Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex

Kin Onn Chan[1], Carl Hutter[2], Perry Lee Jr Wood[2], Lee Grismer[3], and Rafe Brown[2]

[1]National University of Singapore
[2]University of Kansas
[3]La Sierra University

May 5, 2020

Abstract

Most new cryptic species are described using conventional tree- and distance-based species delimitation methods (SDMs), which rely on phylogenetic arrangements and measures of genetic divergence. However, although numerous factors such as spatial population structure and gene flow are known to confound phylogenetic and species delimitation inferences, the influence of these processes on species estimation is rarely evaluated. Using large amounts of exons, introns, and ultraconserved elements obtained using the FrogCap sequence-capture protocol, we compared conventional SDMs with more robust genomic analyses that assesses spatial population structure and gene flow to characterize species boundaries in a Southeast Asian frog complex (Pulchrana picturata). Our results showed that gene flow and introgression can produce phylogenetic patterns and levels of divergence that resemble distinct species (up to 10% divergent in mitochondrial DNA). Hybrid populations were inferred as independent (singleton) clades that were highly divergent from adjacent populations (7–10%) and unusually similar (<3%) to allopatric populations. Such anomalous patterns are not uncommon in Southeast Asian amphibians, which brings into question whether the high cryptic diversity observed in other amphibian groups reflect distinct cryptic species—or, instead, highly structured and admixed metapopulation lineages. Our results also provide an alternative explanation to the conundrum of divergent (sometimes non-sister) sympatric lineages–a pattern that has been celebrated as indicative of true cryptic speciation. Based on these findings, we recommend that species delimitation of continuously distributed "cryptic" groups should not rely solely on conventional SDMs but should necessarily examine spatial population structure and gene flow to avoid taxonomic inflation.

Introduction

Species delimitation plays a pivotal role in biodiversity research, with potential cascading effects in conservation and other applied sciences (Devitt, Wright, Cannatella, & Hillis, 2019; Stanton et al., 2019). As the majority of uncontroversial, obviously distinct lineages have been described, attention is now turning towards identification of "cryptic" species complexes that are composed of two or more phenotypically similar but, ostensibly, genetically divergent closely-related species (Struck et al., 2018). The rise in cryptic species discoveries is largely driven by the expansive use of molecular data and new methods for analyses of increasingly large datasets, which have enabled us to elucidate genetic structure at an unprecedented geographic scale, depth, and resolution. However, most new cryptic species have, to date, been identified or described using tree- and distance-based methods, which rely on phylogenetic arrangements and genetic divergence thresholds (Brown & Stuart, 2012; Fišer, Robinson, & Malard, 2018; Hillis, 2019). This is disconcerting because an increasing number of genomic studies are now demonstrating that phylogenetic estimation (and by implication, most downstream species delimitation inferences) can be biased or misled by factors such as incomplete lineage sorting and gene flow (e.g. Jones, 2018; Leaché et al., 2015; Linkem, Minin, & Leaché, 2016; Long & Kubatko, 2018; Mendes & Hahn, 2018; Roch, Nute, & Warnow, 2019; Xu & Yang, 2016),

thereby obfuscating the distinction between population structure and species divergence (Chan et al., 2017; Drillon, Dufresnes, Perrin, Crochet, & Dufresnes, 2019; Harrison & Larson, 2014; Luo, Ling, Ho, & Zhu, 2018; Maguilla & Escudero, 2016; McFadden et al., 2017; Morales & Carstens, 2018; Quattrini et al., 2019; Supple, Papa, Hines, McMillan, & Counterman, 2015; Surveswaran, Gowda, & Sun, 2018). As such, it remains unclear whether the purportedly high levels of hidden diversity within many cryptic species complexes consist of distinct, undescribed species, or instead, genetically structured metapopulation lineages that are not evolutionarily isolated.

Nevertheless, these confounding factors are usually ignored when delimiting cryptic species, and empirical studies that consider these potentially confounding factors are the exception rather than the norm (Camargo, Morando, Avila, & Sites, 2012; Chambers & Hillis, 2020; Chan et al., 2017; Dufresnes et al., 2020; Morales & Carstens, 2018; Stanton et al., 2019). Therefore, understanding the effects that these processes may have on species delimitation inferences are critical to avoid erroneous estimations of species diversity, particularly in biodiversity hotspots such as Southeast Asia, where cryptic species have been widely interpreted as being responsible for a large portion of this imperilled region's purportedly unrecognized biodiversity (Brown & Stuart, 2012; Inger, Stuart, & Iskandar, 2009; Koh et al., 2013; Sodhi, Koh, Brook, & Ng, 2004; Wilcove, Giam, Edwards, Fisher, & Koh, 2013).

Genomic methods can reveal genetic structure in unparalleled detail (e.g. Benestan et al., 2015; Chan et al., 2017; Lim et al., 2017; Schield et al., 2018), but accurately characterizing species boundaries within a geographically-explicit spatial and evolutionary framework remains challenging. Continuous geographic clines may appear as isolated population clusters if geographic sampling is discontinuous or when the clustering model does not account for confounding spatial processes such as isolation-by-distance (Bradburd, Coop, & Ralph, 2018; Fenderson, Kovach, & Llamas, 2020; Slager et al., 2020; Tonzo, Papadopoulou, & Ortego, 2019). Furthermore, gene flow among populations, and even species, can bias species tree estimation and produce incorrect topologies (Eckert & Carstens, 2008; Edwards, Potter, Schmitt, Bragg, & Moritz, 2016; Ginsberg, Humphreys, & Dyer, 2019; Hahn & Nakhleh, 2016; Hinojosa et al., 2019; Leaché, Harris, Rannala, & Yang, 2014; Solís-Lemus, Yang, & Ané, 2016). These errors can then be exacerbated in downstream species delimitation analyses that are predicated on the species tree, which is assumed to be correct (Talavera, Dincă, & Vila, 2013; Xu & Yang, 2016; Yang & Rannala, 2010). Additionally, performing species delimitation analysis on genome-scale data faces the problem of computational scalability (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012; Fujisawa, Aswad, & Barraclough, 2016; Ogilvie, Heled, Xie, & Drummond, 2016) and distinguishing between population-level structure and species divergence (Barley, Brown, & Thomson, 2018; Chan et al., 2017; Jackson, Carstens, Morales, & O'Meara, 2017; Leaché, Zhu, Rannala, & Yang, 2019; Luo et al., 2018; Sukumaran & Knowles, 2017; Supple et al., 2015).

Most species delimitation methods either disregard gene flow (distance-based methods) or assume that gene flow is absent (e.g. multispecies coalescent methods, MSC; Jackson et al., 2017; Leaché et al., 2019). Consequently, one of the adverse effects of ignoring gene flow is that MSC methods tend to overestimate species numbers by inferring population structure as species divergence (Chambers & Hillis, 2020; Leaché et al., 2019; Luo et al., 2018; Sukumaran & Knowles, 2017; Wagner, Härtl, Vogt, & Oberprieler, 2017; Wagner et al., 2020). To date, few methods jointly estimate and model gene flow into the species delimitation framework; the exceptions, or methods that do characterize gene flow are computationally expensive for larger genomic datasets comprising >3–5 populations (Jackson et al., 2017; Smith & Carstens, 2019). As an alternative, modular approaches that separately test for confounding effects can provide additional independent lines of evidence to differentiate between population and species-level divergence (Chambers & Hillis, 2020; Chan et al., 2017; Dincă, Lee, Vila, & Mutanen, 2019; Dufresnes et al., 2020; Morales & Carstens, 2018; Zheng et al., 2017). Such analyses are not reliant on a single species tree, which can be challenging to estimate accurately (see references above) or may not even be present (Hahn & Nakhleh, 2016). Instead, modular approaches utilize population genetic markers, parameter estimates, or gene trees from thousands of loci, to provide a more unbiased representation of phylogenetic variation (Blischak, Chifman, Wolfe, & Kubatko, 2018; Buerkle, 2005; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014; Leaché et al., 2019). We employed such an approach to infer species boundaries in Southeast Asian Spotted Stream Frogs of the *Pulchrana*

picturata complex, which have been shown to potentially comprise numerous cryptic species (Brown & Siler, 2014).

Currently, *Pulchrana picturata* is considered a single species that exhibits notable but non-discrete (continuous) morphological variation throughout its distribution in southern Thailand, Peninsular Malaysia, Sumatra, and Borneo (Brown & Guttman, 2002; Frost, 2019). High levels of genetic structure and up to 10% mitochondrial divergence (16S rRNA gene) have been detected among strongly-supported and geographically circumscribed clades (Brown & Siler, 2014), suggesting that this complex could comprise multiple cryptic species. Moreover, instead of being nested within the Bornean clade, one population from Borneo was recovered as nested within a separate, Thailand, Peninsular Malaysia, and Indonesia clade with high support (supplementary figure S3 in Brown & Siler, 2014), alluding to the possibility that gene flow may have biased phylogenetic inference in that study using only a handful of loci (Brown & Siler, 2014).

Accordingly, we undertook the present study, using a newly developed target-capture protocol specifically designed for anurans (FrogCap; Hutter et al., 2019) and obtained more than 12,000 informative loci consisting of exons, introns, and ultraconserved elements (UCEs) from representative populations across the distributional range of *P. picturata* to determine whether deep divergences among clades and observed geographically-structured genetic variation correspond with statistically-defensible cryptic species boundaries. Specifically, we test for gene flow among genetically structured populations and assess its effects on phylogenetic and species boundary inferences to determine whether species delimitation based on phylogenetic arrangement and genetic divergence can accurately estimate cryptic species diversity.

Materials and Methods

Sampling and sequencing

Our sampling design is predicated on a Sanger-based molecular phylogenetic analysis with comprehensive geographical sampling by Brown & Siler (2014). Based on their multilocus phylogeny (figure S3 in Brown & Siler, 2014), we strategically selected samples from each notably divergent clade, making sure to include samples from different geographic populations to adequately capture the genomic diversity within this species complex. A total of 24 samples were genotyped using the FrogCap sequence capture marker set (Ranoidea V1 probe set; Hutter et al., 2019) including 10 outgroup samples (*Boophis tephraeomystax, Mantidactylus melanopleura, Cornufer guentheri,* and*Abavorana luctuosa* , *Pulchrana banjarana, P. siberu,* and*P. signata*), and 14 ingroup samples of the *P. picturata*complex from throughout its distribution range in Peninsular Malaysia, Sumatra, and Borneo. For assurances of taxonomic and nomenclatural clarity, we included a sample from the type locality [Mount Kinabalu, Sabah; *sensu* Brown and Guttman's (2002) lectotype designation]. Tissue samples were obtained from the museum holdings of the University of Kansas Biodiversity Institute, Kansas (KU), Field Museum of Natural History, Chicago (FMNH), and La Sierra University Herpetological Collection, California (LSUHC; Table S1). Genomic DNA was extracted using the automated Promega Maxwell® RSC Instrument (Tissue DNA kit) and subsequently quantified using the Promega Quantus® Fluorometer. Library preparation was performed by Arbor Biosciences using the MyBaits v3 protocol and briefly follows: (1) genomic DNA was sheared to 300–500 bp; (2) adaptors were ligated to DNA fragments; (3) unique identifiers were attached to the adapters to later identify individual samples; (4) biotinylated 120mer RNA library baits were hybridized to the sequences for an incubation period of 19 hours and 23 minutes; (5) target sequences were selected by adhering to magnetic streptavidin beads; (6) target regions were amplified via PCR; and (7) samples were pooled and sequenced on an Illumina HiSeq PE-3000 with 150 bp paired-end reads (Hutter et al., 2019). Sequencing was performed at the Oklahoma Medical Research Foundation DNA Sequencing Facility.

