species based on patterns of mito-nuclear discordance. We then test for nuclear introgression by exhaustively applying the “five-taxon” *D*-statistic (*D*FOIL) to all relevant individuals sampled for RADseq data. In our case, this exhaustive approach (dubbed “Ex*D*FOIL”) entails testing up to ~250,000 unique four-taxon combinations of individuals across species. To facilitate use of this Ex*D*FOIL approach, we provide scripts for many relevant tasks, including the selection of appropriate four-taxon combinations, execution of *D*FOIL testsin parallel, and visualization of introgression results in phylogenetic and geographic space. Using Ex*D*FOIL, we find evidence for ancient introgression between the focal species. Furthermore, we reveal geographic variation in patterns of introgression that is consistent with patterns of mito-nuclear discordance and with recurrent introgression. Overall, our study demonstrates that the combination of *D*FOIL and RADseq data can effectively detect introgression under a variety of sampling conditions (for individuals, populations, and loci). Importantly, we also find evidence that batch-specific error and linkage in RADseq data may mislead inferences of introgressionunder certain conditions.

**Keywords:** *D*-statistics, ABBA-BABA, batch effects, hybridization, mito-nuclear discordance

**1 | INTRODUCTION**

Introgressive hybridization is increasingly recognized as a common and influential force of evolution (Mallet, 2005; Schwenk, Brede, & Streit, 2008; Harrison & Larson, 2014). The availability of genomic datasets and new methods for identifying introgression has facilitated a wide variety of new research in this area (reviewed in Twyford & Ennos, 2012; Payseur & Rieseberg, 2016). Nevertheless, it remains difficult to describe the timing, strength, directionality, genomic patterns, and geographic context of introgression, especially for taxa lacking suitable reference genomes.

A variety of methods for inferring introgression using genomic data are now available (reviewed in Sousa & Hey, 2013; Payseur & Rieseberg, 2016). In theory, many of these methods can be applied to “reduced-representation” datasets (e.g., RADseq, genotyping-by-sequencing) without needing a reference genome. The lack of a reference genome remains a common scenario, especially for the many researchers studying introgression in non-model organisms. However, in these cases, mismatches between methodological demands and available data or resources may be more common. For example, many methods assume unlinked sites (e.g., isolation with migration models; Hey, 2010), leading researchers to discard all but a single variant per locus. Other methods rely on well-resolved gene trees from many independent loci (e.g., most phylogenetic species networks; Than, Ruths, & Nakhleh, 2008; Solis-Lemus & Ané, 2016), which can be difficult to obtain using the relatively short sequences characteristic of reduced-representation datasets (e.g., Eaton & Ree 2013). In other cases, methods require a pre-defined set of hypotheses or demographic models to compare (e.g., Gutenkunst, Hernandez, Williamson, & Bustamante, 2009; Cornuet et al., 2014). Such methods may also become computationally intractable as populations or hypotheses are added. Finally, many methods that are otherwise suitable may require that the number of populations and assignment of samples to these populations are known (e.g., *f*-statistics, Reich, Thangaraj, Patterson, Price & Singh, 2009). These tools also rely on estimates of allele frequencies that could be biased by the number of individuals sampled for each population and by the bioinformatic processing of the data.

*D*-statistics (i.e., ABBA-BABA; Green et al., 2010; Durand, Patterson, Reich, & Slatkin, 2011) and related methods offer a simple yet powerful framework for detecting introgression. These statistics typically require only a species-tree hypothesis and site-count patterns for a sufficient number of biallelic sites (e.g., Martin et al., 2013; Eaton & Ree, 2013), and are computationally efficient to calculate. Pease and Hahn (2015) described *D*FOIL, a system of *D*-statistics applicable to a symmetric four-taxon tree (four ingroup taxa [P1-P4], plus an optional outgroup [O]) that can detect introgression and potentially infer its direction. Using simulations, Pease and Hahn (2015) demonstrated that *D*FOIL has high power and low type-I error under a wide range of conditions. By combining the results of four *D*-statistic “components”, *D*FOIL can infer either “ancestral” or “intergroup” introgression. Each component is a four-taxon *D*-statistic comparing three of the ingroup taxa. Ancestral introgression is that between the ancestor of the younger pair of ingroup taxa (P1 and P2) and one other ingroup taxon (P3 or P4). However, the direction of ancestral introgression cannot be inferred with *D*FOIL. In contrast, intergroup introgression signatures can be inferred between any two ingroup taxa, and can estimate directionality. Thus far, *D*FOIL has typically been applied across linkage groups of whole genomes or exomes, to summarize the genomic “landscape” of introgression (e.g., Kumar et al., 2016; Schumer, Cui, Powell, Rosenthal, & Andolfatto, 2016; Sarver et al., 2017). In theory, *D*FOIL will also work with reduced-representation markers (e.g., RADseq), allowing for the recovery of a single “genome-wide” signature of introgression (Pease & Hahn, 2015). However, this particular combination of methods (*D*FOIL and RADseq) has rarely been used (but see Huang, 2016, and Eaton & Ree 2013 for a similar combination of approaches).

Here, we develop a novel application of *D*FOIL that involves exhaustively testing relevant four-taxon combinations of sampled individuals among species, and then summarizing the results over phylogenetic and geographic space. In the present study, this approach (referred to as “Ex*D*FOIL”, https://www.github.com/SheaML/ExDFOIL) is paired with sequence data from double-digest RADseq (“ddRAD” hereafter; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). However, it is well-suited for sequence data from any reduced-representation genomic dataset with multiple individuals per population or species. We demonstrate this approach in a case study involving two lizard species from northeastern México. We also provide scripts that allow this approach to be readily applied by other researchers.

We also address how batch-specific errors and the inclusion of singleton site-pattern counts from RADseq data might influence these analyses of introgression. Sequence error is known to affect *D*-statistics, particularly when it is uneven between samples or populations (Green et al., 2010; Durand et al., 2011). Batch-specific error can arise in RADseq or other genomic data from a wide variety of sources (Mastretta-Yanes et al., 2015), and might be expected to produce non-random patterns of sequence error when combining samples from different batches. However, the potential impact of RADseq batch effects on *D*-statistics (and related approaches) is unexplored with empirical data, to our knowledge. Therefore, we used our exhaustive analyses here to examine the influence of batch identity. Similarly, the inclusion of singleton-site pattern counts (e.g., ABAAA, “singleton counts” hereafter) in *D*FOIL is predicted to generate false positives when counts are highly uneven between samples (Pease & Hahn, 2015). However, this possibility also remains unexplored (both *in silico* and empirically). Therefore, we used our analyses here to empirically evaluate the impact of including singleton counts.

Our study system is a small group of spiny lizard species (genus *Sceloporus)* from eastern and central México. *Sceloporus* is a well-studied genus distributed from Canada to Central America containing ~100 species (Uetz, Freed, & Hošek, 2017). *Sceloporus* research includes many studies focused on hybridization and introgression (e.g., Hall & Selander, 1973; Sites, Davis, Hutchinson, Maurer, & Lara, 1993; Sites, Barton, & Reed, 1995; Leaché & Cole, 2007; Leaché, 2011, Leaché, Harris, Maliska, & Linkem, 2013, Grummer et al., 2015). The present study focuses on a subset of the *torquatus* species group, one of the youngest and most species-rich groups of *Sceloporus* (~17 nominal species, *sensu* Leaché, Banbury, Linkem, & Nieto Montes de Oca, 2016). To our knowledge, this group has not been the target of previous research on hybridization or introgression. The two focal species of this study are the small-bodied, desert-dwelling *S. ornatus* (panel A of Figure 1) and the large-bodied, alpine-dwelling *S. oberon* (s*ensu* Wiens, Reeder, & Nieto Montes de Oca, 1999; Figure 1F, G). These species were synonymized by Martínez-Méndez and Méndez-de la Cruz (2007) using mtDNA alone (although this was inconsistent with other analyses based on mtDNA; Wiens et al. 1999; Wiens, Kuczynski, Arif, & Reeder, 2010; Wiens, Kozak, & Silva, 2013). Here, using nuclear DNA, we find they are reciprocally monophyletic and not sister species.

In this study, we use this new Ex*D*FOIL approach to evaluate nuclear introgression between these *Sceloporus* species. We apply this approach to new RADseq data for the *torquatus* group, using two distinct strategies for *de novo* ddRAD assembly and variant filtering (but identical individual-level sampling for downstream analyses). One strategy is aimed at phylogeny estimation and divergence dating across the clade (“clade-wide” dataset hereafter). The other is aimed at maximizing power to detect introgression in the two targeted species (“targeted” dataset hereafter). We first develop the hypothesis of introgression from patterns of mito-nuclear discordance, based on new mtDNA data and the RADseq phylogeny from the clade-wide dataset. We then test for the expected genome-wide signatures of introgression, using both the clade-wide and targeted datasets. The results demonstrate that *D*FOIL can detect introgression with RADseq data under a wide range of sampling conditions for loci, individuals, and populations. Furthermore, the Ex*D*FOIL approach reveals intra-specific geographic variation in the degree of introgression. Finally, our study reveals that batch-specific error in RADseq data may mislead inferences of introgression under certain circumstances, and that inclusion of singleton counts may also be problematic.

**2 | MATERIALS AND METHODS**

**2.1 | Geographic and taxonomic sampling**

Our sampling focused on a subclade of the *torquatus* group formerly referred to as the *poinsettii* group (sensu Wiens et al., 2010). We emphasized widespread geographic sampling of *S. ornatus* and *S. oberon.* We used samples from Wiens et al. (1999), supplemented with additional fieldwork by NM, UOG, and SML. In order to resolve species-level phylogeny using ddRAD data, we also included representatives of seven other species of the *torquatus* group. These species were selected on the basis of their close relationships to the two focal species in previous phylogenetic studies (e.g., Wiens et al., 2013; Leache et al., 2016) and tissue availability. Many *S. minor* individuals were sampled, as this species will be studied in future work. Vouchers, taxonomic identities, and georeferenced locality information for all samples are provided in Table S2.

There are two parapatric forms of *S. oberon* (Wiens et al., 1999). We refer to them as “*oberon*-black” and “*oberon*-red” based on their distinctive male dorsal coloration (Figure 1, panels F and G, respectively). We excluded ddRAD and mtDNA data from putatively hybrid individuals from the contact zone separating them (see sampling gap in Figure 1). This zone will be examined in future work.

**2.2 | mtDNA data collection**

To better examine mito-nuclear discordance, we supplemented previously published data for the ND4 mtDNA region (Wiens et al., 1999; Martínez-Méndez & Méndez-de la Cruz, 2007) with new ND4 data. Given that *S. ornatus* was represented by two or fewer individuals in previous studies, we added multiple individuals and localities for this species. Clade-wide sampling of the *torquatus* group (and outgroups) relied primarily on sequences from Wiens et al. (1999) and Martínez-Méndez and Méndez-de la Cruz (2007). GenBank accession numbers and taxonomic identities for all sequences used are in Table S1. We focused on ND4 because it contains more informative sites than other mtDNA regions used in this group (e.g., 12S, see Table 3 of Wiens et al., 1999).

