**Appendix S1.** Full description of ddRADseq library preparation and sequencing, variant filtering, error rate estimation, mitigation of batch effects, haplotyping, and removal of discordant phylogenetic signal owing to introgression.

**ddRAD library preparation and sequencing**

We prepared ddRAD libraries generally following Peterson et al. (2012), with departures described here and in *Methods* section **2.2** of the main text. We prepared two separate “batches” of ddRAD libraries, sequenced at different times and sequencing centers (hereafter “batch 1” and “batch 2”). Although the ddRAD protocol was largely identical in both cases, there were several technical differences in library preparation and sequencing.

First, the DNA gel cassette used for size-selection with a Pippin Prep (Sage Science) differed between batches. For batch 1, we used a dye-free Marker K 1.5% agarose cartridge (Product #CDF1510) with internal standards. For batch 2, we used a Marker A 1.5% agarose cartridge (Product #CSD1510) with ethidium bromide dye and an external standard. For batch 1, we entered a size-selection range of 438–538 base pairs (bp). However, when visualizing the fragment length distribution of test libraries for batch 2 using an Agilent 2100 Bioanalyzer, it became clear that the distribution was non-overlapping with that of batch 1, even though an identical size-selection range was entered. As unequal size-selection might preclude the combination of data from each batch, we altered the size-selection window entered for batch 2 to 545–645 bp. This gave a size-selection distribution on the Bioanalyzer close to that observed for batch 1. Peak concentrations were observed at ~625 bp for both batches following final PCR amplification.

For batch 1, Illumina PCR conditions consisted of 10 reactions of 12 cycles each, using 4µL of template, 4µL of nuclease-free H20, 1ul of each primer (10 µM concentration), and 10µL of Phusion High-Fidelity DNA polymerase master mix (Thermo Scientific). For batch 2, conditions consisted of 10 reactions of 12 cycles each, using 2–4µL of template, 4–6µL of nuclease-free H20, 1µL of each primer (10 µM concentration), and 10µl of NEB OneTaq master mix. Following PCR and bead clean-up, we quantified the pools using an Agilent 2100 Bioanalyzer and DNA 7500 Chip. We then pooled each of the 10 PCR-pools in equimolar ratios using these concentrations.

Batch 1 was sequenced at the UT Southwestern Genomics core using one lane of Illumina HiSeq 2500 (rapid-run mode) and 100-cycle paired-end sequencing. Batch 2 was sequenced at the BYU DNA Sequencing Center using one lane of Illumina HiSeq 2500 (high-output mode) and 125-cycle paired end sequencing. Six out of eight PCR barcodes were resequenced from batch 2 at the same facility using the same instrument; the other two did not have sufficient concentrations for resequencing. For all resequenced samples, we simply concatenated the fastq files from both sequencing efforts.

**Variant filtering**

For the clade-wide datasets, we filtered variants based on guidelines set out by the dDocent authors ([www.dDocent.com](http://www.dDocent.com/)), with some modifications. First, we filtered sites based on quality score and depth using vctools v0.1.15 (Danacek et al., 2011), keeping only sites with a minimum quality score of 30 (using the --minQ flag) and a minimum depth of 3 (using the --minDP) flag. Next, we calculated the amount of missing data for each individual using the --missing-indv flag. We then removed individuals with more than 50% missing sites (these samples are indicated in Table S1). Next, we kept only sites with data for 80% or more of individuals, using the --max-missing flag.