Bioinformatics and data filtering

The full bioinformatics pipeline for filtering adapter contamination, assembling markers, and exporting alignments are available at CRH's GITHUB (pipeline V2: https://github.com/chutter/FrogCap-Sequence-Capture). Raw reads were cleaned of adapter contamination, low complexity sequences, and other sequencing artefacts using the program FASTP (default settings; Chen, Zhou, Chen, & Gu, 2018). Next, paired-end reads

were merged using BBMERGE (Bushnell, Rood, & Singer, 2017). Cleaned reads were then assembled *de novo* with SPADES v.3.12 (Bankevich et al., 2012) under a variety of k-mer schemes. Resulting contigs were then matched against reference probe sequences with BLAST, keeping only those that uniquely matched to the probe sequences. The final set of matching loci was then aligned on a marker-by-marker basis using MAFFT.

Alignments were trimmed and saved separately into functional datasets for phylogenetic analyses and data type comparisons. These datasets include (1) Exons: each alignment was adjusted to be in an open-reading frame and trimmed to the largest reading frame that accommodated >90% of the sequences; alignments with no clear reading frame were discarded; (2) Introns: each previously delimited exon was trimmed out of the original contig and both remaining intronic regions were concatenated; (3) Exons-combined: exons from the same gene were concatenated and treated as a single locus (justifiable under the assumption that as they might be linked); and (4) UCEs. We applied internal trimming to the intron and UCE alignments using the program trimAl (automatic1 function; Capella-Gutiérrez et al., 2009). All alignments were externally trimmed to ensure that at least 50 percent of the samples had sequence data present at the alignment edges.

In addition to analysing the unfiltered datasets, we also filtered the data by removing loci with low phylogenetic information, which can introduce noise and increase systematic bias (Mclean, Bell, Allen, Helgen, & Cook, 2019). We used parsimony-informative-sites (PIS) as a proxy for hierarchical structure and phylogenetic information; and removed the lower 50% of loci that contained the least PIS. All datasets were analysed separately to assess phylogenetic congruence. Summary statistics, partitioning, and concatenation of data were performed using the program AMAS (Borowiec, 2016) and custom R scripts.

SNP extraction

To obtain variant data across the target samples, we used GATK v4.1 (McKenna et al., 2010) and followed the recommended best practices when discovering and calling variants (Van der Auwera et al., 2013), using a custom R pipeline available on Carl R Hutter's GitHub (https://github.com/chutter/FrogCap-Sequence-Capture). To discover potential variant data (e.g. SNPs, InDels), we used a consensus sequence from each alignment from the target group as a reference and mapped the cleaned reads back to the reference markers from each sample. We used BWA ("bwa mem" function; Li, 2013) to map cleaned reads to the reference markers, adding the read group information (e.g. Flowcell, Lane, Library) obtained from the fastq header files. We next used SAMTOOLS (H. Li et al., 2009) to convert the mapped reads SAM file to a cleaned BAM file, and merged the BAM file with the unmapped reads as required to be used in downstream analyses. We used the program PICARD to mark exact duplicate reads that may have resulted from optical and PCR artifacts and reformatted the dataset for variant calling. To locate variant and invariant sites, we used GATK4 to generate a preliminary variant dataset using the GATK program *HaplotypeCaller* to call haplotypes in the GVCF format for each sample individually.

After processing each sample, we used the GATK *GenomicsDBImport*program to aggregate the samples from the separate datasets into their own combined database. Using these databases, we used the*GenotypeGVCF* function to genotype the combined sample datasets and output separate ".vcf" files for each marker that contains variant data from all the samples for final filtration. Next, to filter the .vcf files to high quality variants, we used the R package *vcfR* (Knaus & Grünwald, 2017) and selected variants to be used in downstream analyses that had a quality score > 20, and we also filtered out the top and bottom 10% of variants based on their depth and mapping quality to avoid potentially problematic sites.

Phylogenetic estimation and discordance

We used Maximum Likelihood (ML) analysis of concatenated data and coalescent-based methods for phylogenetic estimation. For our ML analysis, we used the program IQ-TREE v1.6 (Chernomor, Von Haeseler, & Minh, 2016; Nguyen, Schmidt, Von Haeseler, & Minh, 2015) and, because of the unprecedented number of loci retrieved with FrogCap, we performed an unpartitioned analysis using the GTR+GAMMA substitution model. Branch support was assessed using 5,000 ultrafast bootstrap replicates (UFB; Hoang et al., 2017) and nodes with UFB >95 were considered strongly-supported. A summary-based species tree analysis

was performed using ASTRAL-III (Zhang, Rabiee, Sayyari, & Mirarab, 2018) because this approach has one of the lowest error rates when the number of informative sites are high and has also been shown to produce more accurate results compared to other summary methods under a variety of conditions including high levels of incomplete lineage sorting (ILS) and low gene-tree estimation error (Davidson, Vachaspati, Mirarab, & Warnow, 2015; Mirarab et al., 2014; Molloy & Warnow, 2017; Ogilvie et al., 2016; Vachaspati & Warnow, 2015, 2018). As input for our ASTRAL analysis, individual marker gene trees were estimated using IQ-TREE, with the best-fit substitution model for each locus determined by the program ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017). Because species boundaries have not been adequately characterized, individual samples were not assigned to species. Finally, the same set of gene trees was used to estimate species trees using the distance-based method ASTRID, which has been shown to outperform ASTRAL when many genes are available and when ILS is very high (Vachaspati & Warnow, 2015).

Phylogenetic analyses were performed separately on the Intron, Exon, Exons-combined, and UCE datasets and we used the program DiscoVista (Sayyari, Whitfield, & Mirarab, 2018) to assess phylogenetic discordance by comparing the relative frequencies of all three topologies surrounding a particular focal branch, in instances in which topological discordance was observed in summary species-tree procedures.

Species delimitation framework

We used a two-step approach to species delimitation, involving independent "discovery" and, subsequent "validation" stages (Hillis, 2019). For our discovery stage, putative evolutionary lineages were inferred from mitochondrial haplotypes derived from originally-inferred, strongly-supported multilocus inferences (Brown & Siler, 2014) and reanalysis of 16S rRNA data in this study. We then used sequence-based (Automatic Barcode Gap Discovery, ABGD; Puillandre, Lambert, Brouillet, & Achaz, 2012) and phylogeny-based (Multi-rate Poisson tree processes, mPTP; Kapli et al., 2017) species delimitation methods to infer putative species boundaries. These single-locus methods have been shown to be effective at delimiting candidate species with uneven sampling (Blair & Bryson, 2017). We used default settings for the ABGD analysis and estimated a maximum likelihood phylogeny with IQ-TREE based on the 16S marker, to use as input for the mPTP analysis. The minimum branch length was automatically detected using the *minbr_auto* function. Two MCMC chains were executed using 10,000,000 iterations with samples saved every 50,000 iterations. Finally, for comparison with previous studies, we examined mitochondrial divergences among reciprocally monophyletic putative species pairs, by comparing distributions of uncorrected p-distances. Putative species were then validated using genomic data, which are explained in detail below.

Population clustering.— We performed dataset dimension-reduction analysis on our SNP dataset to infer and visualize population clusters which might correspond to inferred putative species. A principal component analysis (PCA) was performed to obtain an orthogonal linear transformation of the data using the R package*adegenet* (Jombart & Ahmed, 2011). Additionally, a t-Distributed Stochastic Neighbour-Embedding (t-SNE) method was used to reveal structure at multiple scales (van der Maaten & Hinton, 2008). The t-SNE method is an improvement to traditional linear dimensional reduction methods such as PCA and multidimensional scaling because it is non-linear and is better at capturing structure and presence of clusters in high-dimensional data (W. Li, Cerise, Yang, & Han, 2017; van der Maaten & Hinton, 2008). The t-SNE analysis was performed using the R package *Rtsne* (Krijthe, 2015) under the following settings: dims=3, perplexity=5, theta=0.0, max_iter=1000000.

Population structure.— Next, we examined population structure by calculating ancestry coefficients using a program based on sparse nonnegative matrix factorization (sNMF). This method is comparable to other widely-used programs such as ADMIXTURE and STRUCTURE, but is computationally faster and robust to departures from traditional population genetic model assumptions such as Hardy-Weinberg equilibrium (Frichot et al., 2014). Ancestry coefficients were estimated for 1–10 ancestral populations (K) using 100 replicates for each K. The cross-entropy criterion was then used to determine the best K based on the prediction of masked genotypes. The sNMF analysis was implemented through the R package LEA (Frichot & François, 2015).

Non-spatial clustering methods including sNMF, STRUCTURE, and ADMIXTURE assume that allele frequencies of individuals within a cluster are equal, regardless of their geographic location. This assumption does not account for differentiation caused by continuous processes, such as isolation-by-distance (IBD) and can, consequently, overestimate the number of discrete clusters, especially when geographic sampling is sparse—as is the case, in many empirical studies (Prunier et al., 2013). Therefore, we also performed a spatially-aware model-based clustering analysis (*conStruct*), which also considers IBD as an explanation for genetic variation (Bradburd et al., 2018). We used all ingroup samples and one outgroup taxon (*Pulchrana signata*) for this analysis. The same SNP dataset was used to represent allele frequencies, and geographic coordinates for each sample were converted into a pairwise great-circle distance matrix using the R package*fields* (Nychka, Furrer, Paige, & Sain, 2017). Our*conStruct* analysis was performed with spatial and non-spatial models, using 200,000 MCMC iterations; traceplots were examined to assess convergence. A cross-validation approach was then used to compare different K values between spatial and non-spatial models. Posterior distributions of parameters were estimated using a training partition consisting of 90% randomly selected loci. The predictive accuracy of each value of K was then measured using log-likelihoods of the remaining loci, averaged over the posterior. A total of 8 replicates were used to assess each value of K.

To confirm whether IBD contributed to genetic variation, we implemented a distance-based redundancy analysis (dbRDA), which has been shown to be an improvement over traditional Mantel tests because it uses a principal coordinates analysis to linearize the response variable, thereby removing violations of linearity (Guillot & Rousset, 2013; Kierepka & Latch, 2015). Genetic distances were represented by pairwise population G_{st} (Nei, 1973), which was calculated using the R package *mmod* (Winter, 2012). Geographic distances were transformed into distance-based Moran's eigenvector maps (dbMEM) and used as an independent variable (Legendre, Fortin, & Borcard, 2015). The dbRDA analysis was then performed using the *capscale* function in the R package *vegan* (Oksanen et al., 2017). Statistical significance was assessed using 999 permutations.

Gene flow.— Admixture among populations was confirmed using Bayesian hybrid-index analysis and the python program HyDe. A hybrid-index analysis calculates the proportion of allele copies originating from parental reference populations (Buerkle, 2005), whereas a HyDe analysis detects hybridization using phylogenetic invariants based on the coalescent model with hybridization (Blischak et al., 2018). Different combinations of plausible parental populations were tested, based on results from our population structure and preliminary species delimitation analyses. We implemented the hybrid-index analysis on our SNP dataset using the R package *gghybrid* (Bailey, 2018) after removing loci with a minor allele frequency >0.1 in both parental reference sets. A total of 10,000 MCMC iterations were performed with the first 50% discarded as burnin. The HyDe analysis was performed on sequence data from the intron dataset. First, admixture at the population level was assessed using the *run_ hyde* script that analyses all possible four-taxon configurations consisting of an outgroup (*Pulchrana signata*) and a triplet of ingroup populations comprising two parental populations (P1 and P2) and a putative hybrid population (Hyb). Next, analysis at the individual level was performed using the *individual_ hyde* script to detect hybridization in individuals within populations that had significant levels of genomic material from the parental species. Finally, we performed bootstrap resampling (500 replicates) of individuals within hybrid populations to obtain a distribution of gamma values to assess heterogeneity in levels of gene flow.