For most samples, we extracted genomic DNA (gDNA) from liver or tail tissue using Qiagen DNeasy Blood and Tissue kit and mini-spin column protocol. For others, we used an alternative paramagnetic bead protocol (M. Fujita, pers. comm.) using “Serapure” beads prepared following the protocol of Faircloth & Glenn (2014), modified from Rohland & Reich (2012). To amplify ND4 and adjacent tRNAs, we used the primers ND4 (5’-TGACTACCAAAAGCTCATGTAGAAGC-3’) and LUE (5’-TRCTTTTACTTGGATTTGCACCA-3’), from Forstner et al. (1995). Each reaction totaled 25µL, using 12.5µL of GoTaq® Green Master Mix (Promega), 1–2µL of template DNA, either 4µL at 3µM concentration for each primer, or 1µL of 10µM, and nuclease-free water to 25µL. Reaction conditions generally followed Wiens et al. (1999), but using 35 cycles. Sequencing was performed at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility and the University of Arizona Genetics Core using Applied Biosystems 3730 DNA Analyzers.

**2.3 | ddRAD data collection**

We prepared ddRAD libraries generally following Peterson et al. (2012), with departures described here and in Appendix S1. We used the enzymes SbfI (5'-CCTGCA\*GG -3') and MspI (5'-C\*CGG-3', New England Biolabs) for restriction digest. We extracted gDNA as described above. We used Serapure beads for paramagnetic bead clean-up steps. For each sample, we used 250–500ng of starting DNA. We used a combinatorial barcode scheme to label individuals, with one barcode sequence on the 5’ end of the ligated adapter, and another on the 5’ end of the PCR primer. PCR amplification was performed on pooled libraries of 5’ barcoded samples. Full barcoding schemes and sequences are provided in Table S2.

We prepared two separate “batches” of ddRAD libraries, sequenced at different times and sequencing facilities (hereafter “batch 1” and “batch 2”). Although the ddRAD protocol was similar in both cases, there were several technical differences in library preparation and sequencing. Full details of library preparation and sequencing for both batches are described in Appendix S1.

Differences in library preparation and sequencing for our batches are expected to cause batch-specific error (or “batch effects”; for specific sources of error see Table 1 of Mastretta-Yanes et al., 2015). With this in mind, we included 20 biological replicates (i.e., separate aliquots from the same gDNA stock included in each batch; Table S2). This allowed us to estimate (and optimize) error rates and (together with replicates at the population level) mitigate batch-effects bioinformatically (see Appendix S1). In general, we tried to include samples from the same localities (or regions) in both batches to minimize introduction of real biological differences between the batches, which could be confounded with batch-specific error. However, samples of *S. ornatus* from the four western-most localities are restricted to batch 2. Additionally, several newly sampled localities for *oberon*-black and *oberon*-red were included in batch 2 only. Nevertheless, these sampling differences do not seem to explain any of our results (see sections **3.5** and **4.4**).

**2.4 | ddRAD de-novo assembly and variant calling**

We demultiplexed data using the function process\_radtags in Stacks v1.42 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). We discarded reads for which the expected cut sites and/or barcodes were not found. We filtered PCR duplicates from our data using the function clone\_filter function in Stacks, specifying the use of 8bp unique molecular identifiers located upstream of the barcode on the 5’ adapter end. We filtered data for adapters and low-quality sequence using the quality-filtering step of dDocent v2.24 (Puritz et al., 2014), which uses Trimmomatic v0.3.33 (Bolger, Lohse, & Usadel, 2014). We ran this step before all subsequent assembly, read mapping, or variant-calling steps.

We used two executions of the dDocent pipeline for *de novo* assembly, read mapping, and variant calling. The resulting datasets (referred to as “targeted” and “clade-wide”) differed primarily in the samples used to create the *de novo* reference and in the parameters used to filter variants. Importantly, both datasets ultimately include the same set of individuals for use in *D*FOIL tests. We created a single *de novo* reference for each dataset, using data from multiples species. The reference samples for the clade-wide assembly included eleven total representatives (Table S2) from six species. These samples were selected on the basis of phylogenetic breadth and sequencing depth, and include representatives of *S. torquatus* (*n*=1), *S. poinsettii* (*n*=1), *S. ornatus* (*n*=2), *S. oberon* (*n*=2), *S. cyanogenys* (*n*=1), *S. cyanostictus* (*n*=1), and *S. minor* (*n*=3). The assembly used a minimum-read count (k1) of 3 and minimum-individual count (k2) of 4. The reference samples for the targeted assembly included all available individuals of *S. oberon* and *S. ornatus* from batch 2, and used values of k1=3 and k2=3. As we did not include individuals of *S. cyanogenys* or *S. cyanostictus* in the targeted reference panel, an increased rate of mapping error may occur for these taxa. In theory, this could lead to inflated type-I error rates for tests involving these taxa (Durand et al., 2011). However, because relatively few individuals are available for these taxa, this bias would likely persist even if the samples were included in the reference panel. Consequently, we interpreted with caution those tests using the targeted dataset and one individual of *S. cyanostictus* or *S. cyanogenys* (see section **3.5**).

For both *de novo* assemblies, we mapped reads using BWA v0.7.15-r1142-dirty (Li & Durbin, 2010), with a match score of 1, mismatch score of 3, and gap score of 5. We called variants using the population-informed model of freebayes v1.0.2-33-gdbb6160 (Garrison & Marth, 2012), with populations defined by sampling locality (Table S2). For the clade-wide assembly, we mapped reads and called variants for all available individuals, including those from batch 1. These individuals had better overall sequencing depth but shorter read length (Table S2). However, these individuals were not included in either reference panel, as we observed that “mixed” reference panels resulted in higher error rates for biological replicates. For the targeted assembly for DFOIL analyses, we called variants for all available individuals (including both batches) of *S. oberon*, *S. ornatus*, *S. cyanogenys,* and *S. cyanostictus.* Based on our mtDNA analyses (Figure 2, Fig S1), one sample of *S. minor* was also included to use as the outgroup for our *D*FOIL analyses. We selected an individual of *S. minor* rather than a more distantly related taxon to maximize the number of comparable sites for analysis.

We applied distinct variant-filtering pipelines to the raw clade-wide variants and targeted variants (contained in the TotalRawSNPs.vcf files produced by dDocent). The clade-wide variants underwent relatively stringent filtering, which greatly reduced between-batch error rates (described in Appendix S1). In contrast, the targeted variants underwent relatively minimal filtering. Using vcftools v0.1.15 (Danacek et al., 2011), we filtered all sites with quality scores <30, and all genotypes with coverage <3. We then filtered any variants that were not bi-allelic SNPs or that had missing data for >99% of individuals. We converted the resulting vcf file to fasta format using the vcf\_to\_tab function of vcftools and a publicly available perl script by C.M. Bergey (https://code.google.com/archive/p/vcf-tab-to-fasta/). To characterize the effect of sampling additional individuals, we further split the data into “reduced” (Figure 2) and “full” (Fig S1) sets of individuals. For reduced sampling, we kept a small number of the best-sequenced individuals from each sampled population. For full sampling, we used these individuals plus any remaining individuals of *S. cyanostictus*, *S. cyanogenys*, *S. oberon,* or *S. ornatus* that had <50% missing sites. This threshold was applied to a minimally filtered version of the clade-wide TotalRawSNPs.vcf file (see Appendix S1 for filtering parameters).

**2.5 | Phylogenetic analyses and divergence dating**

For mtDNA data, we aligned sequences using the CLUSTAL W (Thompson, Higgins, & Gibson, 1994) plugin in Geneious v6 (Kearse et al., 2012). We collapsed identical haplotypes using the “Find Duplicates” function of Geneious. Collapsed sequences are indicated in Table S1. We manually inspected the protein-coding section of the ND4 alignment to ensure open reading frames for all sequences. In one case, we replaced an apparent stop-codon early in the protein-coding sequence of an outgroup sample (*S. jarrovii*, #AF154210) with three ambiguous base pairs (“NNN”).

We estimated phylogenetic relationships and divergence times with the ND4 alignment using a relaxed-clock Bayesian framework with BEAST2 v2.4.6 (Bouckaert et al., 2014), executed using the CIPRES Scientific Gateway v3.3 (Miller, Pfeiffer, & Schwartz, 2010). We used bModelTest (Bouckaert & Drummond, 2017) to infer and marginalize models of substitution and rate heterogeneity. As bModelTest has only been tested with sequences as short as 500bp, we chose not to use separate partitions for each codon position and the tRNA region, which would all be <250bp. Instead we treated the entire sequence as a single partition (but using models that incorporated rate heterogeneity among sites). Given the scarcity of fossils within the *torquatus* group, we used one secondary calibration point with a range of dates from two recent studies (Bayesian relaxed-clock trees of Wiens et al., 2013; Leaché et al., 2016). We calibrated the node corresponding to the most recent common ancestor of *S. grammicus* and the *torquatus* group, the root of the tree in the present study. We used a uniform prior from 12.9–18.0 million years ago (Mya), bracketed using these two point estimates for this node. We selected a single uncorrelated lognormal relaxed clock model (Drummond, Ho, Phillips, & Rambaut, 2006), and a calibrated Yule-tree prior (Heled & Drummond, 2012). We ran 40 million generations, sampling once every 10,000 generations, and ran three replicate analyses. We assessed convergence using Tracer v1.6 (Rambaut, Suchard, Xie, & Drummond, 2014). We first checked that all parameter and prior estimates had effective sample sizes (ESS) >200 for each replicate analysis. We then compared replicate analyses, ensuring that each provided similar estimates for all parameters. Using the results of one analysis, we generated a maximum clade-credibility tree using common ancestor heights, after discarding the first 10% of generations as burn-in, using TreeAnnotator v.2.4.5.

For phylogenetic analyses of ddRAD data, we used RAxML v8.2.9 (Stamatakis, 2014) on the University of Arizona High Performance Computing (UA HPC) cluster. We treated each concatenated matrix as one partition, using the GTRCAT model of evolution. We used the -f a option to search for an optimal tree and to assess support using 100 rapid bootstrap replicates. We conducted two analyses with the clade-wide alignment, one with full- and one with reduced-individual sampling. We did not include biological replicate sample pairs in phylogenetic analyses. Instead, we retained the replicate with more overall sequence data. Several samples were excluded from these analyses after being identified as having strongly discordant signal, putatively owing to introgression. Removal of these samples allowed for consistent placement of *S. cyanogenys,* but had no effect on placement of *S. cyanostictus*, *S. oberon,* or *S. ornatus* (see Appendix S1 for details).