We then used a modified version of the dDocent\_filters.sh script, provided by the dDocent authors <<https://github.com/jpuritz/dDocent/raw/master/scripts/dDocent_filters>>, to further filter sites. This script executes a series of filters designed to remove problematic variants, and relies on vcftools and vcflib (<https://github.com/ekg/vcflib>). These filters are described and justified in detail in the Supplemental Information of Puritz et al. (2016). These filters include filtering based on allelic balance at heterozygous sites, reference and alternate allele mapping qualities, read pair-status, read-overlap, and maximum depth. We modified a version of dDocent\_filters.sh downloaded on June 11, 2017. We adjusted the values for filtering based on allele balance, keeping sites with allelic balance between 0.30 and 0.70 (modified from 0.20 and 0.80) based on both error rate and the number of retained sites. We removed the filter based on the ratio of site quality to site depth, as we found that this filter removed a large number of sites, while at the same time not improving error rates based on replicate samples (see *Error-rate estimation* below). We modified the filter that was based on mapping quality of reference and alternate alleles, keeping loci with an alternate:reference mapping quality ratio of > 0.25 and < 1.75, modified from 0.5 and 1.5, respectively. All other filters were left unchanged. A copy of our modified dDocent\_filters.sh file is available in the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL.

Following filtering with dDocent\_filters.sh, we filtered out variants from sites within repetitive or low-complexity regions. We identified such regions using the web version of RepeatMasker (Smit et al. 2013), using the “cross\_match” search engine. We then converted the RepeatMasker output into a .bed file, and excluded variants from within the identified region using the --exclude-bed flag of vcftools. Next, we decomposed complex (multi-site) variants into single nucleotide polymorphisms (SNPs) using the vcfallelicprimitive function of vcflib. Finally, we removed indels and kept only sites with a minor allele count of at least 3, using the --remove-indels and --mac flags of vcftools, respectively.

**Error-rate estimation**

We included 21 replicate samples in batch 2. These samples were taken from the same gDNA extractions as used in batch 1 for each corresponding individual. The inclusion of these samples allowed for the estimation of error rates (e.g., Mastretta-Yanes et al., 2015). We used a mismatch error rate, defined as the raw distance between replicate samples, calculated at shared biallelic sites using vcf files and the snpgdsIBS function of the R package SNPRelate (Zheng et al., 2012). Six out of 21 replicate pairs had at least one replicate excluded during our variant filtering pipeline (for having >50% missing sites, see ***Variant filtering***), and were therefor not used for the estimation of error rates. Because we did not call variants on individuals of *S. minor* for the targeted dataset, a further reduced set of nine replicates was used to estimate error rates for the targeted variants.

We examined the influence of our assembly and variant filtering steps on error rate, using error rate to help guide filtering parameters in some cases. However, we did not attempt to strictly minimize error rate, as we also wanted to retain as many informative sites as possible. Error rates for the clade-wide dataset, at various stages of filtering, and the targeted dataset, after all filtering (described in the main text), are provided as Supplemental Tables S5–S10.

**Batch-effect mitigation**

Because of differences in library preparation and sequencing, we expected our data to contain “batch effects,” including batch-specific recovery of loci, and batch-specific errors in genotyping (Mastretta-Yanes et al. 2015). To identify and mitigate these batch effects, we used Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010), executed using the adegenet package v2.0.1 (Jombart, 2008; Jombart & Ahmed, 2011) in R v3.4.1, generally following the methodology described in Appendix 3 of Deagle et al. (2015). By comparing replicate samples, and samples from populations with individuals included in both batch 1 and batch 2, DAPC can be used to find loci that effectively distinguish the batches, when batch identity is provided as the grouping variable for the discriminant function analysis.

We used a total of 87 samples to identify batch effects, including 15 replicates and 72 samples from populations sequenced in both batches. The full list of samples used is available in the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL. We retained the first 20 principal components, as they explained ~60% of the cumulative variance, and the retention of too many principal components can lead to model over-fitting and instability (<http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf>). Finally, we removed all sites with alleles that had with loading values in the top 5%. In total, 526 sites were removed. The .bed file used to exclude these sites with vcftools is available in the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL.