Genealogical divergence index.— Finally, we used the genealogical divergence index (*gdi*) to determine whether putative species boundaries corresponded to species-level divergences (Chan & Grismer, 2019; Leaché et al., 2019). First, an A00 analysis in BPP was used to estimate the parameters τ and ϑ with the *thetaprior* = 3 0.002 e and *tauprior* = 3 0.004 (Flouri, Jiao, Rannala, Yang, & Yoder, 2018). Species assignments were based on putative species boundaries inferred from the discovery step. Because BPP performs best on neutrally evolving loci, we conducted the analysis only on our intron dataset. For the analysis to be computationally tractable, we further filtered these data to include only loci with full taxon representation (1,515 loci). Two separate runs were performed (100,000 MCMC iterations each) and converged runs were concatenated to generate posterior distributions for the multispecies coalescent parameters that were used subsequently to calculate *gdi* following the equation: $gdi = 1 - e^{-2\tau/\vartheta}$ (Jackson et al., 2017; Leaché et al., 2019). Population A is distinguished from population B using the equation $2\tau_{AB}/\vartheta_A$, whereas $2\tau_{AB}/\vartheta_B$ is

used to differentiate population B from population A. Populations are considered distinct species when *gdi* values are >0.7, and low *gdi*values (<0.2) indicate two populations belong to the same species. Values of 0.2> *gdi* < 0.7 indicate ambiguous species status (Jackson et al., 2017; Pinho & Hey, 2010).

Results

Data collection, phylogeny estimation, and topological discordance

Summary statistics for retained loci are presented in Table 1. In general, almost 12,000 intronic and exonic markers were obtained; UCEs numbered 625 and were on average the longest (713 bp), whereas exons were shortest (212 bp). After exons from the same gene were identified and combined, a total of 2,186 markers remained (average length 617 bp). Introns exhibited the most informative sites, with more than 2.6 million variable sites and over 950,000 PIS (Table 1).

Two different topologies (T1 and T2) were obtained across all phylogenetic analyses and datasets (Fig. 1). In general, regional populations (Peninsular Malaysia, Sumatra, Borneo) formed highly supported clades except for two Bornean samples (ND 7479 and ND 7056 from Sarawak), which we designated as putative hybrids (H1 and H2; Fig. 1) based on their anomalous placement within the Peninsular Malaysia + Sumatra clade. For most datasets (Exons-combined, Introns, UCEs), these two samples were recovered as the first-branching lineages within the Peninsular Malaysia + Sumatra clade, with high support across all analyses (topology T1; Fig. 1). However, for the Exon dataset, one of those samples (ND 7479) was recovered as the first-branching lineage of the Bornean clade, with high support across all analyses (topology T2; Fig. 1). Complete details of all phylogenetic trees from analyses of each datasets are provided in Supplementary material.

The relative frequency of alternate topologies surrounding a discordant branch revealed that the number of gene trees supporting the main topology was only slightly more (<3%) than those supporting an alternate topology, indicating a high level of discordance and a lack of overwhelming support for a particular topology (Fig. 2). These outcomes were most evident in datasets that had relatively fewer markers (Exons-combined, 2,186; UCEs, 625) and in which the primary topology was supported by not more than 20 additional gene trees.

Putative species boundaries

The topology of the mitochondrial phylogeny estimated for the mPTP analysis was the same as the topology from analyses of our Exons dataset (topology T2; Figs. 1, 3A). Excluding the outgroup (*Pulchranasignata*), the mPTP analysis inferred a total of five species (Fig. 3A). The first species (Sp1) comprised samples from Peninsular Malaysia, Sumatra, and one of the putative hybrids (Hybrid 1; ND 7056 from Sarawak). Putative species Sp2 included samples from Sabah, Borneo (FMNH 230864 from Lahad Datu ND 8281 from Tawau; Fig. 1), which were the sister lineage to True *P. picturata* (exemplified by topotype ID 7750 from Mount Kinabalu, Sabah). Other Bornean populations were split into two distinct clades but these were not strongly-supported as distinct species (average support value 0.62) and were therefore considered a single putative species (Sp3). Our mPTP analysis also delimited the putative Hybrid 2 as a distinct species with strong support. These five putative species (True *P. picturata* , Hybrid 2, Sp1, Sp2, Sp3) were also delimited by the ABGD analysis, again with strong support. A comparison of mitochondrial p-distances showed that the level of divergences within Sp1 (including Hybrid 1) and Sp3 were relatively low at [?] 3% (Fig. 3B); in comparison, divergences among putative species were high (>5%).

Validation using genomic data

Population structure.— A total of 11,490 SNPs were obtained and used for clustering (PCA, t-SNE), population structure (sNMF, conStruct), and gene flow (Bayesian hybrid index, HyDe) analyses. In our PCA analysis, the outgroup (*Pulchrana signata*) and populations from Peninsular Malaysia and Sumatra formed two distinct clusters that were distantly separated—markedly more so than Bornean populations, which showed less separation (Fig. 4A). The t-SNE analysis showed similar results but with more diffusion within clusters (Fig. 4B).

The cross-entropy criterion of the sNMF analysis inferred K=3 and K=4 as the best-predicted numbers of ancestral populations, with K=3 being only marginally better (Fig. 4C). At K=3, populations from Peninsular Malaysia and Sumatra (Sp1) were clustered as a single population with no admixture (Fig. 5A). Similarly, populations from far east Borneo (True*P. picturata* + Sp2) also formed a single, non-admixed cluster. Other Bornean populations (Sp3, Hybrid 1, Hybrid 2) exhibited a cline of admixture with the two putative hybrid samples being the most admixed. At K=4, the putative hybrid samples were characterized as highly admixed and Sp3 formed a distinct non-admixed group (Fig. 5A).

The *conStruct* analysis also inferred K=3 and K=4 as ideal numbers of ancestral populations, with K=4 slightly better. Model comparison demonstrated that the spatial model fitted the data slightly better than the non-spatial model at K=3, but the two had similar scores at K=4 (Fig. 4D). This was corroborated by the dbRDA analysis (p -value = 0.2736; R^2 = 0.2236), indicating that IBD was not a significant factor affecting genetic variation. In general, these assignments of individuals to population clusters were similar to results from the sNMF analysis, but with higher levels of admixture (Fig. 5B). Notably, Sp1 was inferred to contain low levels of admixture from Bornean genotypes and Bornean populations were inferred to be more admixed.

Our True *Pulchrana picturata* clade and Sp2 showed relatively high levels of admixture, whereas Sp3 had dissimilar levels of admixture. One Sp3 sample from far west Borneo was considerably admixed, while the other two samples from east Borneo were not (Fig. 5B).

Gene flow and species delimitation.— Based on results from our population clustering and structure analyses, we inferred Sp1 and either Sp3 or True *P. picturata*+ Sp2 to be potential parental populations, due to their dominant representation in ancestry coefficients. When Sp1 and True *P. picturata* +Sp2 were designated as parental references, the genome of Sp3 and the putative hybrid samples showed a mixture of alleles from both parent taxa (Fig. 6A). A similar result was achieved when Sp1 and Sp3 were designated as parental populations and, in both scenarios, the hybrid index of the putative hybrids was considerably higher (Fig. 6B).

The HyDe analysis at the population level produced a similar, but more nuanced, characterization of hybridization. Using different ingroup configurations, significant hybridization was detected in all Bornean populations (Table 2). The Sp2 population exhibited the lowest level of hybridization (Gamma=0.9), whereas Hybrid and Sp3 populations displayed moderate to high levels of hybridization (Gamma=0.2–0.8). Furthermore, this analysis showed that hybridization was not limited to Sp1 and True*P. picturata* as parental populations, but also between Hybrid/True *P. picturata* , Sp1/Sp2, Sp1/Sp3, Hybrid/Sp2, and True*P. picturata* /Sp3. Analysis at the individual level showed the Hybrid population to be a mixture of True *P. picturata* , Sp1, Sp2, and to a lesser extent Sp3, whereas individuals from Sp3 were a mixture of True *P. picturata* , Hybrid, Sp2, and Sp1. Individuals from Sp2 were the least admixed (Gamma=0.9; Table 2).

Our *gdi* analysis was performed on a reduced subset of 1,515 loci, but with full taxon representation. Additionally, to avoid bias, two putative hybrid samples were removed from this dataset due to their phylogenetic uncertainty and high levels of gene flow. Our results indicate that populations from Peninsular Malaysia and Sumatra (Sp1) are a distinct species, supported by high confidence (Fig. 7C; mean*gdi* =0.91). However, the specific status of all other populations (those from Borneo) remain uncertain (mean *gdi P. picturata* = 0.59; Sp2 = 0.57; Sp3 = 0.55), and so we conservatively consider them conspecific at the present time.

Because our results revealed high levels of gene flow among multiple populations, we also inferred a phylogenetic network that accounts for ILS and hybridization using the program PhyloNet v.3.8 (Wen, Yu, Zhu, & Nakhleh, 2018). To facilitate computation, we used all 625 single-locus gene trees from the UCE dataset (outgroups removed) to infer a species network using the Minimizing Deep Coalescence (MDC) criterion, with the maximum number of reticulations set to five. A total of five runs were performed and all other parameters were set to default values. The best inferred network was congruent with results from the *conStruct*and HyDe analyses and provided deeper insights at the individual level. Gene flow was detected among most Bornean populations and specifically, between H1 and a sample from Sumatra (FMNH 266944).

Gene flow involving the H1 and H2 samples were also older compared to gene flow among other Bornean populations (Fig. 7). More nuanced admixture was not detected by the sNMF analysis, suggesting that sNMF may not be as efficient at detecting fine-scale admixture when sampling is discontinuous.

Discussion

Confounding effects of gene flow

Our results showed that gene flow/introgression can produce confounding phylogenetic and divergence patterns that can be positively misleading when analyzed using conventional species delimitation procedures. Two of the most highly introgressed hybrids (Hybrid 1 and Hybrid 2) were from Borneo but were inferred in most analyses as independent lineages that were more closely related to the Peninsular Malaysia + Sumatra clade to the exclusion of the Bornean clade. Consequently, the hybrid samples were highly divergent from adjacent Bornean populations (7–10% mitochondrial divergence), but remarkably similar (<3%) to allopatric populations from Peninsular Malaysia and Sumatra (Hybrid 1). High mitochondrial divergence could be due to mitochondrial gene flow, a phenomenon where introgressed mitochondrial DNA from another species reflects past introgressive events as opposed to lineage isolation (Ballard & Whitlock, 2004; Linnen & Farrell, 2007; Ruane, Bryson, Pyron, & Burbrink, 2014). Using a more robust population genomics approach, we showed that the genomic makeup of the hybrid samples contained relatively equal proportions of alleles from both Borneo and Peninsular Malaysia/Sumatra (Sp 1) lineages. These results also provide an alternative explanation for the conundrum of highly divergent (sometimes non-sister) sympatric/parapatric lineages–a pattern that has been celebrated as an archetypal sign of genuine cryptic speciation (Brown, 2015; Cobos et al., 2016; Cooke, Chao, & Beheregaray, 2012; Grismer et al., 2015; Ladner & Palumbi, 2012; McLeod, 2010). Such anomalous patterns are not uncommon in amphibians and are present in virtually every Southeast Asian frog family that has been touted to harbor pronounced cryptic diversity: Bufonidae (Chan & Grismer, 2019), Dicroglossidae (Matsui et al., 2016; McLeod, 2010), Ichthyophiidae (Nishikawa et al., 2012), Megophryidae (J. M. Chen et al., 2018, 2017; Rowley et al., 2015), Ranidae (Lu, Bi, & Fu, 2014; Stuart, Inger, & Voris, 2006), and Rhacophoridae (Chan, Grismer, & Brown, 2018; Poyarkov et al., 2015). Our results demonstrate that high levels of genetic divergence between sympatric lineages could be an artefact of introgression as opposed to divergence via natural selection.