We did not use the targeted dataset for phylogeny estimation, for two major reasons. First, invariant sites were not retained. Their exclusion can negatively influence topological and branch-length inference (Lewis, 2001; Bertels, Silander, Pachkov, Rainey, & van Nimwegen, 2014), even when corrections for acquisition bias are applied (Leaché, Banbury, Felsenstein, Nieto Montes de Oca, & Stamatakis, 2015). Second, only *S. oberon* and *S. ornatus* individuals were used to create the *de novo* reference assembly. This should bias the retained loci towards those with variants within or between these particular species. This is useful for detecting introgression between these species using *D*FOIL, but not for estimating branch-lengths or topologies.

For divergence dating using ddRAD data, we used penalized likelihood (Sanderson, 2002) implemented using treePL v1.0 (Smith & O'Meara, 2012). This approach estimates divergence times based on an existing topology and set of branch lengths (here from the concatenated likelihood analysis). Use of BEAST to simultaneously estimate the topology and divergence times was not practical for the ddRADseq data, given the large number of loci. We used a fixed-point calibration on the root, representing the crown-node of the *torquatus* group (sensu Leaché et al., 2016). We set this age to 11.8 Mya, based on the Bayesian estimate from Leaché et al. (2016). We used the leave-one-out cross-validation procedure (Sanderson, 2002) to choose the optimal smoothing parameter (10), comparing 6 smoothing values between 0.1–10,000 in tenfold increments.

**2.6 | Exhaustive application of DFOIL (‘Ex*D*FOIL’)**

Our approach to assessing introgression with *D*FOIL involved applying this test exhaustively to all sets of individuals that matched the assumptions of *D*FOIL. To do this, we first wrote a custom function in R v3.4.1 (R Core Team, 2017) that accepts a phylogenetic tree and a list of taxa, and returns all unique sets of taxa {P1, P2, P3, P4} such that subsets {P1,P2} and {P3,P4} are reciprocally monophyletic, and the most recent common ancestor of subset {P1,P2} is younger than that of subset {P3,P4}. The latter assumption is met by all pairs of clades with non-identical clade age, but *D*FOIL expects that the younger clade is listed first. In cases of identical clade ages, the function will still accept the combination as valid and return the clades in an arbitrary order. However, no such cases exist for our tree. This function depends on the R packages *ape* (Paradis, Claude, & Strimmer, 2004), *phytools* (Revell, 2012), and *combinat* (Chasalow, 2012), and is available as Supplementary File S1. We applied this function using each of our two ddRAD-based phylogenetic hypotheses (reduced-individual sampling, Figure 2; and full-individual sampling, Fig S1). We then filtered these sets of taxa, retaining only sets with representatives of both *S. oberon* and *S. ornatus,* resulting in 32,368 unique sets (reduced) and 237,600 (full) unique sets. For every test, we used the same outgroup, a sample of *S. minor* from batch 2 with highly complete data (voucher EPR743, Table S2). We used the default significance cutoff of 0.01 for each *D*FOIL component, and defaults for all other settings.

We ran the *D*FOIL pipeline on each list of unique sets, for both the clade-wide and targeted datasets, on the UA HPC cluster. We executed the *D*FOIL pipeline using custom shell scripts and a publicly available perl script (selectSeqs.pl; <http://raven.iab.alaska.edu/~ntakebay/teaching/programming/perl-scripts/selectSeqs.pl>), parallelized across 28 processors with GNU parallel (Tange, 2011). We then used custom shell and R scripts to collate the *D*FOIL test results, associate sample information with each test result, calculate summary statistics, and visualize results. Example scripts and input files for these steps are available in the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL.

For our primary analyses, we used the dfoilalt option of dfoil.py to ignore singleton-site pattern counts (e.g., ABAAA, “singleton counts” hereafter). These counts can be included in DFOIL to reflect the fact that introgression could transfer either the derived or ancestral allele (Pease & Hahn 2015). Their inclusion is predicted to be potentially problematic when error or substitution rates are skewed (Pease & Hahn, 2015), but this prediction has not been formally tested. Therefore, we repeated all tests while including singleton counts using the default dfoil flag (see section **2.8**).

*D*FOIL uses Chi-square tests of significance, rather than Z-scores from bootstrap or block-jackknife resampling, as in previous applications of *D*-statistics (e.g., Green et al. 2010; Eaton & Ree, 2013). An assumption of the Chi-square test is that the sites considered are independent. This assumption is clearly violated in most data sets (including our own). However, using simulation, Pease and Hahn (2015) showed that D-statistics will nevertheless approximate a Chi-square distribution when a sufficient number of sites are considered. They concluded that the number of sites required can be estimated using the parameter *ρ* = 4*N*e*rL*, with *Ne*=effective population size, *r*=recombination rate, *L*=the number of sites considered, and a minimum value of *ρ* of ~4000. Using values of *Ne*=106 and *r*=10-8, Pease and Hahn (2015) showed that the number of sites required is ~100kb. Although we lack formal estimates of effective population sizes or recombination rates for our focal taxa, we believe that we are likely sampling sufficient sites to avoid serious issues. Our clade-wide dataset samples from >400kb, and our targeted dataset samples from >5Mb. Moreover, while a recombination rate of 10-8 may be reasonable for sites within a single genomic window, the per-site recombination rate in our ddRAD dataset should be much higher, as ddRAD loci are scattered across the genome. Given these considerations, we do not believe that linkage within or between our RAD loci is likely to inflate our type-I error rate, particularly for the targeted dataset. Nevertheless, we lack formal estimates of effective population size or recombination rate for our data. To more directly address the possible effects of linkage on our inferences, we also conducted Ex*D*FOIL analyses using a single site per locus (to remove the effects of linkage within loci) and bootstrapped datasets (randomly resampling loci, to reduce the effects of linkage between loci). A full description of the methods for these analyses is found in Appendix S2.

**2.7 | Examination of batch-specific effects**

To examine the influence of batch identity, we first compared proportions of positive tests for introgression for each of the four primary datasets (using Supplemental Tables S3–S6). We did this for tests that included: (1) only batch 2 samples for the four ingroup individuals (P1–P4); (2) only batch 1 samples for these individuals; and (3) any combination of batch 1 and batch 2 samples for ingroup individuals. We then used stacked bar plots to visualize test results for our targeted dataset with reduced-individual sampling, grouped by “batch signature.” We defined the batch signature as a string giving the batch of origin for each of the four ingroup individuals (P1–P4). For example, we used “1122” for a test with P1 and P2 from batch 1 and P3 and P4 from batch 2.

**2.8 | Examination of singleton-count effects**

Pease and Hahn (2015) predicted that *D*FOIL may return false positives when including singleton counts if sample-specific error, or substitution rate, is high enough. The predicted mechanism for these false positives is an inflated distance from all other taxa for a taxon that has a relatively high error (or substitution rate) leading to false inference of introgression for that taxon’s sister lineage.

For this reason, *D*FOIL warns if either of the singleton count ratios P1:P2 or P3:P4 is >1.25 or <0.75 for a given test.

To examine the influence of including singleton counts on our *D*FOIL results, we first compared proportions of positive tests within each dataset, both for tests excluding singleton counts (executed with dfoilalt flag) and tests including them (using dfoil flag). We also used a grouped bar plot to examine ratios of singleton counts in P3 to singleton counts in P4 (using the test results from our targeted dataset with reduced-individual sampling; Figure 6). This plot compares the average P3:P4 ratio for tests with singleton counts excluded vs. included, across each possible *D*FOIL result. Highly uneven ratios may cause erroneous inference of introgression for the taxon with a lower singleton count (Pease & Hahn, 2015). We chose to examine the P3:P4 ratio because introgression signatures returned by *D*FOIL can involve only one or the other of these taxa, allowing for a clearer examination of the relationship between singleton count and involvement in introgression.

**2.9 | Comparison of Ex*D*FOIL and TreeMix**

To compare Ex*D*FOIL with an alternative methodology, we conducted analyses using the program TreeMix (Pickrell & Pritchard, 2012) and our targeted dataset with full-individual sampling. TreeMix creates population networks in a two-step process, where a population tree is first inferred and migration edges are subsequently added to populations that do not fit the tree model well. A detailed description of our TreeMix methods is found in Appendix S3.

**3 | RESULTS**

**3.1 | mtDNA alignment and phylogenetic results**

The final ND4 alignment contained 122 unique haplotypes, 915 total sites, and 310 parsimony-informative sites. We recovered a mostly well-resolved topology (Figure 2) for our species of interest (i.e., those sequenced for ddRAD): ((*cyanostictus*+*cyanogenys*),(*minor*(*oberon+ornatus*)). As we were unable to successfully sequence samples of *S. serrifer* for ddRAD, we do not focus on *S. serrifer* here. However, in agreement with Martínez-Méndez and Méndez-de la Cruz (2007), we find that *S. serrifer* is polyphyletic, with one clade related to *S. minor* and another to *S. cyanogenys*. *Sceloporus cyanogenys* and *S. cyanostictus* were each strongly supported as monophyletic, with posterior probabilities (PP) of 1.00. *Sceloporus minor* was supported as monophyletic with PP=0.89, and weakly supported (PP=0.72) as the sister group to a well-supported clade (PP=0.92) containing all *S. oberon* and *S. ornatus* samples. This clade is comprised of two strongly supported subclades (PP=1.00). One contains only *S. ornatus* samples, including five individuals from localities near the contact zone with *S. oberon* (see Figure 1) and all samples from localities further west (see Figure 4). The other contains all *S. oberon* individuals, and all remaining *S. ornatus* individuals from localities near the contact zone. Within this “mixed” clade, *oberon*-red is strongly supported as monophyletic (PP=1.00), but *oberon*-black and *S. ornatus* are paraphyletic with respect to each other (with strong support for relevant nodes).

**3.2 | ddRAD sequencing, assembly, and alignment results**

Read counts for all samples are in Table S2 (batch 1: mean=1,145,317, range=11,973–3,712,752, standard deviation=814,844; batch 2: mean=983,548, range=1,546–6,765,507, standard deviation=1,306,692). The clade-wide reference assembly totaled 3,728 loci and 922,077 sites. After all filtering steps (see Appendix S1) we retained 1,712 loci and 404,966 sites. The clade-wide alignments used for phylogenetic analysis contained 2,950 (reduced individual) and 3,389 (full individual) parsimony-informative sites. For full and reduced-individual sampling of clade-wide variants, we retained 10,095 biallelic sites, including heterozygous sites (*D*FOIL derives site-pattern counts from biallelic sites with fixed differences only). The targeted reference assembly totaled 22,433 loci and 5,538,162 sites. After filtering based on quality and depth, we retained 205,101 biallelic sites from 21,818 loci, for both full and reduced-individual sampling. We did not retain any loci that were entirely invariant across all individuals for either dataset.

**3.3 | ddRAD phylogenetic results**

Using our clade-wide ddRAD dataset with reduced individual sampling, we recovered a strongly supported species-level hypothesis: (*minor*(*oberon*(*cyanogenys*(*cyanostictus*+*ornatus*)))). Bootstrap support was >90% for each inter-specific split (Figure 2). All nominal species were strongly supported as monophyletic, excepting *S. minor* (monophyletic in the best tree, but with bootstrap <50%) and *S. cyanogenys* (only one individual included). When the full set of individuals was used in tree estimation, bootstrap support for the placement of *S. cyanogenys* fell to 83% (Figure S1). However, the species-level topology remained identical and otherwise strongly supported (Fig S1). Notably, the topology from nuclear data shares none of the four interspecific splits from the mtDNA tree for the same five species (Figure 2).