**Haplotyping**

Following all filtering steps, and the removal of sites contributing to the batch effect, we used the perl script rad\_haplotyper v1.1.5 (Willis et al., 2017) to phase our ddRAD loci, and further filter loci that may represent paralogs. Another advantage of this script is that it provides the full sequences of retained loci, including invariant sites, which are important for topological and branch-length accuracy when using RADseq data to infer phylogeny (Leaché et al., 2015). Briefly, this script samples raw reads for each individual and locus. The script then attempts to bin the raw reads into a maximum of two haplotypes, phased across paired-end reads. Recovery of more than two haplotypes indicates a potential paralog. We filtered loci for which more than two haplotypes were recovered in five or more individuals, using the --mp flag, as well as loci for which haplotypes could not be recovered in more than 50% of individuals, using the --m flag. We specified that output files be written in vcf and IMa2 format, and left all other parameters at default.

We converted the IMa2 output of rad\_haplotyper to phylip format using PGDSpider v2.1.1 (Lischer & Excoffier 2012). We then used custom shell scripts and publicly available perl scripts (http://raven.iab.alaska.edu/~ntakebay/teaching/programming/perl-scripts/perl-scripts.html) to obtain a concatenated, consensus sequence alignment in .fasta format, with heterozygotes coded as IUPAC ambiguity codes, for use in phylogenetic analysis. Final alignment files are available at the Dryad Digital Repository (available upon acceptance) and at https://www.github.com/SheaML/ExDFOIL.

**Removal of discordant phylogenetic signal** **owing to introgression**

In the initial phylogenetic analyses using maximum-likelihood in RAxML, we noticed weak support for the placement of *S. cyanogenys*, and for two populations of *S. minor*, one from San Luis Potosi and the other from Tamaulipas (populations “minor6” and “minor7”, Fig. S2). Given the geographic position of these *S. minor* populations, both abutting the range of *S. cyanogenys* on the eastern side of the Sierra Madre Oriental, we suspect that they may have experienced a history of introgression with *S. cyanogenys.* After removing all samples of *S. minor* from these localities from our alignment, we observed much stronger (but still moderate) support for the placement of *S. cyanogenys* (0.82 bootstrap support, Fig. S3). However, support for monophyly of *S. oberon* became weaker (0.89 bootstrap support, down from 0.95; Figs. S2,S3). Given that one of the two *S. cyanogenys* samples included was from a locality overlapping the range of *S. oberon* in Nuevo León (JJW593), we suspected that this individual may have experienced introgression with *S. oberon*. We were further motivated to remove this sample by unpublished mitochondrial and nuclear evidence for a complicated history of introgression between *S. oberon*, *S. cyanogenys*, and an apparent sister lineage to *S. cyanogenys,* found in the nearby Cerro de la Silla, Nuevo León, potentially representing the taxon “*S. jarrovii* *cyaneus*” (Treviño-Saldaña 1988). This history of introgression will be explored in future work. After removing the *S. cyanogenys* sample from Nuevo Leon, support for the monophyly of *S. oberon* and the placement of *S. cyanogenys* increased to >0.90 bootstrap support (Fig. S4).

**Literature Cited**

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

Jombart, T., Devillard, S., Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. doi: 10.1186/1471-2156-11-94

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405. doi: 10.1093/bioinformatics/btn129

Jombart, T. & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 17, 3070-3071. doi: 10.1093/bioinformatics/btr521

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

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

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298-299. doi: 10.1093/bioinformatics/btr642

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

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

Puritz, J. B., Portnoy, D. S., & Gold, J. R. (2016). Fine-scale partitioning of genomic variation among recruits in an exploited fishery: causes and consequences. *Scientific Reports*, 6, 36095. doi:10.1038/srep36095

Treviño-Saldaña, C. H. (1988). A new montane lizard (*Sceloporus jarrovii cyaneus*) from Nuevo León, México. *Revista de Biologia Tropical*, 36, 407-411.

Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Portnoy, D. S., & Gold, J. R. (2017). Haplotyping across RAD loci as an efficient method to characterize and filter paralogs and genotyping errors. *Molecular Ecology Resources*, 17, 955–965. doi:10.1111/1755-0998.12647

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C. & Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28, 3326-3328. doi:10.1093/bioinformatics/bts606.