Although distinct, highly divergent sympatric or parapatric cryptic species do undoubtedly exist (Pulido-Santacruz, Aleixo, & Weir, 2018), they usually consist of relatively old lineages that (1) are highly fragmented and whose phylogeographic structure was facilitated by cyclical climatic fluctuations (repeated contraction and expansion of refugia; Grismer et al., 2015); (2) diverged in isolation, followed by subsequent secondary contact (Chan & Brown, 2019); or (3) exhibit varying levels of niche partitioning, for example through contrasting phenologies (Amato et al., 2007; Scriven, Whitehorn, Goulson, & Tinsley, 2016) or small-scale habitat segregation (Muangmai, Von Ammon, & Zuccarello, 2016). However, the purported existence of high numbers of undescribed sympatric/parapatric cryptic species in numerous Southeast Asian amphibian complexes are mostly represented by relatively young (<5 million years old; e.g. Chen et al., 2018, 2017), widespread, and continuously occurring lineages that are ecologically similar (see references above). Taking these characteristics into account (and disregarding the possibility of sympatric speciation, which remains a controversial and hotly debated topic; Foote, 2018), we hypothesize that most recently diverged and purportedly cryptic species correspond to incipient species in the grey zone of speciation (see below), and that the majority of young and highly divergent sympatric lineages (e.g. Brown, 2015; Garrick, Dickinson, Reppel, & Yi, 2019; Giska, Sechi, & Babik, 2015; McLeod, 2010) can be explained by introgression. Therefore, a re-analysis of such cases using more robust methods that assesses spatial population structure and gene flow is warranted.

Cryptic species as a window on diversity—or slippery slope towards taxonomic inflation?

According to the most widely-adopted definition, cryptic species are (1) genetically but not morphologically distinguishable; and (2) are, or have been, classified as a single nominal species (Bickford et al., 2007). Other researchers have specified that cryptic species should also be recently diverged, occur in sympatry, or exhibit

reproductive isolation (Chenuil, Cahill, Délémontey, du Luc, & Fanton, 2019; Struck et al., 2018). Viewing speciation as a continuous, gradual, and protracted process (Rosindell, Cornell, Hubbell, & Etienne, 2010; Sukumaran & Knowles, 2017), recently diverged lineages that are morphologically similar but genetically divergent may also be associated with the "grey zone" of the speciation continuum—a region of diversification in which there is conflict among operational species criteria (de Queiroz, 2005; Roux et al., 2016). Early diverging lineages in the grey zone can be referred to as incipient species, and we suspect that many, if not most, previously-identified cryptic species may fall within this category. These are lineages that have begun to diverge but still exchange genes or maintain signatures of recent gene flow (Marques et al., 2016; Schield et al., 2015; Supple et al., 2015). At this stage of speciation, species boundaries are ephemeral and incipient species can continue to diverge and eventually form distinct species (complete reproductive isolation), or merge back into a single species (Feder, Egan, & Nosil, 2012; Harrison & Larson, 2014; Mallet, 2008). Therefore, it is critical for cryptic species delimitation to be scrutinized for evidence that lineages are on diverging trajectories of ancestor–descendant series of populations, among which independent lineage status or cohesion has been achieved via postzygotic incompatibilities (Pulido-Santacruz et al., 2018) or prezygotic isolation mechanisms such as ecographic segregation (Dufresnes et al., 2020; Slager et al., 2020; Sobel & Streisfeld, 2015), environmental adaptation (Rundle & Nosil, 2005), and behavioural/mate recognition differentiation (Boake, Andreadis, & Witzel, 2000; Drillon et al., 2019; Köhler et al., 2017). These criteria are more robust, compatible with evolutionary theory and species concepts, and reflective of lineage separation; and thus, should be included as part of a more informed, modern, multidisciplinary statistical species delimitation framework to avoid unnecessary taxonomic inflation.

Our study also suggests that hybridization (ancient, intermittent or ongoing) may play a significant role in the evolutionary history and biodiversity of "cryptic" species (Taylor & Larson, 2019), particularly in the Sunda region where large landmasses and island archipelagos have been periodically connected and separated due to climatic changes or geological events (Hall, 2013; Yumul, Dimalanta, Marquez, & Queaño, 2009; Yumul et al., 2004). Advancements in high-throughput sequencing has enabled us to move beyond classical criteria for species delimitation such as phylogenetic arrangements and divergence thresholds, and build towards a more process-based understanding of how cryptic species boundaries are formed and maintained (Smith & Carstens, 2019; Struck et al., 2018). This includes critical questions such as: (1) how prevalent is hybridization in phylogeographically structured species complexes? (2) how does hybridization affect species boundaries and biodiversity estimates? and (3) is hybridization context dependent? (hybrid zones facilitated by landscape features or temporally induced by intermittent habitat corridors during past climatic/geologic events). Although quantifying biodiversity is crucial to many fields in biology and conservation, it is becoming increasingly evident that tree- and distance-based criteria are poor proxies for species divergence when gene flow is present, and that general/global thresholds (Fouquet et al., 2007; Vieites et al., 2009) should not be used to justify the discovery of new species.

Systematics and biogeography

All analyses in both discovery and validation steps showed a clear distinction between populations from Borneo (True *P. picturata* , Sp2, Sp3, H1, and H2) and Peninsular Malaysia + Sumatra (Sp1). This was further corroborated by the *gdi* analysis that inferred Sp1 as a distinct species from the true *P. picturata* from Borneo. It is also noteworthy that although the H1 and H2 samples were inferred as sister lineages to Sp1 and were paraphyletic with regard to other Bornean populations, this phylogenetic configuration was not in agreement with patterns of spatial and genetic structure inferred from genomic validation analyses. The phylogenetic network analysis provided additional insight by showing that the two hybrid samples introgressed specifically with a population from Sumatra (FMNH 266944). This explains the anomalous phylogenetic placement of H1 and H2 and indicates that their inferred affinity to the Sp1 clade is due to genetic admixture as opposed to shared ancestry. According to our results, both H1 and H2 samples should be considered part of the Bornean clade. However, doing so would render the Bornean lineage paraphyletic, which highlights the importance of excluding hybrid samples from species delimitation (and probably species tree) analyses and, additionally, exposes the limitation of phylogeny-based classification that can yield paraphyletic groups in the presence of gene flow (Kumar et al., 2017; Ma et al., 2017). However, it is also

possible that additional sampling, especially from central and southern Sumatra could change the species tree topology. In any case, our results unequivocally demonstrate the presence of marked gene flow among populations from Borneo and Sumatra, and the absence of gene flow with populations from Peninsular Malaysia. Additionally, the phylogenetic network analysis also showed that gene flow involving the H1 and H2 samples were older, as opposed to more recent/ongoing gene flow among the other Bornean populations. This spatio-temporal pattern of gene flow alludes to an ancestral admixture event(s) facilitated by a more southerly habitat corridor, probably along the Karimata Strait via the Bangka-Belitung arc as opposed to a more northerly route through the Riau Archipelago (Fig. 7)—a pattern that has also been documented in numerous other vertebrate groups (Inger & Voris, 2001; Mason, Helgen, & Murphy, 2019; Nijman & Nekaris, 2010). Although we were unable to estimate the timing of diversification in this study, a previous study estimated the diversification of major clades in the *Pulchrana signata/picturata* complex during the Miocene and Pliocene (Chan & Brown, 2017), during which there was land connection between Borneo and Sumatra, and before the fragmentation of these land masses at the onset of the Pleistocene (Hall, 2013). Therefore, it is likely that ancient introgression between Sumatra and Borneo lineages occurred during the Miocene and Pliocene and that the cessation of gene flow (and subsequent allopatric diversification) was caused by the inundation of land bridge corridors during the Pleistocene. Subsequently. cyclical Pleistocene glaciations exposed intermittent land bridges, which could have re-established gene flow. This is congruent with the spatio-temporal patterns of introgression inferred by our phylogenetic network analysis, which showed the occurrence of ancient as well as more recent introgression.

In summary, all lines of evidence indicate that at least three distinct evolutionary lineages are present within the *Pulchranapicturata* complex. One lineage comprises populations from Peninsular Malaysia and Sumatra, and two others occur in Borneo. These lineages exchanged genes in the past and Bornean lineages also exchanged genes with each other, generating a hybrid swarm that created a potentially confusing mirage of what might have been interpreted as "cryptic" species given the inferred phylogenetic structure and high genetic divergence among admixed populations. However, despite high levels of genetic structure and divergence, our results clearly demonstrated that Bornean populations/lineages cannot be unambiguously distinguished from one another morphologically or genetically. Furthermore, they occur continuously across the landscape with no evidence of pre- or postzygotic isolation and thus, should be considered a single species under the name *Pulchrana picturata* . Conversely, Sp1 is sufficiently divergent from Bornean lineages and is physically prevented from exchanging genes with them. Therefore, considering allopatry as a proxy for reproductive isolation or lineage separation, in conjunction with sufficient genetic distinction and reduced gene flow, Sp1 from Sumatra and Peninsular Malaysia can be considered a distinct species.

Acknowledgements

We thank the University of Kansas', Office of the Provost Research Investment Council (RIC Level II Award No. 2300207, to RMB and R. G. Moyle), and KU's Docking Scholar Fund for support to RMB; and KU's Genome Sequencing Core support to CRH and RMB; we also acknowledge U.S. National Science Foundation GRF support to CRH (1540502, 1451148, and 0907996), and DEB 1654388, 1557053, 0743491 to RMB. We thank I. Das (U. Malaysia, Sarawak), A. Resetar, H. Voris, and the late R. Inger (FMNH) for access to genetic resources. This paper is contribution number 923 of the Auburn University Museum of Natural History.

References

Amato, A., Kooistra, W. H. C. F., Levialdi Ghiron, J. H., Mann, D. G., Pröschold, T., & Montresor, M. (2007). Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* ,*158* (2), 193–207. doi: 10.1016/j.protis.2006.10.001

Bailey, R. I. (2018). gghybrid: Evolutionary analysis of hybrids and hybrid zones. *R Package v. 0.0.0.9* . Retrieved from https://github.com/ribailey/gghybrid

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology* , *13* (4), 729–744. doi: 10.1046/j.1365-294X.2003.02063.x

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* , *19* (5), 455–477. doi: 10.1089/cmb.2012.0021

Barley, A. J., Brown, J. M., & Thomson, R. C. (2018). Impact of Model Violations on the Inference of Species Boundaries under the Multispecies Coalescent. *Systematic Biology* , *67* (2), 269–284. doi: 10.1093/sysbio/syx073

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular Ecology* , *24* (13), 3299–3315. doi: 10.1111/mec.13245

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* ,*22* (3), 148–155. doi: 10.1016/j.tree.2006.11.004

Blair, C., & Bryson, R. W. (2017). Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: *Phrynosoma*). *Molecular Ecology Resources* , *17* (6), 1168–1182. doi: 10.1111/1755-0998.12658

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A python package for genome-scale hybridization detection. *Systematic Biology* , *67* (5), 821–829. doi: 10.1093/sysbio/syy023

Boake, C. R. B., Andreadis, D. K., & Witzel, A. (2000). Behavioural isolation between two closely related Hawaiian *Drosophila*species: The role of courtship. *Animal Behaviour* , *60* (4), 495–501. doi: 10.1006/anbe.2000.1509

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* , *4* , e1660. doi: 10.7717/peerj.1660

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics* , *210* , 33–52. doi: 10.1534/genetics.118.301333

Brown, R. M. (2015). A new species of stream frog of the genus*Hylarana* from the mountains of southern Mindanao Island, Philippines. *Herpetologica* , *71* (3), 223–233. doi: 10.1655/herpetologica-d-14-00075

Brown, R. M., & Guttman, S. I. (2002). Phylogenetic systematics of the Rana signata complex of Philippine and Bornean stream frogs: reconsideration of Huxley's modification of Wallace's Line at the Oriental–Australian faunal zone interface. *Biological Journal of the Linnean Society* , *76* , 393–461.