**3.4 | Exhaustive DFOIL results**

Summaries for all *D*FOIL test results using the dfoilalt method (excluding singleton counts) are provided in Tables S3–S6. We found that dfoilalt tests including two representatives of *oberon-*red resulted in very few introgression signatures (~1% of such tests for the targeted dataset with reduced-individual sampling, calculated from Table S3). This is consistent with the monophyly of *oberon-*red in our mtDNA dataset (Figure 2). Furthermore, we found very few introgression signatures in tests including any representative of *S. cyanostictus* or *S. cyanogenys* using the dfoilalt method (<2% of such tests for the targeted dataset with reduced-individual sampling, calculated from Table S3). Therefore, in Tables 1–3, we report proportions of positive tests considering only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these being *oberon*-black). For each dataset, we provide the percentage of positive tests for ancestral and inter-group introgression in Table 1, raw numbers for each unique *D*FOIL signature in Table 2, and percentage of introgression signatures categorized by the species involved in Table 3, and the median, range, and standard deviation of non-singleton site-pattern counts in Table 4.

Overall, introgression was inferred much more frequently using the targeted datasets than the clade-wide datasets (39.4% vs 1.6% for reduced-individual sampling; Table 1), but this is unsurprising given that the targeted dataset had roughly fourfold more site-pattern counts per test (Table 4). Most positive results were for signatures of ancestral introgression, particularly involving the ancestors of pairs of *S. ornatus* (as P1 and P2) and *oberon*-black. Intergroup signatures (between terminal taxa) were recovered less frequently, and most often indicated introgression from *S. ornatus* into *S. oberon*, opposite the direction inferred from mtDNA (Table 3). However, we show in the next section that these intergroup signatures may be spurious and attributable to batch effects.

Phylogenetic and geographic patterns of introgression for the targeted dataset with reduced individual sampling are visualized in Figures 3 and 4, respectively. We stress that the proportions provided merely indicate the proportion of positive tests, and do not provide information on (for example) the quantity of introgression. These visualizations demonstrate that introgression signatures involving *S. ornatus* and *oberon*-black were detected in similar proportions under multiple phylogenetic and geographic contexts, as expected if introgression was ancient. However, tests involving the western-most populations of *S. ornatus* recovered introgression slightly less often, potentially indicating recurrent introgression involving the easternmost populations of *S. ornatus*.

We expected that for our targeted dataset, individuals of *S. cyanostictus* and *S. cyanogenys* could have increased rates of mapping error relative to individuals of *S. ornatus* and *S. oberon*, as they were not used in the reference panel (see section **2.4**). A bias in rates of mapping error could theoretically lead to false positives for *D*-statistics (Durand et al., 2011). However, we observed the opposite pattern, with tests involving *S. cyanostictus* or *S. cyanogenys* recovering very few positive results. This result could be caused by an increased rate of allelic dropout (or failure to map) for these taxa. Consistent with this idea, tests involving *S. cyanostictus* or *S. cyanogenys* had on average fewer counts than tests without these taxa. For the targeted dataset with reduced-individual sampling, tests involving these taxa had a median of 315 non-singleton counts, while tests without these taxa had a median of 505 non-singleton counts (calculated using Table S7).

The results of our single-site-per-locus analyses and locus-level bootstrapping analyses, which were designed to address the potential effects of linkage on our inferences using *D*FOIL, are summarized in Table S11 and Figures S6 and S7. Overall, we recovered introgression in ~85% of these resampled datasets relative to the full dataset. This pattern suggests that introgression may be somewhat overestimated due to linkage, but that the overall effect is relatively small and should not overturn our conclusions. More detailed reporting and discussion of these results are found in Appendix S2, and complete results can be found at https://www.github.com/SheaML/ExDFOIL/AppendixS2\_Materials.

**3.5 | Batch-specific effects**

Results for each dataset separated by batch identity (all batch 1 samples, all batch 2 samples, or a mix) are presented in Table 5. We focus here on the targeted dataset with reduced-individual sampling. For ancestral introgression signatures (Table 5), the proportion of positive tests was much higher for tests using only batch 2 samples (70.6%) than tests using only batch 1 samples (38.8%), or a combination of batch 1 and 2 samples (35.0%). Intergroup signatures of introgression were also inferred in a larger proportion using only batch 2 samples (6.1%), compared to batch 1 (0.9%), or mixed samples (3.7%; Table 6). When applying the full-individual sampling to the targeted dataset, there was also a reduction in the proportion of positive tests when using batch 1 or mixed samples For the clade-wide dataset, batch 1 had a higher proportion of positive tests than batch 2, but the proportions were very low for both.

A visual comparison of the test results for each possible “batch signature” is provided in Figure 5, using the targeted dataset with reduced individual sampling. These results show several instances of over-representation of a particular introgression signature for one or more batch signatures. Most notably, the introgression signature “P1→P4” is over-represented in tests with a batch signature of “1211” or “1221” and the introgression signature “P2→P4” is over-represented in tests with a batch signature of “2111” or “2121”. In these cases, we argue that each of these patterns is the result of type I error induced by batch effects (see Discussion). Also evident in Figure 5 is the over-representation of “P1/P2←→P3” introgression in tests with a batch signature “1112”. In this case, however, the over-representation is more likely due to a sampling difference between the batches for the reduced-individual datasets (see Discussion).

The inclusion of some sampled populations in only a single batch may raise concerns that our results are confounded by batch effects. In the case of *S. oberon*, several localities were restricted to batch 2 only (oberonS1, oberonS3, oberonS4, oberonS6; Figure 4). However, there is no apparent pattern in the introgression results for these populations vs. populations included in both batches. In the case of *S. ornatus*, the four western-most localities were exclusively from batch 2. These localities also show a reduction in the percentage of tests returning introgression (Figure 4). However, this geographic pattern is still apparent when results from western and eastern *S. ornatus* localities are compared using only batch 2 (Figure S5). Given this result, we think this geographic pattern is not a product of batch effects.

**3.6 | Singleton-count effects**

For simplicity, results in this section are only from the targeted dataset with reduced-individual sampling (results for all datasets in Table 6). Importantly, we included all possible tests in Table 6 and Figure 6. The inclusion of singleton counts massively increased the proportion of positive tests involving representatives of *S. cyanostictus* or *S. cyanogenys*, or without a representative of *oberon*-black (Table 6).

The effects of including singleton counts were dramatic: 45.0% of tests returned introgression signatures when singleton counts were included and 15.9% when they were not (Table 6). We found many more signatures of directional introgression when including singleton counts (8.1% vs. 1.3%), and a much higher proportion of introgression signatures for tests that lacked representatives of *oberon*-black(43.8% vs. 1.1%).

Comparison of singleton-count ratios for taxa P3 and P4 across introgression signatures (Figure 6) showed that when singleton counts were included, average count ratios were often well beyond the limits suggested by the *D*FOIL authors (>0.75 and <1.25, indicated by dotted red lines), which may lead to type-I error (Pease & Hahn, 2015). In comparison, tests excluding these counts typically had ratios much closer to 1.

**3.7 | TreeMix Results**

Our TreeMix analysis inferred introgression between *oberon*-black and *S. ornatus* (Fig S8). Analyses including additional migration edges (up to 5) improved model fit with diminishing returns, and did not reveal further introgression events between these taxa. The inferred migration edge suggests that introgression occurred from within the *oberon*-black clade into the ancestors of *S. ornatus*, with fractional genomic contribution of *S. oberon* to *S. ornatus* of ~0.2. A detailed description of our TreeMix results and discussion of the relative advantages and disadvantages of TreeMix vs. Ex*D*FOIL are found in Appendix S3. Overall, the pattern of introgression inferred from TreeMix was concordant with that inferred by Ex*D*FOIL and the mtDNA data.

**4 | Discussion**

**4.1 | Advantages of combining ddRAD and Ex*D*FOIL**

The combination of ddRAD and the exhaustive *D*FOIL (Ex*D*FOIL) approach developed here for detecting introgression has several attractive properties. These include cost-efficiency and ease of execution for laboratory and computational aspects alike. *D*FOIL also allows for maximal use of data from ddRAD markers by concatenating all biallelic sites from all loci, as long as a sufficient number of sites are sampled for the D-statistics to approximate a Chi-square distribution (Pease & Hahn, 2015). Many popular methods have assumptions about linkage that are potentially problematic for ddRAD and similar data types. For instance, some methods assume that each variant is effectively unlinked (e.g., Pickrell & Pritchard, 2012), which could require the sub-sampling of only one variant per ddRAD locus. Other methods may assume that variants within each locus are completely linked, but loci are unlinked (e.g., Hey 2010). Either condition could be violated in ddRAD or similar datasets, and these assumptions can be difficult to assess, especially without a reference genome.

The Ex*D*FOIL approach is also free of several assumptions or intermediate steps required by “population-based” methods for inferring introgression (e.g., *f*-statistics: Reich et al., 2009). These methods require the number of populations, assignment of samples to these populations, and allele frequencies to be known. This makes the Ex*D*FOIL approach more readily applicable to datasets with highly uneven sampling between taxa and/or uncertainty about the number and composition of populations, as in our dataset. Moreover, by virtue of its exhaustive nature, Ex*D*FOIL provides a richer view of the overall signal and noise in the data, as compared to a relatively small number of population-based hypothesis tests.

Despite its exhaustive nature, Ex*D*FOIL remains computationally efficient, especially if multiple processors are available for use. The independent execution of a set of *D*FOIL tests is a readily parallelizable problem, with no communication required between processors. This means that computational time should decrease in a roughly linear manner with the number of available processors. This also means that the computational load of Ex*D*FOIL could easily be split across entirely independent machines, and the results later collated. The calculation of D-statistics scales easily with the number of sites. Therefore, the major source of computational load for most Ex*D*FOIL analyses will be the number of tests considered. This number will depend on the total number of taxa and the structure of the phylogenetic tree used. In our case, an increase in sampling of 25 individuals, from 36 (reduced-individual sampling) to 62 (full-individual sampling), generated 236,110 additional tests, or roughly 10^4 tests per individual added.

Our results show that the combination of ddRAD and *D*FOIL can recover signatures of introgression using only a small fraction of genome-wide variation. Given an estimate of 2.73pg for the genome size of a congener (*S. magister*; De Smet, 1981), and a conversion factor of 978 x 106 base-pairs per picogram (Doležel, Bartos, Voglmayr, & Greilhuber, 2003), our targeted dataset sampled only ~0.2% of the genome (~5.5 x 106 bp). We detected introgression between the expected taxa with as few as 69 non-singleton site-pattern counts for the clade-wide dataset, and 113 for the targeted dataset (Table 4). Furthermore, we were able to detect introgression even when including individual samples with large amounts of missing data (as in the full-individual sampling). For example, tests including a sample of *S. ornatus* with >75% missing data in the targeted dataset (NM170p1) still recovered introgression in >20% of tests (calculated from Table S4). This result suggests that *D*FOIL may perform well despite missing data or limited intra-population sampling.

The signals of introgression that we recovered between our two focal species (*S. oberon, S. ornatus*) were phylogenetically ancient and geographically widespread (Figures 3,4). However, we did discover subtle but sensible geographic variation in introgression within *S. ornatus* (discussed below). Presumably, the exhaustive application of *D*FOIL demonstrated here could reveal stronger variation in the phylogenetic, geographic, and/or temporal evidence of introgression in other cases. Nevertheless, an important shortcoming of Ex*D*FOIL is that the quantity of introgression, in terms of genomic contribution from each species, is not estimated. This could be addressed instead population-based approaches (e.g., Reich et al., 2009; Pickrell & Pritchard, 2012).

Finally, we note that we used Ex*D*FOIL here to investigate a single a-priori hypothesis of introgression. However, we believe the approach may be particularly well-suited as a tool for *de novo* discovery of introgression as it produces a richly informative view of the signal (and noise) in a given dataset with relatively few assumptions.

**4.2 | Mito-nuclear discordance indicates recurrent introgression**

We argue that the rampant paraphyly of mtDNA from *S. ornatus* and *S. oberon* that we observe here is best explained by repeated introgression betweenthese species (Figure 2). This introgression may have been influenced by Pliocene climate change and/or Pleistocene climate cycles and associated range shifts or population-size fluctuations (although this would require detailed analyses to test, beyond the scope of the present study). We think it is unlikely that incomplete lineage sorting alone could produce these patterns, a conclusion that is in agreement with similar studies of mtDNA discordance in related lizard clades (e.g., McGuire et al., 2007; Jezkova, Leal, & Rodríguez-Robles, 2013). First, it is less likely that a species will be non-monophyletic in mtDNA due to incomplete lineage sorting, given the much smaller effective population size and rapid evolution of mtDNA (e.g., Wiens & Penkrot, 2002; Hudson & Turelli, 2003; Zink & Barrowclough, 2008). Second, there is a geographically biased distribution of putatively introgressed haplotypes, centered on the contact zone between *oberon*-black and *S. ornatus* (Figure 1). This pattern is not expected under incomplete lineage sorting alone (Funk & Omland, 2003).

We suggest that mtDNA introgression most likely occurred from *oberon*-black into *S. ornatus,* based onthe co-occurrence of two distinct haplogroups within single *S. ornatus* localities near the range of *oberon*-black (Figure 1, panel D)*.* Under this proposal, one haplogroup represents “native” *S. ornatus* haplotypes, more closely related to *S. ornatus* samples from localities further west (depicted in blue in panels C and D of Figure 1). The other haplogroup represents introgressed haplotypes, more closely related to geographically proximate *oberon*-black (depicted in gray in panels C and D of Figure 1).

If introgression occurred in the opposite direction (from *S. ornatus* into *oberon*-black), we would need an alternative explanation for the retention of two distinct ND4 haplogroups in single localities of *S. ornatus* for at least ~4.64 Mya (the crown-node ancestor of all *S. oberon* and *S. ornatus*, 95% highest posterior density [HPD] 2.21–6.49 Mya). Several selection-based scenarios could explain polymorphism of ancient mtDNA, including negative frequency-dependent selection (Kazancıoğlu & Arnqvist, 2014) or sex-ratio distorting bacterial infection (e.g., Jiggins & Tinsley, 2005). Nevertheless, because introgression from *oberon*-black into *S. ornatus* does not require the invocation of any additional evolutionary forces, we consider it the more likely direction of mtDNA introgression.

Our analyses suggest that mtDNA introgression between *S. oberon* and *S. ornatus* occurred at least once, in an event as old as 2.29 Mya (crown-node ancestor of all *S. oberon* mtDNA samples, 95% HPD interval 1.31–3.38 Mya). One or two additional instances of mtDNA introgression are indicated by younger discordant relationships (grouping *S. ornatus* and *oberon-*black) as young as 0.61 Mya (youngest strongly-supported node with both *S. oberon* and *S. ornatus* descendants, 95% HPD 0.24–1.04 Mya; Figure 2). More ancient mtDNA introgression, involving the ancestors of *S. oberon* and *S. ornatus,* is also suggested by the discordance between mtDNA and ddRAD estimates for the phylogenetic position and age of the most recent *oberon*–*ornatus* ancestor (Figure 2)*.* If the mtDNA clade of *oberon-ornatus* is caused by introgression, then introgression may have occurred as long ago as ~5.29 Mya (stem-node of *S. oberon* and *S. ornatus* in the mtDNA tree, 95% HPD 3.34–7.26; Figure 2).

**4.3 | Historical introgression revealed using RADseq data and Ex*D*FOIL**

Our exhaustive *D*FOIL approach (Ex*D*FOIL) revealed phylogenetic and geographic patterns of introgression between *S. oberon* and *S. ornatus*. Introgression occurred anciently, between the ancestors of *S. ornatus* and *oberon*-black, and is broadly detectable across the range of both taxa. This includes populations from the western extent of the range of *S. ornatus*, >150 km from the range of *S. oberon* (Figures 3,4). However, a slight reduction in the frequency of positive tests is evident for these western-most populations. This may indicate that these populations harbor less introgressed ancestry than their eastern counterparts, which may have participated in more recent introgression. This result is consistent with the geographic patterns of mito-nuclear discordance we observed in *S. ornatus,* where only eastern-most populations contain recently-introgressed mtDNA haplotypes (see above). Although the western-most *S. ornatus* localities are exclusively from batch 2, we do not believe that this result is compromised by batch effects, as the geographic pattern is still apparent when comparing only batch 2 samples (Supplemental Figure S5).

Introgression involving *S. ornatus* and *oberon*-black is rarely inferred in tests that include one representative of *S. cyanostictus* or *S. cyanogenys* and one of *S. ornatus* (Figure 3). As discussed in section **3.4**, this result may be caused by higher rates of allelic dropout for *S. cyanostictus* and *S. cyanogenys*, at least for the targeted dataset. Closer examination of these tests, however, reveals that individual *D*FOIL components frequently indicate introgression involving *S. ornatus* and *oberon*-black, but also introgression between *oberon*-red and *S. cyanostictus* or *S. cyanogenys* (Tables S3–S6). It may be that the combination of introgression between e.g., P1↔P3 and P2↔P4 prevents *D*FOIL from inferring either introgression type in these cases.

We were able to detect introgression using the clade-wide dataset, but only rarely (<1% overall; Table 1), despite sampling >1,700 loci and >400,000 bp. This result is not necessarily surprising, however. Many tests using the clade-wide data may not have sufficient count data (e.g., zero observations for one or more site patterns; Table 4, Tables S5, S6). Clearly, our clade-wide dataset has low power to detect historical introgression. Nevertheless, this dataset is more appropriate for resolving phylogeny using ddRAD data, as it contains more accurate variants (see Appendix S1, Tables S12–S16), and does not share the same potential for mapping error bias of the targeted dataset (see section **2.4**).

We did not detect intergroup introgression signatures in large frequencies overall when excluding singleton counts (Tables 1–3). Moreover, many of these signatures may be compromised by batch-specific error (see next section). Simulations show that *D*FOIL may fail to polarize introgression when directional introgression is weak, occurs close to the divergence time of P1 and P2 (Figure 5 of Pease & Hahn, 2015), or is bi-directional (Schumer et al., 2016). The level of asymmetry in nuclear introgression between *oberon*-black and *S. ornatus* is therefore considered unknown.

**4.4 | Batch effects in RADseq data may mislead *D*-statistics and *D*FOIL**

The potential impact of batch effects on analyses of RADseq data is rarely discussed in the literature (but see Deagle, Faux, Kawaguchi, Meyer, & Jarman, 2015; Mastretta-Yanes et al., 2015). We found that batch effects apparently reduced power for detecting introgression, at least for the targeted datasets designed for analyzing introgression (Table 5). For the clade-wide datasets, batch 1 recovered slightly higher numbers of positive tests overall. This may be due to the greater number of individuals sequenced in batch 1 (especially for *oberon*-black and *S. ornatus* from near the contact zone), but the power of the clade-wide data was also consistently low (Table 5). As batch 1 had a shorter read length (100bp vs. 125bp), some reduction in power for tests using batch 1 samples was expected *a priori*.

We also found clear associations between particular introgression signatures and batch signatures (Figure 5, see Results). In some cases, these associations seem to likely reflect some level of type-I error induced by batch effects. In the case of “P1/P2←→P3” introgression and batch signature “1112”, the potential for batch effects to influence the results seems obvious, as introgression is inferred to involve the three taxa from batch 1, while the representative from batch 2 is not involved. However, the majority of these signatures correspond to introgression involving two representatives of *oberon*-black (~75%; Table S3). As batch 2 retained only one individual of *oberon*-black in the reduced-individual sampling, this type of introgression was not detectable using only batch 2 samples. In this case, confounding of batch and sampling effects prevents us from making a firm interpretation.

In order to understand how batch effects might influence “P1→P4” and “P2→P4” introgression signatures, a consideration of the unique *D*FOIL signatures (sensu Pease & Hahn, 2015) underlying these results is required. Briefly, *D*FOIL signatures are defined by the individual results for each of the four *D*FOIL components. Each component is a four-taxon *D*-statistic that can be significantly different than zero in either the positive (+) or negative (-) direction (each implying a different taxon is involved in introgression), or not significantly different than zero (0). For example, one potential signature would be “--+0”.

Importantly, both the “P1→P4” and “P2→P4” *D*FOIL signatures (“--0+” and “--0-”) differ from a *D*FOIL signature of ancestral introgression (“--00”) by only one *D*FOIL component (“*D*OL”). This means that a false positive result for *D*OL will imply directional introgression when placed against the background of a “true” signature of ancestral introgression. The *D*OL component compares the site-pattern counts for P1, P2, and P4. If batch-specific error is generating false positives, this should occur when P1 and P2 come from different batches, and P4 comes from the same batch as the donor taxon, while the batch identity of P3 should not matter. As expected, “P1→P4” introgression is overrepresented in tests with batch signatures of “1211” and “1221”, and “P2→P4” introgression is overrepresented in tests with batch signatures “2111” and “2121” (Figure 5). Specifically, we find that P1 and P2 are from different batches in 81.3% (334/411) of tests with “--0+” or “--0-” introgression signatures, and that P4 is from the same batch as the donor taxon in 95.8% of these cases (320/334), which is 77.8% of the total instances of directional introgression obtained using a mix of batch 1 and 2 samples (320/411, proportions calculated from Table S3). Unlike the situation described above for “P1/P2←→P3” introgression, no difference in sampling regime between the batches could fully explain this bias. We therefore conclude that some intergroup signatures we recovered are spuriously produced by batch effects, although we cannot say for sure which (or how many). This finding has important implications for other RADseq studies that focus on introgression using data from multiple batches.