Brown, R. M., & Siler, C. D. (2014). Spotted stream frog diversification at the Australasian faunal zone interface, mainland versus island comparisons, and a test of the Philippine "dual-umbilicus" hypothesis. *Journal of Biogeography* ,*41* (1), 182–195. doi: 10.1111/jbi.12192

Brown, R. M., & Stuart, B. L. (2012). Patterns of biodiversity discovery through time: an historical analysis of amphibian species discoveries in the Southeast Asian mainland and adjacent island archipelagos. In D. J. Gower, K. Johnson, J. Richardson, B. Rosen, L. Ruber, & S. Williams (Eds.), *Biotic Evolution and Environmental Change in Southeast Asia* (pp. 348–389). Cambridge: Cambridge University Press.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. a., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* , *29* (8), 1917–1932. doi: 10.1093/molbev/mss086

Buerkle, C. A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* , *5* (3), 684–687. doi: 10.1111/j.1471-8286.2005.01011.x

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PLoS ONE* , *12* (10), 1–15. doi: 10.1371/journal.pone.0185056

Camargo, A., Morando, M., Avila, L. J., & Sites, J. W. (2012). Coalescent-based methods with ABC and other coalescent-based methods : a yest of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata : Liolaemidae). *Evolution* , *66* (9), 2834–2849. doi: 10.5061/dryad.4409k652

Capella-Gutierrez, S., Silla-martinez, J. M., & Gabaldon, T. (2009). trimAl : a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* , *25* (15), 1972–1973. doi: 10.1093/bioinformatics/btp348

Chambers, E. A., & Hillis, D. M. (2020). The Multispecies Coalescent Over-Splits Species in the Case of Geographically Widespread Taxa.*Systematic Biology* , *69* (1), 184–193. doi: 10.1093/sysbio/syz042

Chan, K. O., Alexander, A. M., Grismer, L. L., Su, Y.-C., Grismer, J. L., Quah, E. S. H., & Brown, R. M. (2017). Species delimitation with gene flow: a methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs.*Molecular Ecology* , *26* , 5435–5450. doi: 10.1111/mec.14296

Chan, K. O., & Brown, R. M. (2017). Did true frogs 'dispersify'?*Biology Letters* , *13* (8), 20170299. doi: 10.1098/rsbl.2017.0299

Chan, K. O., & Brown, R. M. (2019). Elucidating the drivers of genetic differentiation in Malaysian torrent frogs (Anura : Ranidae : Amolops): a landscape genomics approach. *Zoological Journal of the Linnean Society* , *zlz151* .

Chan, K. O., & Grismer, L. L. (2019). To split or not to split? Multilocus phylogeny and molecular species delimitation of southeast Asian toads (family: Bufonidae). *BMC Evolutionary Biology* ,*3* , 1–12.

Chan, K. O., Grismer, L. L., & Brown, R. M. (2018). Comprehensive multi-locus phylogeny of Old World tree frogs (Anura: Rhacophoridae) reveals taxonomic uncertainties and potential cases of over- and under-estimation of species diversity. *Molecular Phylogenetics and Evolution* , *127* , 1010–1019.

Chen, J. M., Poyarkov, N. A., Suwannapoom, C., Lathrop, A., Wu, Y. H., Zhou, W. W., ... Che, J. (2018). Large-scale phylogenetic analyses provide insights into unrecognized diversity and historical biogeography of Asian leaf-litter frogs, genus Leptolalax (Anura: Megophryidae).*Molecular Phylogenetics and Evolution* , *124* , 162–171. doi: 10.1016/j.ympev.2018.02.020

Chen, J. M., Zhou, W. W., Poyarkov, N. A., Stuart, B. L., Brown, R. M., Lathrop, A., ... Che, J. (2017). A novel multilocus phylogenetic estimation reveals unrecognized diversity in Asian horned toads, genus*Megophrys* sensu lato (Anura: Megophryidae). *Molecular Phylogenetics and Evolution* , *109* , 28–43. doi: 10.1016/j.ympev.2016.11.019

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34* (17), i884–i890. doi: 10.1093/bioinformatics/bty560

Chenuil, A., Cahill, A. E., Delemontey, N., du Luc, E., & Fanton, H. (2019). Problems and questions posed by cryptic species. A framework to guide future studies. In E. Casetta, J. da Silva, & D. Vecchi (Eds.),*From Assessing to Conserving Biodiversity: Conceptual and Practical Challenges* (pp. 77–106). doi: 10.1007/978-3-030-10991-2_4

Chernomor, O., Von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices.*Systematic Biology* , *65* (6), 997–1008. doi: 10.1093/sysbio/syw037

Cobos, A., Grismer, L. L., Wood, P. L. J., Quah, E. S. H., Anuar, S., & Muin, M. A. (2016). Phylogenetic relationships of geckos of the*Hemiphyllodactylus harterti* group, a new species from Penang Island, Peninsular Malaysia, and a likely case of true cryptic speciation. *Zootaxa* , *4107* (May), 367–380.

Cooke, G. M., Chao, N. L., & Beheregaray, L. B. (2012). Five cryptic species in the Amazonian Catfish *Centromochlus existimatus*Identified based on biogeographic predictions and genetic data.*PLoS ONE* , *7*

(11). doi: 10.1371/journal.pone.0048800

Coyne, J. A., & Orr, H. A. (2004). Speciation. In *Sunderland, MA: Sinauer Associates* . Sunderland, MA: Sinauer Associates.

Davidson, R., Vachaspati, P., Mirarab, S., & Warnow, T. (2015). Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. *BMC Genomics* ,*16* (Suppl 10), S1. doi: 10.1186/1471-2164-16-S10-S1

de Queiroz, K. (2005). A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences* , *56* (1), 196–215. doi: 10.1073/pnas.0502030102

Devitt, T. J., Wright, A. M., Cannatella, D. C., & Hillis, D. M. (2019). Species delimitation in endangered groundwater salamanders: Implications for aquifer management and biodiversity conservation.*Proceedings of the National Academy of Sciences of the United States of America* , *116* (7), 2624–2633. doi: 10.1073/pnas.1815014116

Dincă, V., Lee, K. M., Vila, R., & Mutanen, M. (2019). The conundrum of species delimitation: A genomic perspective on a mitogenetically super-variable butterfly. *Proceedings of the Royal Society B: Biological Sciences* , *286* (1911). doi: 10.1098/rspb.2019.1311

Drillon, O., Dufresnes, G., Perrin, N., Crochet, P. A., & Dufresnes, C. (2019). Reaching the edge of the speciation continuum: Hybridization between three sympatric species of *Hyla* tree frogs.*Biological Journal of the Linnean Society* , *126* (4), 743–750. doi: 10.1093/biolinnean/bly198

Dufresnes, C., Pribille, M., Alard, B., Gonçalves, H., Amat, F., Crochet, P.-A., ... Martínez-Solano, I. (2020). Integrating hybrid zone analyses in species delimitation: lessons from two anuran radiations of the Western Mediterranean. *Heredity* , *124* , 423–438. doi: 10.1038/s41437-020-0294-z

Eckert, A. J., & Carstens, B. C. (2008). Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution* , *49* (3), 832–842. doi: 10.1016/j.ympev.2008.09.008

Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences* ,*113* (29), 8025–8032. doi: 10.1073/pnas.1601066113

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics* , *28* (7), 342–350. doi: 10.1016/j.tig.2012.03.009

Fenderson, L. E., Kovach, A. I., & Llamas, B. (2020). Spatiotemporal landscape genetics: Investigating ecology and evolution through space and time. *Molecular Ecology* , *29* , 218–246. doi: 10.1111/mec.15315

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* , (July 2017), 613–635. doi: 10.1111/mec.14486

Flouri, T., Jiao, X., Rannala, B., Yang, Z., & Yoder, A. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution* , 1–9. doi: 10.1093/molbev/msy147

Foote, A. D. (2018). Sympatric Speciation in the Genomic Era.*Trends in Ecology and Evolution* , *33* (2), 85–95. doi: 10.1016/j.tree.2017.11.003

Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M., & Gemmell, N. J. (2007). Underestimation of species richness in neotropical frogs revealed by mtDNA analyses. *PLoS ONE* , *2* (10), e1109. doi: 10.1371/journal.pone.0001109

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* ,*6* (8), 925–929. doi: 10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* , *196* (4), 973–983. doi: 10.1534/genetics.113.160572

Frost, D. R. (2019). Amphibian Species of the World: an Online Reference. Version 6.0 (accessed 10 June 2019).

Fujisawa, T., Aswad, A., & Barraclough, T. G. (2016). A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology* , *65* (5), 759–771. doi: 10.1093/sysbio/syw028

Garrick, R. C., Dickinson, T., Reppel, D. K., & Yi, R. N. (2019). Two divergent genetic lineages within the horned passalus beetle,*Odontotaenius disjunctus* (Coleoptera: Passalidae): An emerging model for insect behavior, physiology, and microbiome research.*Insects* , *10* (6). doi: 10.3390/insects10060159

Ginsberg, P. S., Humphreys, D. P., & Dyer, K. A. (2019). Ongoing hybridization obscures phylogenetic relationships in the*Drosophila subquinaria* species complex. *Journal of Evolutionary Biology* , *00* , 1–13. doi: 10.1111/jeb.13512

Giska, I., Sechi, P., & Babik, W. (2015). Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology* ,*15* (1), 1–13. doi: 10.1186/s12862-015-0488-9

Grismer, L. L., Wood, P. L., Anuar, S., Quah, E. S. H., Muin, M. A., Chan, K. O., ... Loredo, A. I. (2015). Repeated evolution of sympatric, palaeoendemic species in closely related, co-distributed lineages of *Hemiphyllodactylus* Bleeker, 1860 (Squamata: Gekkonidae) across a sky-island archipelago in Peninsular Malaysia.*Zoological Journal of the Linnean Society* , *174* (4), 859–876. doi: 10.1111/zoj.12254

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests.*Methods in Ecology and Evolution* , *4* (4), 336–344. doi: 10.1111/2041-210x.12018

Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution* , *70* (1), 7–17. doi: 10.1111/evo.12832

Hall, R. (2013). The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *Journal of Limnology* , *72* (S2), 1–17. doi: 10.4081/jlimnol.2013.s2.e1

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* ,*105* (S1), 795–809. doi: 10.1093/jhered/esu033

Hillis, D. M. (2019). Species Delimitation in Herpetology. *Journal of Herpetology* , *53* (1), 3–12. doi: 10.1670/18-123

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., & Vila, R. (2019). A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly Thymelicus sylvestris. *Molecular Ecology* , *28* , 3857–3868. doi: 10.1111/mec.15153

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Le, S. V. (2017). UFBoot2: improving the ultrafast bootstrap approximation.*Molecular Biology and Evolution* , *35* (2), 518–522. doi: 10.1093/molbev/msx281

Hutter, C. R., Cobb, K. A., Portik, D. M., Travers, S. L., Wood, P. L., & Brown, R. M. (2019). FrogCap : A modular sequence capture probe set for phylogenomics and population genetics for all frogs , assessed across multiple phylogenetic scales. *BioRxiv* , *825307* . doi: 10.1101/825307

Inger, R. F., Stuart, B. L., & Iskandar, D. T. (2009). Systematics of a widespread Southeast Asian frog, *Rana chalconota* (Amphibia: Anura: Ranidae). *Zoological Journal of the Linnean Society* ,*155* , 123–147. doi: 10.1111/j.1096-3642.2008.00440.x

Inger, R. F., & Voris, H. K. (2001). The Biogeographical Relations of the Frogs and Snakes of Sundaland. *Journal of Biogeography* ,*28* (7), 863–891.

Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology* ,*66* (5), 799–812. doi: 10.1093/sysbio/syw117

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* , *27* (21), 3070–3071. doi: 10.1093/bioinformatics/btr521

Jones, G. R. (2018). Divergence Estimation in the Presence of Incomplete Lineage Sorting and Migration. *Systematic Biology* , *68* (1), 19–31. doi: 10.1093/sysbio/syy041

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* , *14* (6), 587–589. doi: 10.1038/nmeth.4285

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* , *33* (11), 1630–1638. doi: 10.1093/bioinformatics/btx025

Kierepka, E. M., & Latch, E. K. (2015). Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources* , *15* (3), 512–525. doi: 10.1111/1755-0998.12332

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* , *17* (1), 44–53. doi: 10.1111/1755-0998.12549

Koh, L. P., Kettle, C. J., Sheil, D., Lee, T. M., Giam, X., Gibson, L., & Clements, G. R. (2013). Biodiversity State and Trends in Southeast Asia. In S. Levin (Ed.), *Encyclopedia of Biodiversity: Second Edition* (Vol. 1). doi: 10.1016/B978-0-12-384719-5.00357-9

Köhler, J., Jansen, M., Rodríguez, A., Kok, P. J. R., Toledo, L. F., Emmrich, M., ... Vences, M. (2017). The use of bioacoustics in anuran taxonomy: Theory, terminology, methods and recommendations for best practice. In *Zootaxa* (Vol. 4251). doi: 10.11646/zootaxa.4251.1.1

Krijthe, J. H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation. *URL: Https://Github.Com/Jkrijthe/Rtsne* .

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* (March), 46487. doi: 10.1038/srep46487

Ladner, J. T., & Palumbi, S. R. (2012). Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Molecular Ecology* , *21* (9), 2224–2238. doi: 10.1111/j.1365-294X.2012.05528.x

Leaché, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D., & Linkem, C. W. (2015). Phylogenomics of Phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution* ,*7* (3), 706–719. doi: 10.1093/gbe/evv026

Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study.*Systematic Biology* , *63* (1), 17–30. doi: 10.1093/sysbio/syt049

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The Spectre of Too Many Species. *Systematic Biology* , *68* (1), 168–181. doi: 10.1093/sysbio/syy051

Legendre, P., Fortin, M.-J., & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* , *6* (11), 1239–1247. doi: 10.1111/2041-210X.12425

Li, H. (2013). Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM. *ArXiv* , *1303.3997v* .

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25* (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, W., Cerise, J. E., Yang, Y., & Han, H. (2017). Application of t-SNE to human genetic data. *Journal of Bioinformatics and Computational Biology* , *15* (04), 1750017. doi: 10.1142/s0219720017500172

Lim, H. C., Gawin, D. F., Shakya, S. B., Harvey, M. G., Rahman, M. A., & Sheldon, F. H. (2017). Sundaland's east-west rain forest population structure: Variable manifestations in four polytypic bird species examined using RAD-Seq and plumage analyses. *Journal of Biogeography* . doi: 10.1111/jbi.13031

Linkem, C. W., Minin, V. N., & Leaché, A. D. (2016). Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Systematic Biology* , *65* (3), 465–477. doi: 10.1093/sysbio/syw001

Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion*(Hymenoptera: Diprionidae) sawflies. *Evolution* , *61* (6), 1417–1438. doi: 10.1111/j.1558-5646.2007.00114.x

Long, C., & Kubatko, L. (2018). The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology* ,*67* (5), 770–785. doi: 10.1093/sysbio/syy020

Lu, B., Bi, K., & Fu, J. (2014). A phylogeographic evaluation of the*Amolops mantzorum* species group: cryptic species and plateau uplift. *Molecular Phylogenetics and Evolution* , *73* (1), 40–52. doi: 10.1016/j.ympev.2014.01.008

Luo, A., Ling, C., Ho, S. Y. W., & Zhu, C.-D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology* , *67* (5), 830–846. doi: 10.1093/sysbio/syy011

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., ... Liu, J. (2017). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex.*Proceedings of the National Academy of Sciences of the United States of America* , *115* (2), 236–243. doi: 10.1073/pnas.1713288114

Maguilla, E., & Escudero, M. (2016). Cryptic species due to hybridization: A combined approach to describe a new species (carex: Cyperaceae). *PLoS ONE* , *11* (12), 1–24. doi: 10.1371/journal.pone.0166949

Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation.*Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* , *363* (1506), 2971–2986. doi: 10.1098/rstb.2008.0081

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genetics* ,*12* (2), 1–34. doi: 10.1371/journal.pgen.1005887

Mason, V. C., Helgen, K. M., & Murphy, W. J. (2019). Comparative phylogeography of forest-dependent mammals reveals paleo-forest corridors throughout sundaland. *Journal of Heredity* ,*110* (2), 158–172. doi: 10.1093/jhered/esy046

Matsui, M., Kuraishi, N., Eto, K., Hamidy, A., Nishikawa, K., Shimada, T., ... Hossman, M. Y. Bin. (2016). Unusually high genetic diversity in the Bornean Limnonectes kuhlii-like fanged frogs (Anura: Dicroglossidae). *Molecular Phylogenetics and Evolution* ,*102* , 305–319. doi: 10.1016/j.ympev.2016.06.009

McFadden, C. S., Haverkort-Yeh, R., Reynolds, A. M., Halàsz, A., Quattrini, A. M., Forsman, Z. H., ... Toonen, R. J. (2017). Species boundaries in the absence of morphological, ecological or geographical differentiation in the Red Sea octocoral genus Ovabunda (Alcyonacea: Xeniidae). *Molecular Phylogenetics and Evolution* , *112* , 174–184. doi: 10.1016/j.ympev.2017.04.025

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning* ,*20* , 1297–1303. doi: 10.1101/gr.107524.110.20

Mclean, B. S., Bell, K. C., Allen, J. M., Helgen, K. M., & Cook, J. A. (2019). Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of Ground Squirrels (Xerinae: Marmotini). *Systematic Biology* , *68* (2), 298–316. doi: 10.1093/sysbio/syy064

McLeod, D. S. (2010). Of Least Concern? Systematics of a cryptic species complex: *Limnonectes kuhlii* (Amphibia: Anura: Dicroglossidae).*Molecular Phylogenetics and Evolution* , *56* (3), 991–1000. doi: 10.1016/j.ympev.2010.04.004

Mendes, F. K., & Hahn, M. W. (2018). Why concatenation fails near the anomaly zone. *Systematic Biology* , *67* (1), 158–169. doi: 10.1093/sysbio/syx063

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., S. Swenson, M., & Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* , *30* (17), 541–548. doi: 10.1093/bioinformatics/btu462

Molloy, E. K., & Warnow, T. (2017). To include or not to include: the impact of gene filtering on species tree estimation methods.*Systematic Biology* , *67* (April), 285–303. doi: 10.1093/sysbio/syx077

Morales, A. E., & Carstens, B. C. (2018). Evidence that *Myotis lucifugus* "Subspecies" are five nonsister species, despite gene flow.*Systematic Biology* , *67* (5), 756–769. doi: 10.1093/sysbio/syy010

Muangmai, N., Von Ammon, U., & Zuccarello, G. C. (2016). Cryptic species in sympatry: Nonrandom small-scale distribution patterns in*Bostrychia intricata* (Ceramiales, Rhodophyta). *Phycologia* ,*55* (4), 424–430. doi: 10.2216/16-5.1

Nei, M. (1973). Analysis of gene diversity in subdivided populations.*Proceedings of the National Academy of Sciences of the United States of America* , *70* (12), 3321–3323.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* ,*32* (1), 268–274. doi: 10.1093/molbev/msu300

Nijman, V., & Nekaris, K. A. I. (2010). Checkerboard patterns, interspecific competition, and extinction: Lessons from distribution patterns of Tarsiers (*Tarsius*) and Slow Lorises (*Nycticebus*) in insular Southeast Asia. *International Journal of Primatology* , *31* (6), 1147–1160. doi: 10.1007/s10764-010-9458-7

Nishikawa, K., Matsui, M., Yong, H. Sen, Ahmad, N., Yambun, P., Belabut, D. M., ... Shimada, T. (2012). Molecular phylogeny and biogeography of caecilians from Southeast Asia (Amphibia, Gymnophiona, Ichthyophiidae), with special reference to high cryptic species diversity in Sundaland. *Molecular Phylogenetics and Evolution* ,*63* (3), 714–723. doi: 10.1016/j.ympev.2012.02.017

Nychka, D., Furrer, R., Paige, J., & Sain, S. (2017). "fields: Tools for spatial data." *R Package v 9.8-6* . doi: 10.5065/D6W957CT

Ogilvie, H. A., Heled, J., Xie, D., & Drummond, A. J. (2016). Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Systematic Biology* , *65* (3), 381–396. doi: 10.1093/sysbio/syv118

Oksanen, J., Blanchet, F. G., M., F., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2017). *Vegan: community ecology package. R package version 2.4-4* .

Pinho, C., & Hey, J. (2010). Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics* ,*41* (1), 215–230. doi: 10.1146/annurev-ecolsys-102209-144644

Poyarkov, N. A., Orlov, N. L., Moiseeva, A. V, Galoyan, E. A., Nguyen, T. T., & Gogoleva, S. S. (2015). Sorting out moss frogs: mtDNA data on taxonomic diversity and phylogenetic relationships of the Indochinese species of the genus *Theloderma* (Anura, Rhacophoridae).*Russian Journal of Herpetology* , *22* (December), 241–280.

Prunier, J. G., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., & Lena, J. P. (2013). Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: Towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology* , *22* (22), 5516–5530. doi: 10.1111/mec.12499

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation.*Molecular Ecology* , *21* (8), 1864–1877. doi: 10.1111/j.1365-294X.2011.05239.x

Pulido-Santacruz, P., Aleixo, A., & Weir, J. T. (2018). Morphologically cryptic amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* , *285* , 20172081. doi: 10.1098/rspb.2017.2081

Quattrini, A. M., Wu, T., Soong, K., Jeng, M., Benayahu, Y., & Mcfadden, C. S. (2019). A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. *BMC Evolutionary Biology* , *19* , 116 (2019).