**4.5 | The inclusion of singleton-counts may mislead *D*FOIL**

Pease and Hahn (2015) highlighted the theoretical potential for the inclusion of singleton counts to mislead *D*FOIL, but this issue has not been explored using simulated or empirical data (to our knowledge). We found that the inclusion of singleton counts in *D*FOIL tests dramatically increased the total proportion of positive tests recovered, and substantially altered the phylogenetic composition of the recovered introgression signatures (Table 6, Tables S7–S10). Specifically, adding singleton counts greatly increased the proportion of positive introgression results for tests that did not include any representative of *oberon*-black (from ~1% to >40% for the targeted dataset), and increased the overall proportion of intergroup introgression signatures recovered (from ~1% to ~6–8% for the targeted dataset). We consider these results suspect, independent of the actual singleton counts, for two reasons. First, most of the remaining intergroup signatures inferred when excluding singleton counts are potentially compromised by batch effects (see section **4.4**). Second, although we cannot rule out historical introgression between *oberon*-red and *S. ornatus*, they are not geographically adjacent, and *oberon*-red is monophyletic with respect to *S. ornatus* in our mtDNA tree (Figure 2). These results contrast with those for *S. ornatus* and *oberon*-black. Thus, we suspect that inferred introgression between *S. ornatus* and *oberon*-red is artifactual.

Our Ex*D*FOIL analyses also allow a direct appraisal of singleton-count ratios and associated results for thousands of tests. Using the targeted dataset with reduced-individual sampling, we compared average P3:P4 count ratios for tests excluding vs. including singleton counts separately for each *D*FOIL result (Figure 6). This comparison demonstrates that average count ratios when including singleton counts are typically much more extreme than count ratios when excluding singleton counts. In many cases, the average count ratio when including singleton counts exceeds the recommended bounds of 0.75 or 1.25 (Figure 6). Furthermore, count ratios were consistently skewed such that the taxon with a lower singleton count was inferred to be involved in introgression, matching the predictions of Pease and Hahn (2015). These results provide empirical support for the idea that the inclusion of singleton counts maybe problematic (Pease & Hahn, 2015).

**4.6 | Linkage in RADseq data may mislead *D*FOIL**

We found that using single-site-per-locus sampling and locus-level bootstrapping approaches reduced the number of introgression events observed when compared to the Chi-square method (Appendix S2; Table S11). Although we found that ~85% of cases were robust to this putative linkage issue, our results nevertheless indicate that the Chi-square method may be susceptible to the effects of linkage, even when sampling from ~5.5 x 106 bp (as in our targeted dataset). Thus, we encourage users to conduct similar re-sampling sensitivity analyses when combining RADseq data and *D*FOIL, and provide scripts to replicate our analyses at <https://www.github.com/SheaML/ExDFOIL/AppendixS2>.

**5 | Conclusions**

We demonstrate a novel application of the *D*FOIL method (Pease & Hahn, 2015) for detection of introgression. The approach (‘Ex*D*FOIL’) involves exhaustively applying *D*FOIL to hundreds of thousands of unique four-taxon combinations of individuals, here sequenced using a reduced-representation protocol (ddRAD). We demonstrate Ex*D*FOIL in an empirical system in which mito-nuclear discordance independently suggests recurrent introgression. We find that DFOIL can detect introgression under a broad range of genomic and geographic sampling conditions. Furthermore, the Ex*D*FOIL approach reveals subtle intra-specific geographic variation in introgression that is also consistent with observed patterns of mito-nuclear discordance and a hypothesis of recurrent introgression. Our results may also provide the first empirical evidence that batch effects in RADseq data can mislead inferences of introgression. We also find empirical support for the predictions of Pease and Hahn (2015) that the inclusion of singleton counts in *D*FOIL analyses may yield problematic results. Finally, we found that linkage between sites and/or loci in RADseq data may slightly inflate the rate of introgression recovery by *D*FOIL, even when sampling many sites (~5.5x106 in our case). We provide scripts to apply our approach to other datasets at https://www.github.com/SheaML/ExDFOIL.

**Acknowledgements**

We thank Harsimran Brar, Liz Miller, and Tereza Jezkova for help in the lab, and Anthony Baniaga, Tim O’Connor, and Tod W. Reeder for help in the field. We thank Nicholas Barton and an anonymous reviewer for helpful comments that improved the manuscript. We thank Kathryn Chenard for the artwork of *S. oberon* and *S. ornatus* in Figures 2, 3, and 4. This work was partially funded by awards from the Society for the Study of Evolution (Rosemary Grant Award) and Society of Systematic Biologists (Graduate Student Research Award) to SML. This material is based upon work supported by the U.S. National Science Foundation, including a Graduate Research Fellowship to SML (Grant No. 2012117663) and NSF DEB 1655690 to JJW. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

**REFERENCES**

Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B., & van Nimwegen, E. (2014). Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Molecular Biology and Evolution*, 31, 1077–1088. doi: 10.1093/molbev/msu088

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537. doi: 10.1371/journal.pcbi.1003537

Bouckaert, R. R., & Drummond, A. J. (2017). Bmodeltest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17, 42. doi: 10.1186/s12862-017-0890-6

Catchen, J., Hohenlohe, J. P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. doi: 10.1111/mec.12354

Chasalow, S. (2012). combinat: combinatorics utilities. R package version 0.0-8.

<https://CRAN.R-project.org/package=combinat>

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J. M., & Estoup, A. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187–1189. doi: 10.1093/bioinformatics/btt763

Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., DePristo, M. A., Handsaker, R., Lunter, G., Marth, G., Sherry, S.T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. doi: 10.1093/bioinformatics/Fbtr330

Deagle, B. E., Faux, C., Kawaguchi, S., Meyer, B., & Jarman, S. N. (2015). Antarctic krill population genomics: apparent panmixia, but genome complexity and large population size muddy the water. *Molecular Ecology,* 24, 4943–4959. doi: 10.1111/mec.13370

De Smet, W. H. O. (1981). The nuclear Feulgen-DNA content of the vertebrates (especially reptiles), as measured by fluorescence cytophotometry, with notes on the cell and chromosome size. *Acta Zoologica et Pathologica Antverpiensia*, 76, 119–167.

Doležel, J., Bartos, J., Voglmayr, H., & Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry Part A,* 51, 127–128. doi: 10.1002/cyto.a.10013

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, e88. doi: 10.1371/journal.pbio.0040088

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28, 2239–2252. doi: 10.1093/molbev/msr048

Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, 62, 689–706. doi: 10.1093/sysbio/syt032

Faircloth, B. C., & Glenn, T. C. (2014). Protocol: Preparation of an AMPure XP substitute (AKA Serapure). doi: 10.6079/J9MW2F26

Funk, D. J., & Omland, K. E. (2003). Species level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, 34, 397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, 328, 710–722. doi: 10.1126/science.1188021

Grummer, J. A., Calderón-Espinosa, M. L., Nieto Montes de Oca, A.., Smith, E. N., Méndez-de la Cruz, F. R., & Leaché, A. D. (2015). Estimating the temporal and spatial extent of gene flow among sympatric lizard populations (genus *Sceloporus*) in the southern Mexican highlands. *Molecular Ecology*, 24, 1523–1542. doi: 10.1111/mec.13122

Gutenkunst, R. N., Hernandez, R. G., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5: e10000695. doi: 10.1371/journal.pgen.1000695

Hall, W. P., & Selander, R. K. (1973). Hybridization of karyotypically differentiated populations in the *Sceloporus grammicus* complex (Iguanidae). *Evolution*, 27, 226–242. doi: 10.2307/2406963

Harrison, R. G. & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105, 795–809. doi: 10.1093/jhered/esu033

Heled, J., & Drummond, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, 61, 138-49. doi: 10.1093/sysbio/syr087

Hey, J., & Sousa, V. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, 14, 404–414. doi:10.1038/nrg3446

Hey, J. (2010). Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, 27, 905–920. doi: 10.1093/molbev/msp296

Huang, J.-P. (2016). Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (*Dynastes*, Dynastinae). *Molecular Ecology*, 25, 5513–5526. doi: 10.1111/mec.13849

Hudson, R. R., & Turelli, M. (2003). Stochasticity overrules the ‘‘three-times rule’’: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, 57, 182–190. doi: 10.1111/j.0014-3820.2003.tb00229.x

Jezkova, T., Leal, M., & Rodríguez-Robles, J. A. (2013). Genetic drift or natural selection? Hybridization and asymmetric mitochondrial introgression in two Caribbean lizards (*Anolis pulchellus* and *Anolis krugi*). *Journal of Evolutionary Biology*, 26, 1458–1471. doi: 10.1111/jeb.12149

Jiggins, F. M., & Tinsley, M. C. (2005). An ancient mitochondrial polymorphism in *Adalis bipunctata* linked to a sex-ratio-distorting bacterium. *Genetics*, 171, 1115–1124. doi: [10.1534/genetics.105.046342](https://doi.org/10.1534/genetics.105.046342)

Kazancıoğlu, E., & Arnqvist, G. (2014). The maintenance of mitochondrial genetic variation by negative frequency-dependent selection. *Ecology Letters*, 17, 22–27. doi: 10.1111/ele.12195

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M. A., & Janke, A. (2017). The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*, 7, 46487. doi: 10.1038/srep46487

Leaché, A. D. (2011). Multi-locus estimates of population structure and migration in a fence lizard hybrid zone. *PLoS ONE,* 6, e25827. doi: 10.1371/journal.pone.0025827

Leaché, A. D., Banbury, B. L., Felsenstein, J., Nieto Montes de Oca, A., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 64, 1032–1047. doi: 10.1093/sysbio/syv053

Leaché, A. D., Banbury, B. L., Linkem, C. W., & Nieto Montes de Oca, A. (2016). Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in North American spiny lizards (Genus *Sceloporus*)? *BMC Evolutionary Biology*, 16, 63. doi: 10.1186/s12862-016-0628-x

Leaché, A. D., & Cole, C. J. (2007). Hybridization between multiple fence lizard lineages in an ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology. *Molecular Ecology*, 16, 1035–1054. doi: 10.1111/j.1365-294X.2006.03194.x

Leaché, A. D., Harris, R. B., Maliska, M. E., & Linkem, C. W. (2013). Comparative species divergence across eight triplets of spiny lizards (*Sceloporus*) using genomic sequence data. *Genome Biology and Evolution*, 5, 2410–2419. doi: 10.1093/gbe/evt186

Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116. doi: 10.1111/2041-210X.12410

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50, 913–925. doi: 10.1080/106351501753462876

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics. Epub 2010 Jan 15. doi: 10.1093/bioinformatics/btp698

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23, 1817–1828. doi: 10.1101/gr.159426.113

Martínez-Méndez, N., & Méndez de la Cruz, F. R. (2007). Molecular phylogeny of the *Sceloporus torquatus* species-group (Squamata: Phrynosomatidae). *Zootaxa*, 1609.1, 53–68.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15, 28–41. doi:10.1111/1755-0998.12291

McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinal, J., Riddle, B. R., & Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution*, 61, 2879–2897. doi: 10.1111/j.1558-5646.2007.00239.x

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1–8.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. doi: 10.1093/bioinformatics/btg412

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, 25, 2337–2360. doi: 10.1111/mec.13557

Pease, J. B., & Hahn, M. W. (2015). Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, 64, 651–662. doi: 10.1093/sysbio/syv023

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. doi: 10.1371/journal.pone.0037135

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967. doi: 10.1271/journal.pgen.1002967

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431. doi: 10.7717/peerj.431

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. (2014). Tracer v1.6, Available from <http://tree.bio.ed.ac.uk/software/tracer/>

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. Nature 461: 489–494.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22, 939–946. doi: 10.1101/gr.128124.111

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, 19, 101–109. doi: 10.1093/oxfordjournals.molbev.a003974

Sarver, B. A. J., Keeble, S., Cosart, T., Tucker, P. K., Dean, M. D., & Good, J. M. (2017). Phylogenomic insights into mouse evolution using a pseudoreference approach. *Genome Biology and Evolution*, 9, 726–739. doi: 10.1093/gbe/evx034

Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G., & Andolfatto, P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology*, 25, 2661–2679. doi: 10.1111/mec.13602

Schwenk, K., Brede, N., & Streit, B. (2008). Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363, 2805–2811. doi: 10.1098/rstb.2008.0055

Sites, J. W. Jr., Davis, S. K., Hutchinson, D. W., Maurer, B. A., & Lara, G. (1993). Parapatric hybridization between chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae): structure of the Tulancingo transect. *Copeia*, 1993, 373–398. doi: 10.2307/1447137

Sites, J. W. Jr., Barton, N. H., & Reed, K. M. (1995). The genetic structure of a hybrid zone between two chromosome races of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in central Mexico. *Evolution*, 49, 9–36. doi: 10.1111/j.1558-5646.1995.tb05955.x

Smith, S. A., & O’Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28, 2689–2690. doi: 10.1093/bioinformatics/bts492

Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, 12, e1005896. doi: 10.1371/journal.pgen.1005896

Stamatakis A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, 9, 322. doi: 10.1186/1471-2105-9-322

Tange, O. (2011). GNU Parallel - The Command-Line Power Tool;login: The USENIX Magazine, February 2011, 42–47.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Twyford, A. D., & Ennos, R. A. (2012). Next-generation hybridization and introgression. *Heredity*, 108, 179–189. doi: 10.1038/hdy.2011.68

Uetz, P., Freed, P., & Hošek, J. (eds.). “The Reptile Database”. [http://www.reptile-database.org](http://www.reptile-database.org/), accessed October 2nd, 2017

Wiens, J. J., Kuczynski, C. A., Arif, S., & Reeder, T. W. (2010). Phylogenetic relationships of phrynosomatid lizards based on nuclear and mitochondrial data, and a revised phylogeny for *Sceloporus*. *Molecular Phylogenetics and Evolution*, 54, 150–161. doi: 10.1016/j.ympev.2009.09.008

Wiens, J. J., & Penkrot, T. L. (2002). Delimiting species based on DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Systematic Biology*, 51, 69–91. doi: 10.1080/106351502753475880

Wiens, J. J., Kozak, K. H., & Silva, N. (2013). Diversity and niche evolution along aridity gradients in North American lizards (Phrynosomatidae). *Evolution*, 67, 1715–1728. doi: 10.1111/evo.12053

Wiens, J. J., Reeder, T. W., & Nieto Montes de Oca, A. (1999). Molecular phylogenetics and evolution of sexual dichromatism among populations of the Yarrow’s spiny lizard (*Sceloporus jarrovii*). *Evolution*, 53, 1884–1897. doi: 10.2307/2640448

Zink, R. M., & Barrowclough, G. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17, 2107–2121. doi: 10.1111/j.1365-294X.2008.03737.x

**FIGURE AND TABLE LEGENDS**

**Figure 1**. **A.** Adult male *S. ornatus*. **B.** Representative habitat of *S. ornatus*: rocky slopes with Chihuahuan desert-scrub vegetation.**C.**A median-joining haplotype network of ND4 sequences for *S. ornatus* (in blue and gray), *S. oberon “*black” (in black) and *S. oberon* “red” (in red) made using PopART (Leigh & Bryant, 2015). The number of samples for each haplotype is indicated by the circle size (see scale in top-left). **D.** Map of the contact zone between *S. oberon* and *S. ornatus* in northeastern Mexico. Black circles indicate sampled localities of *oberon*-black, red circles indicate sampled localities of *oberon*-red. Localities for *S. ornatus* indicate the proportion of putatively introgressed (gray) and putatively native (blue) ND4 haplotypes, as in the haplotype network. Map is a composite of digital elevation models and satellite imagery made in QGIS v2.18 (http://qgis.osgeo.org/). Landsat 7 imagery courtesy of the U.S. Geological Survey. Note that this map encompasses the entire range of *S. oberon*, but only the eastern portion of the range of *S. ornatus*. **E.**Representative habitat of *S. oberon* “black”: high-elevation pine-oak forest. **F.** Adult male *S. oberon* “black”. **G.** Adult male *S. oberon* “red”. **H.** Inset map and legend for map in panel D. Photo A by John Wiens, photos B and E by Anthony Baniaga, photos F and G by Shea Lambert.

**Figure 2.** Time-calibrated trees for ddRAD data using the clade-wide dataset and reduced-individual sampling, using RAxML and treePL (above) and the mtDNA gene ND4 (below), using BEAST2. Blue shading indicates groups of samples assigned to *S. ornatus*, gray shading indicates groups of samples assigned to *S. oberon* “black”, and red shading indicates groups of samples assigned to *S. oberon* “red”. For the ND4 tree, support values indicated on nodes are posterior probabilities, and for the ddRAD tree, support values are the percentage of bootstrap replicates supporting that bipartition. For both trees, support values of less than 50 / 0.5 are not shown, and asterisks indicate support values of 100 / 1.

**Figure 3.** Visualization of the proportions of positive tests for introgression over phylogenetic space, using the targeted dataset with reduced-individual sampling. Tree is based on the clade-wide dataset (Figure 2). Pies above nodes indicate results when that node was the most recent common ancestor of taxa P3 and P4 (the older pair of taxa); pies below nodes indicate results when that node was the most recent common ancestor of taxa P1 and P2 (the younger pair of taxa). Only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these *oberon*-black) are considered here, as in Tables 1-3.. Gray indicates no introgression, blue indicates ancestral introgression involving two *S. ornatus* and one *S. oberon*, red indicates introgression from *S. ornatus* into *S. oberon*, orange indicates ancestral introgression involving two *S. oberon* and one S*. ornatus*, yellow indicates introgression from *S. oberon* into *S. ornatus*, and green represents ancestral introgression involving two *S. oberon* and one *S. cyanostictus*. Batch identities are indicated at the end of sample names (b1 = batch 1, b2 = batch 2). Prefixes indicate the species and locality for each sample, as seen in Figure 4 and Table S2.

**Figure 4.** Visualization of the proportions of positive tests for introgression over geographic space, using the targeted dataset with reduced-individual sampling. Localities of *oberon*-red are in red, localities of *oberon*-black are in black, and localities of *S. ornatus* are in blue. For simplicity, we consider tests where P1 and P2 were drawn from *S. ornatus*, and P3 and P4 were drawn from one representative each of *oberon*-black and *oberon*-red. Futhermore, we combine test results for individual localities regardless of if the locality was used for P1, P2, P3 or P4.. Green indicates ancestral introgression involving two *S. ornatus* and one *S. oberon*, andblue indicates introgression from *S. ornatus* into *S. oberon*. All remaining tests returned no introgression.

**Figure 5.** Stacked bar plots displaying all positive test results for the targeted dataset with reduced-individual sampling, comparing tests with each possible “batch signature”, defined simply as the batch identity of samples P1, P2, P3, P4 in that order (e.g., “1111” indicates that all four taxa belong to batch 1). Test results are color coded. For example, “P4 → P2” represents intergroup introgression from taxon P4 into taxon P2.

**Figure 6.** Comparison of the ratio of singleton counts in taxa P3:P4 for tests including (gray) or excluding (black) singleton counts, for the targeted dataset with reduced-individual sampling. Dashed lines are drawn at values of 1.25 and 0.75 for the P3:P4 ratio, as tests with count ratios that exceed these bounds are considered to be potentially problematic (i.e., *D*FOIL prints a warning). Bar plots are grouped by test results, indicated on the x-axis. For example, “123” representing ancestral introgression involving P1 / P2 and P3, and “13” indicating intergroup introgression from P1 into P3. We note that count ratios for introgression results of “23” or “na” are not compared here, as there were no results of “23” when singleton counts were excluded, and no results of “na” when singleton counts were included. Otherwise, all tests of the targeted dataset with reduced individual sampling are considered.

**TABLE 1** Coarse-scale summaries of the proportions of positive tests for introgression using *D*FOIL, for each of the four primary datasets. “Ancestral” signatures involve the ancestor of taxa P1 and P2, and one of either P3 or P4, and do not have directionality inferred. Inter-group signatures involve any two non-sister terminal taxa, with directionality inferred. Only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these *oberon*-black) are considered here.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Number of tests | Proportion of positive tests | Proportion of “ancestral” signatures | Proportion of “inter-group” signatures |
| Targeted, Full | 117,600 | 0.350 | 0.331 | 0.019 |
| Targeted, Reduced | 12,376 | 0.394 | 0.359 | 0.035 |
| Clade-wide, Full | 117,600 | 0.007 | 0.007 | 0 |
| Clade-wide, Reduced | 12,376 | 0.016 | 0.016 | 0 |

**TABLE 2** Raw counts of *D*FOIL signatures recovered for each of the four primary datasets. Signatures are indicated by the taxa involved (e.g., “4” = P4) with directionality indicated by the arrow; “12” indicates the ancestor of P1 and P2. The *D*FOIL signature (sensu Pease & Hahn, 2015) is indicated in parentheses, and corresponds to the results of each of the four *D*FOIL components (*D*FO, *D*IL, *D*FI, and *D*OL, respectively; + indicates significantly positive, - indicates significantly negative, and 0 indicates not significantly different than zero). Only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these *oberon*-black) are considered here.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Signatures | Targeted, Full | Targeted, Reduced | Clade-wide, Full | | Clade-wide, Reduced | |
| 12 ↔ 3 (++00) | 38,533 | 183 | | 876 | | 199 |
| 12 ↔ 4 (--00) | 426 | 4,262 | | 0 | | 0 |
| 1 → 3 (+++0) | 1,312 | 10 | | 0 | | 0 |
| 1 → 4 (--0+) | 4 | 222 | | 0 | | 0 |
| 2 → 3 (++-0) | 911 | 0 | | 0 | | 0 |
| 2 → 4 (--0-) | 2 | 189 | | 0 | | 0 |
| 3 → 1 (+0++) | 20 | 1 | | 0 | | 0 |
| 3 → 2 (0+--) | 7 | 2 | | 0 | | 0 |
| 4 → 1 (-0++) | 0 | 3 | | 0 | | 0 |
| 4 → 2 (0---) | 2 | 1 | | 0 | | 0 |

**TABLE 3** Proportions of positive tests for introgression recovered by *D*FOIL, collated by the species involved, for each primary dataset. Only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these *oberon*-black) are considered here.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | *S. ornatus / S. ornatus ↔ S. oberon* | *S. oberon / S. oberon ↔ S. ornatus* | *S. ornatus → S. oberon* | *S. oberon → S. ornatus* |
| Targeted, Full | 0.313 | 0.017 | 0.019 | < 0.001 |
| Targeted, Reduced | 0.340 | 0.019 | 0.033 | 0.001 |
| Clade-wide, Full | 0.007 | 0 | 0 | 0 |
| Clade-wide, Reduced | 0.016 | 0 | 0 | 0 |

**TABLE 4** Median, range, and standard deviation of the number of non-singleton counts for the targeted and clade-wide datasets. These values are compared for all tests, tests that recovered no introgression, and tests that did infer introgression.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Tests considered | Median of non-singleton counts | Range of non-singleton counts | Standard deviation of non-singleton counts |
| Targeted, Reduced | All tests | 406 | 153 – 1313 | 146 |
|  | No introgression | 388 | 153 – 1313 | 147 |
|  | Introgression | 508 | 214 – 1232 | 121 |
|  |  |  |  |  |
| Clade-wide, Reduced | All tests | 89 | 29 – 207 | 32 |
|  | No introgression | 88 | 29 – 207 | 32 |
|  | Introgression | 118 | 77 – 159 | 16 |
|  |  |  |  |  |
| Targeted, Full | All tests | 359 | 74 – 1504 | 126 |
|  | No introgression | 349 | 74 – 1504 | 129 |
|  | Introgression | 395 | 113 – 1297 | 103 |
|  |  |  |  |  |
| Clade-wide, Full | All tests | 91 | 9 – 279 | 33 |
|  | No introgression | 91 | 9 – 279 | 33 |
|  | Introgression | 113 | 69 – 160 | 16 |

**TABLE 5** Proportions of positive *D*FOIL tests for each dataset, comparing tests using only batch 1 samples, tests using only batch 2 samples, and tests using a mix of samples from batches 1 and 2. Only tests including two individuals of *S. ornatus* and two individuals of *S. oberon* (with at least one of these *oberon*-black) are considered here.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Batch identity | Number of tests | Proportion of positive tests | Proportion of “ancestral” signatures | Proportion of “inter-group” signatures |
|  |  |  |  |  |  |
| Targeted, Reduced | Batch 1 only | 1260 | 0.397 | 0.388 | 0.009 |
|  | Batch 2 only | 180 | 0.767 | 0.706 | 0.061 |
|  | Mixed | 10936 | 0.387 | 0.350 | 0.037 |
|  |  |  |  |  |  |
| Targeted, Full | Batch 1 only | 18473 | 0.354 | 0.349 | 0.005 |
|  | Batch 2 only | 1045 | 0.487 | 0.467 | 0.009 |
|  | Mixed | 98082 | 0.348 | 0.327 | 0.022 |
|  |  |  |  |  |  |
| Clade-wide, Reduced | Batch 1 only | 1260 | 0.018 | 0.018 | 0 |
|  | Batch 2 only | 180 | 0.011 | 0.011 | 0 |
|  | Mixed | 10936 | 0.016 | 0.016 | 0 |
|  |  |  |  |  |  |
| Clade-wide, Full | Batch 1 only | 18473 | 0.012 | 0.012 | 0 |
|  | Batch 2 only | 1045 | 0.003 | 0.003 | 0 |
|  | Mixed | 98082 | 0.007 | 0.007 | 0 |

**TABLE 6** Proportions of positive *D*FOIL tests for each dataset, comparing tests with singleton counts excluded against tests with singleton counts included.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Singleton counts | Proportion of positive tests | Proportion ancestral | Proportion inter-group | Proportion of positive results for tests without *oberon*-black |
|  |  |  |  |  |  |
| Targeted, Reduced | excluded | 0.159 | 0.145 | 0.013 | 0.011 |
|  | included | 0.450 | 0.369 | 0.081 | 0.438 |
|  |  |  |  |  |  |
| Targeted, Full | excluded | 0.165 | 0.157 | 0.009 | 0.011 |
|  | included | 0.403 | 0.342 | 0.061 | 0.405 |
|  |  |  |  |  |  |
| Clade-wide, Reduced | excluded | 0.006 | 0.006 | 0 | 0 |
|  | included | 0.094 | 0.091 | 0.003 | 0.121 |
|  |  |  |  |  |  |
| Clade-wide, Full | excluded | 0.003 | 0.003 | 0 | < 0.001 |
|  | included | 0.103 | 0.098 | 0.005 | 0.110 |

**Data accessibility**

All alignments and tree files are available in the Dryad Digital Repository (available upon acceptance). Scripts and example files used to select taxa for *D*FOIL, execute *D*FOIL in parallel, collate results, calculate summary statistics and visualize results are found in the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL. Genbank accession numbers for all previously published ND4 sequences are found in Table S1 (accession numbers for newly generated sequences will be made available upon acceptance). Raw sequence reads and quality scores for all samples used will be made available on the NCBI Sequence Read Archive upon acceptance.

**Author Contributions**

SML designed the study, conducted fieldwork, laboratory work, bioinformatic processing and downstream analyses. JWS conducted laboratory work and provided technical supervision in the lab. MCF-R provided technical supervision in the lab. UOG, ANMO, NM, and JJW conducted fieldwork and provided tissue samples. SML wrote the manuscript and created figures. All authors contributed to editing the manuscript.

**Supporting Information**

**Appendix S1.** Description of ddRAD library preparation and sequencing, variant filtering, haplotyping, error-rate estimation, mitigation of batch effects, and removal of discordant phylogenetic signal related to introgression.

**Appendix S2.** Bootstrapping and single site per locus Ex*D*FOIL analyses; testing the impacts of linkage and frequency of putative type-I error.

**Appendix S3.** Comparison of Ex*D*FOIL with TreeMix.

**Supplemental File S1.** Custom R function for selecting taxa from a list that meet the assumptions of *D*FOIL, given a phylogenetic hypothesis containing all of the taxa.

**Supplemental File S2.** Meta-data file for Supplemental Tables S3 and S4

**Figure S1.** Chronogram made using the clade-wide alignment with full individual sampling. Rapid bootstrap support values from RAxML are provided as node labels.

**Figure S2.** Maximum-likelihood tree obtained using RAxML and the clade-wide dataset with reduced individual sampling, before removing two populations of *S. minor* (“minor6” and ”minor7”) with a potential history of introgression with *S. cyanogenys*. Rapid bootstrap support values are provided as node labels.

**Figure S3.** Maximum-likelihood tree obtained using RAxML and the clade-wide dataset with reduced individual sampling, after removing all samples from localities “minor6” and “minor7”. Rapid bootstrap support values are provided as node labels.

**Figure S4.**  Maximum-likelihood tree obtained using RAxML and the clade-wide dataset with reduced individual sampling, after removing the sample of *S. cyanogenys* from within the range of *S. oberon* in Nuevo Leon, JJW593. Rapid bootstrap support values are provided as node labels.

**Figure S5.** Visualization of introgression results over geographic space, for the targeted dataset with reduced individual sampling, and only batch 2 samples considered. For simplicity, we do not include tests that involved only *S. oberon* “black” or *S. oberon* “red” representatives in this visualization. Furthermore, we only consider tests where P1 and P2 were drawn from *S. ornatus*, and combine test results for individual localities regardless of if the locality was used for P1 or P2. Red indicates ancestral introgression involving two *S. ornatus* and one *S. oberon*, pink indicates introgression from *S. ornatus* into *S. oberon*, and blue indicates no introgression. As in Figure 3, a geographic pattern is apparent, where western-most populations of *S. ornatus* recover introgression less often.

**Figure S6.** Results of pilot bootstrapping analyses using 100 tests that returned ancestral introgression and 100 tests that returned no introgression using the Chi-square method and targeted dataset with reduced individual sampling. Panel on the left shows the number of tests that remained non-significant (no introgression) and panel on the right shows the number of tests that remained significant for ancestral introgression. At each level of bootstrap replication (indicated on the x-axis), the corresponding number of replicates was sampled 100 times and the number of significant/non-significant results was tallied each time, and the resulting distributions used to create boxplots.

**Figure S7.** Barplot comparing the number of results for each of the five most common introgression signatures using 1) full data and Chi-square method 2) locus bootstrapping and *Z*-scores, and 3) single SNP per locus analyses and Chi-square method. Includes only tests with two representatives of *S. ornatus* and two representatives of *S. oberon* (with at least one *oberon*-black) from the targeted dataset with reduced-individual sampling.

**Figure S8.** Results of our TreeMix analyses with one and two migration edges added, indicating introgression between the *oberon*-black clade and ancestors of *S. ornatus*. Horizontal branch lengths indicate the relative amount of genetic drift occuring on each branch, and the scale bar indicates ten times the average standard error of the sample covariance matrix. Migration strength is indicated by the color of the arrow, and corresponds to fraction of alleles in the recipient population that originated in the donor population.

**Table S1.** Genbank accession numbers for all previously published and newly generated ND4 sequences (Excel spreadsheet).

**Table S2.** ddRAD sample information and read counts after quality filtering with dDocent (Excel spreadsheet).

**Table S3.** Full results of reduced-individual sampling Ex*D*FOIL analyses.

**Table S4.** Full results of full-individual sampling Ex*D*FOIL analyses.

**Table S5.** Error rates for the clade-wide dataset, before any filtering steps (using the TotalRawSNPs.vcf file).

**Table S6.** Error rates for the clade-wide dataset, after filtering sites for a minimum quality of 30 (Phred33) and a minimum depth of 3 (Excel format spreadsheet).

**Table S7.** Error rates for the clade-wide dataset, after all filtering steps, except for the removal of batch effects, minor allele count filter, and haplotyping with rad\_haplotyper.pl.(Excel format spreadsheet).

**Table S8.** Error rates for the clade-wide dataset, after removal of batch effects and filtering of sites with minor allele counts less than three. (Excel format spreadsheet).

**Table S9.** Error rates for the clade-wide dataset, after haplotyping and filtering of potential paralogs with rad\_haplotyper.pl. (Excel format spreadsheet).

**Table S10.** Error rates for the targeted dataset, after filtering sites for a minimum quality of 30 and a minimum depth of 3. (Excel format spreadsheet).

**Table S11.** Summary of results using bootstrapping and single SNP per locus sampling approaches, collated by the species involved (as in Table 3). Results shown are for the targeted dataset with reduced individual sampling, including only tests with two representatives of *S. ornatus* and two representatives of *S. oberon* (with at least one *oberon*-black), as in Tables 1–3.