Roch, S., Nute, M., & Warnow, T. (2019). Long-Branch Attraction in Species Tree Estimation: Inconsistency of Partitioned Likelihood and Topology-Based Summary Methods. *Systematic Biology* , *68* (2), 281–297. doi: 10.1093/sysbio/syy061

Rosindell, J., Cornell, S. J., Hubbell, S. P., & Etienne, R. S. (2010). Protracted speciation revitalizes the neutral theory of biodiversity.*Ecology Letters* , *13* (6), 716–727. doi: 10.1111/j.1461-0248.2010.01463.x

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology* , *14* (12), 1–22. doi: 10.1371/journal.pbio.2000234

Rowley, J. J. L., Tran, D. T. A., Frankham, G. J., Dekker, A. H., Le, D. T. T., Nguyen, T. Q., ... Hoang, H. D. (2015). Undiagnosed cryptic diversity in small, microendemic frogs (*Leptolalax*) from the Central Highlands of Vietnam. *PLoS ONE* , *10* (5), 1–21. doi: 10.1371/journal.pone.0128382

Ruane, S., Bryson, R. W., Pyron, R. A., & Burbrink, F. T. (2014). Coalescent species delimitation in Milksnakes (Genus*Lampropeltis*) and impacts on phylogenetic comparative analyses.*Systematic Biology* , *63* (2), 231–250. doi: 10.1093/sysbio/syt099

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters* , *8* (3), 336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Sayyari, E., Whitfield, J. B., & Mirarab, S. (2018). DiscoVista: Interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution* , *122* (February), 110–115. doi: 10.1016/j.ympev.2018.01.019

Schield, D. R., Card, D. C., Adams, R. H., Corbin, A., Jezkova, T., Hales, N., ... Castoe, T. A. (2018). Cryptic genetic diversity, population structure, and gene flow in the Mojave rattlesnake (*Crotalus scutulatus*). *Molecular Phylogenetics and Evolution* , *127* (July 2017), 669–681. doi: 10.1016/j.ympev.2018.06.013

Schield, D. R., Card, D. C., Adams, R. H., Jezkova, T., Reyes-Velasco, J., Proctor, F. N., ... Castoe, T. a. (2015). Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (Crotalus atrox). *Molecular Phylogenetics and Evolution* , *83* , 213–223. doi: 10.1016/j.ympev.2014.12.006

Scriven, J. J., Whitehorn, P. R., Goulson, D., & Tinsley, M. C. (2016). Niche partitioning in a sympatric cryptic species complex. *Ecology and Evolution* , *6* , 1328–1339. doi: 10.1002/ece3.1965

Slager, D. L., Epperly, K. L., Ha, R. R., Rohwer, S., Wood, C., Van Hemert, C., & Klicka, J. (2020). Cryptic and extensive hybridization between ancient lineages of American crows. *Molecular Ecology* ,*00* , 1–14. doi: 10.1111/mec.15377

Smith, M. L., & Carstens, B. C. (2019). Process-based species delimitation leads to identification of more biologically relevant species. *Evolution* , *66* , 37–39. doi: 10.1111/evo.13878

Sobel, J. M., & Streisfeld, M. A. (2015). Strong premating reproductive isolation drives incipient speciation in *Mimulus aurantiacus* .*Evolution* , *69* (2), 447–461. doi: 10.1111/evo.12589

Sodhi, N. S., Koh, L. P., Brook, B. W., & Ng, P. K. L. (2004). Southeast Asian biodiversity: an impending disaster. *Trends in Ecology and Evolution* , *19* (12), 654–660. doi: 10.1016/j.tree.2004.09.006

Solís-Lemus, C., Yang, M., & Ané, C. (2016). Inconsistency of species tree methods under gene dlow. *Systematic Biology* , *65* (5), 843–851. doi: 10.1093/sysbio/syw030

Stanton, D. W. G., Frandsen, P., Waples, R. K., Heller, R., Russo, I. R. M., Orozco-terWengel, P. A., ... Bruford, M. W. (2019). More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservation Genetics* ,*20* (1), 101–113. doi: 10.1007/s10592-019-01149-5

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology and Evolution* , *33* (3), 153–163. doi: 10.1016/j.tree.2017.11.007

Stuart, B. L., Inger, R. F., & Voris, H. K. (2006). High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology Letters* , *2* (3), 470–474. doi: 10.1098/rsbl.2006.0505

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* , *114* (7), 1607–1612. doi: 10.1073/PNAS.1607921114

Supple, M. A., Papa, R., Hines, H. M., McMillan, W. O., & Counterman, B. A. (2015). Divergence with gene flow across a speciation continuum of*Heliconius* butterflies. *BMC Evolutionary Biology* ,*15* (1), 1–12. doi: 10.1186/s12862-015-0486-y

Surveswaran, S., Gowda, V., & Sun, M. (2018). Using an integrated approach to identify cryptic species, divergence patterns and hybrid species in Asian ladies' tresses orchids (Spiranthes, Orchidaceae).*Molecular Phylogenetics and Evolution* , *124* (February), 106–121. doi: 10.1016/j.ympev.2018.02.025

Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey.*Methods in Ecology and Evolution* , *4* (12), 1101–1110. doi: 10.1111/2041-210X.12107

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature.*Nature Ecology and Evolution* , *3* (2), 170–177. doi: 10.1038/s41559-018-0777-y

Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology* , (December 2018), 1–18. doi: 10.1111/mec.15189

Vachaspati, P., & Warnow, T. (2015). ASTRID: Accurate species TRees from internode distances. *BMC Genomics* , *16* (Suppl 10), 1–13. doi: 10.1186/1471-2164-16-S10-S3

Vachaspati, P., & Warnow, T. (2018). SVDquest: Improving SVDquartets species tree estimation using exact optimization within a constrained search space. *Molecular Phylogenetics and Evolution* ,*124* (March), 122–136. doi: 10.1016/j.ympev.2018.03.006

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. In *Current Protocols in Bioinformatics* . doi: 10.1002/0471250953.bi1110s43

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE.*Journal of Machine Learning Research* , *9* , 2579–2605.

Vaughan, T. G. (2017). IcyTree: Rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* , *33* (15), 2392–2394. doi: 10.1093/bioinformatics/btx155

Vieites, D. R., Wollenberg, K. C., Andreone, F., Kohler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences* , *106* (20), 8267–8272. doi: 10.1073/pnas.0810821106

Wagner, F., Härtl, S., Vogt, R., & Oberprieler, C. (2017). "Fix Me Another Marguerite!": Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae). *Molecular Ecology* , *26* (16), 4260–4283. doi: 10.1111/mec.14180

Wagner, F., Ott, T., Schall, M., Lautenschlager, U., Vogt, R., & Oberprieler, C. (2020). Taming the Red Bastards: Hybridisation and species delimitation in the *Rhodanthemum arundanum* - group (Compositae, Anthemideae). *Molecular Phylogenetics and Evolution* ,*144* (2020), 106702. doi: 10.1016/j.ympev.2019.106702

Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Systematic Biology* , *67* (4), 735–740. doi: 10.1093/sysbio/syy015

Wilcove, D. S., Giam, X., Edwards, D. P., Fisher, B., & Koh, L. P. (2013). Navjot's nightmare revisited: Logging, agriculture, and biodiversity in Southeast Asia. *Trends in Ecology and Evolution* ,*28* (9), 531–540. doi: 10.1016/j.tree.2013.04.005

Winter, D. J. (2012). MMOD: An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* , *12* (6), 1158–1160. doi: 10.1111/j.1755-0998.2012.03174.x

Xu, B., & Yang, Z. (2016). Challenges in species tree estimation under the multispecies coalescent model. *Genetics* , *204* (4), 1353–1368. doi: 10.1534/genetics.116.190173

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* , *107* (20), 9264–9269. doi: 10.1073/pnas.0913022107

Yumul, G. P., Dimalanta, C. B., Marquez, E. J., & Queaño, K. L. (2009). Onland signatures of the Palawan microcontinental block and Philippine mobile belt collision and crustal growth process: A review.*Journal of Asian Earth Sciences* , *34* (5), 610–623. doi: 10.1016/j.jseaes.2008.10.002

Yumul, G. P., Dimalanta, C. B., Tamayo, R. D., Maury, R. C., Bellon, H., Polve, M., ... Cotten, J. (2004). Geology of the Zamboangan Peninsula, Mindanao, Philippines: an enigmatic South China continental fragment? In J. Malpas, C. J. N. Fletcher, J. R. Ali, & J. C. Aitchison (Eds.), *Aspects of the Tectonic Evolution of China* (pp. 289–312). London.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* , *19* (Suppl 6), 15–30. doi: 10.1186/s12859-018-2129-y

Zheng, H., Fan, L., Milne, R. I., Zhang, L., Wang, Y., & Mao, K. (2017). Species delimitation and lineage separation history of a species complex of aspens in China. *Frontiers in Plant Science* ,*8* (March), 1–17. doi: 10.3389/fpls.2017.00375

Data accessibility

Relevant data generated from this project are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.zw3r2284d [Not released]. For peer-review, please use the following link:

https://datadryad.org/stash/share/5W7nS3nyHbIseL22eUmxopYVuuiuhMLFvrdyNVKHGwk

Bioinformatic scripts can be obtained from <https://github.com/chutter/FrogCap-Sequence-Capture/>

Author contributions

RMB conceived of this project and, together with KOC and PLW, designed and implemented the study; CRH developed FrogCap resources, data processing, and SNP analysis pipelines; PLW oversaw sample preparation. KOC performed analyses, and composed the manuscript, with input from all authors, who approved this paper in its final form.

Figures and Tables

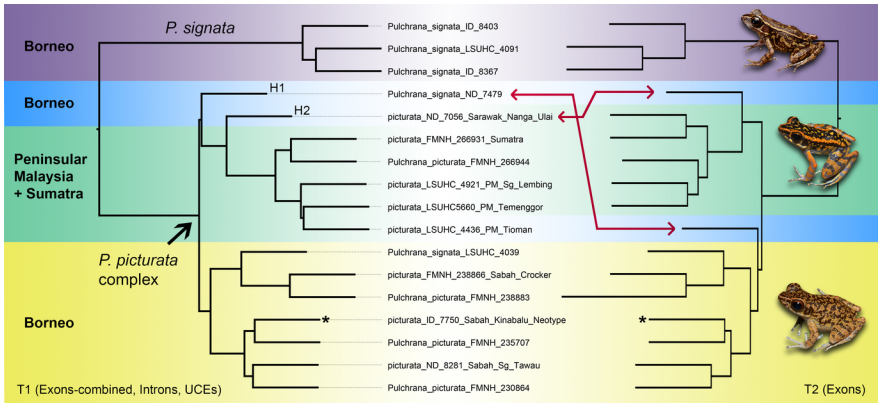


Fig. 1. Two species tree summary topologies (T1, T2), inferred by ASTRAL-III, based on the unfiltered Exons-combined (2,186 markers), Introns (11,935), UCEs (625; left), and Exons datasets (11,978). All nodes were supported by 1.0 local posterior probabilities and placements of discordant samples (putative hybrids: H1, H2) are indicated by red arrows. IQ-TREE and ASTRID analyses produced the same topologies for the corresponding datasets. *=topotype specimen for *Pulchranapicturata* . See supplementary material for trees with full taxon representation (including outgroups). Inset photos by A. Haas (top and bottom) and KOC (middle).

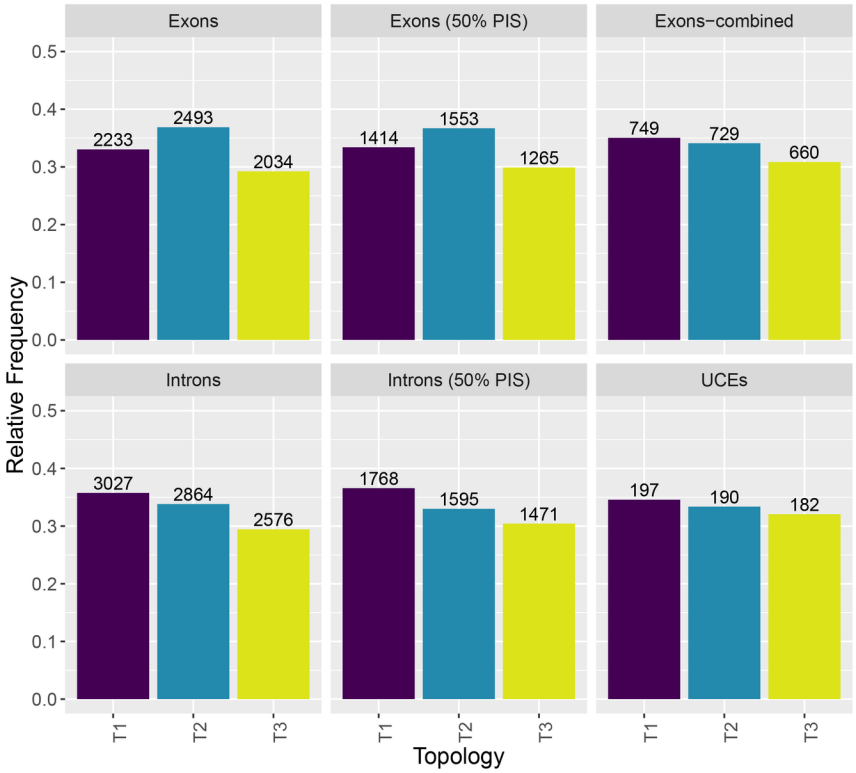


Fig. 2. Relative frequencies of alternate gene tree topologies for each dataset. Numbers on top of bars represent the actual number of gene trees supporting that particular topology. The T1 and T2 topologies are presented in Fig. 1, while the T3 gene tree topology was not recovered in any of our phylogenetic species tree analyses.

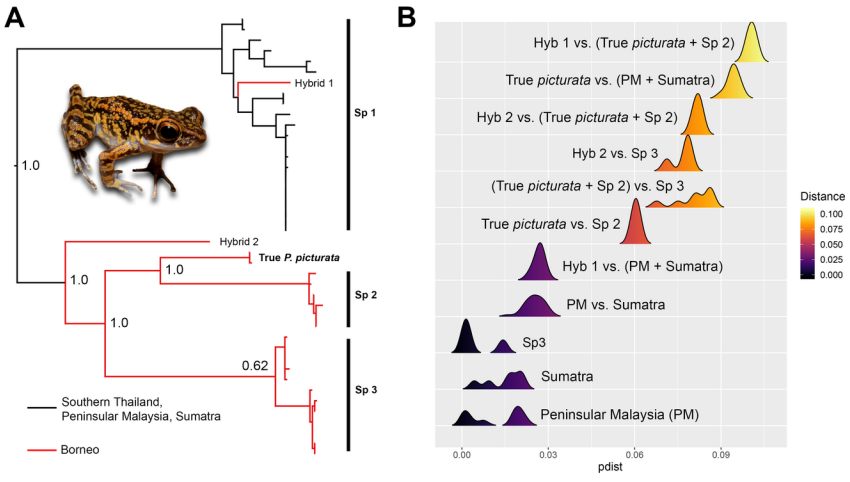


Fig. 3. A. Putative species delimitation using mPTP analysis, based on 16S rRNA data. Support values at nodes indicate the fraction of sampled delimitations in which a node was part of the speciation process. The analysis strongly supported the discovery-step delimitation of putative candidate lineages labelled here as

Sp1, Sp2, Sp3, True*P. picturata* , and Hybrid 2 ("Hyb 2") as distinct species. The ABGD analysis produced the same preliminary candidate species discovery results. **B.** Distribution of uncorrected p -distances among pairs of taxa/populations/samples, based on the 16S rRNA gene. Distributions labelled "Sp3," "Sumatra," and "Peninsular Malaysia (PM)" presumably represent intraspecific genetic variation. Inset photo by KOC.

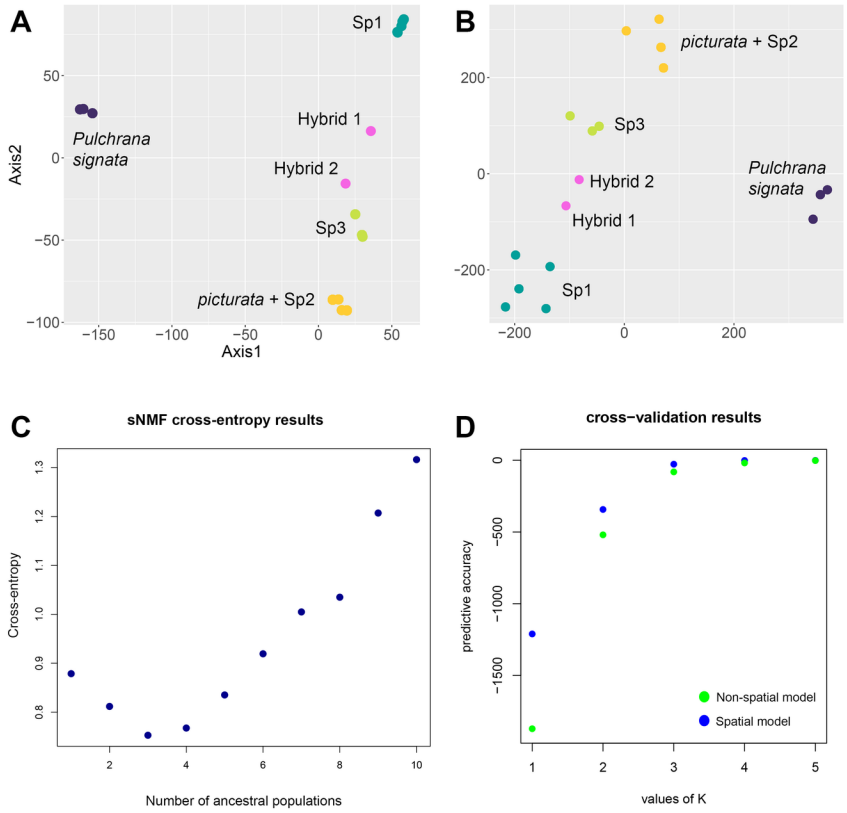


Fig. 4. A. Results of Principal Components Analysis and**B.** t-distributed Stochastic Neighbour Embedding (t-SNE) Analysis, demonstrating population clustering after dimension-reduction of SNP data. **C.** Cross-entropy results of K 1–10 (lower cross-entropy scores correspond to the highest predictive accuracy) from the sparse non-negative matrix factorization (sNMF) analysis.**D.** cross-validation results from *conStruct* analysis, using non-spatial and spatial models (Ks of highest log-likelihood scores correspond to highest predictive accuracy).

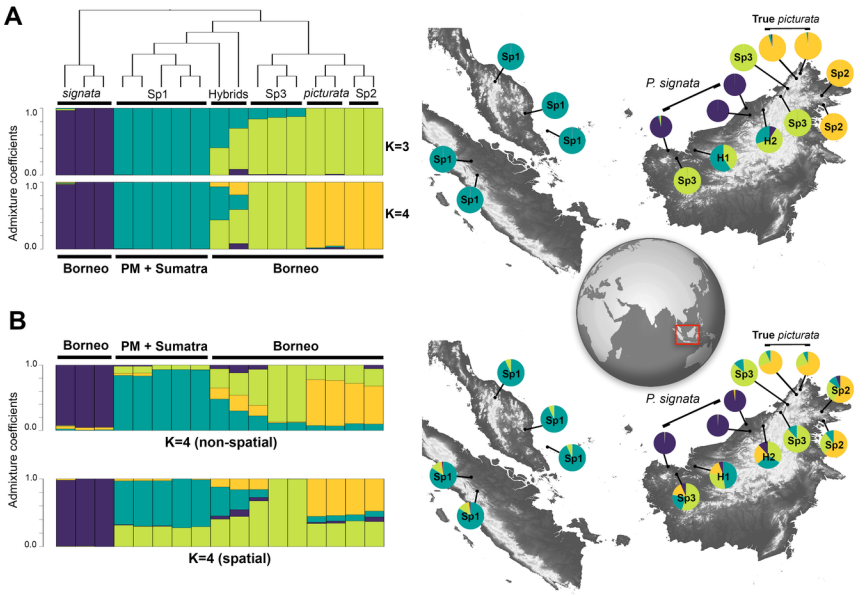


Fig. 5. A. Barplots of admixture coefficients from the sparse non-negative matrix factorization (sNMF) analysis at K=3 and K=4, juxtaposed with a cladogram depicting our T1 topology (refer to Fig. 1). Population labels correspond to putative species inferred from species discovery stage analysis of 16S rRNA. Maps (right panels) depict locations of each sample and pie charts of admixture ratios for K=4.**B.** Results of spatial and non-spatial *conStruct* analysis and corresponding distribution map showing admixture ratios for K=4. H1 and H2 represent the putative hybrid samples. The location of the study region is outlined in the red box on the global inset map.

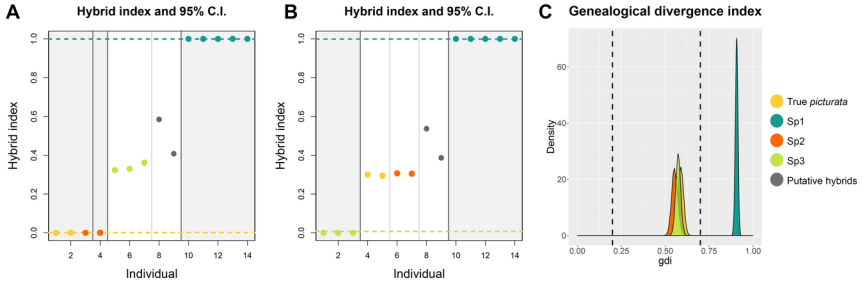


Fig. 6. Bayesian hybrid-index plots, with Sp1, True *P. picturata* + Sp2 (**A**) and Sp1, Sp3 (**B**) as parental references. Dotted lines demarcate 95% confidence intervals.**C.** Density plots of *gdi* values. We interpret species validation to be accomplished in cases of $gdi > 0.7$, whereas $0.2 < gdi < 0.7$ indicate uncertain species status.

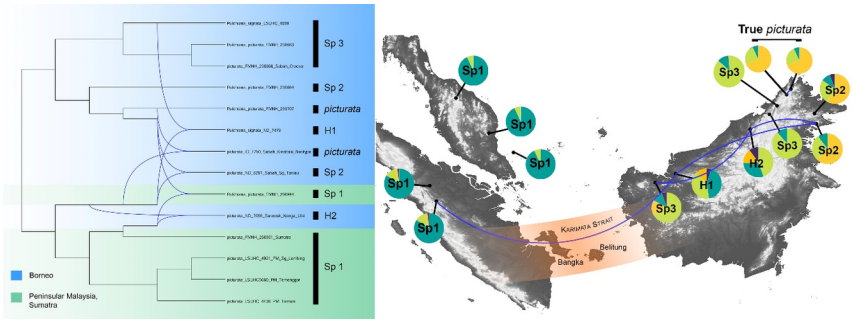


Fig. 7. Left. Results of the PhyloNet phylogenetic network analysis depicted using IcyTree (Vaughan, 2017). **Right.** Results from the *conStruct* analysis at K=4. Blue lines connecting populations on the map correspond to blue lines depicting reticulations on the phylogenetic network. Orange shading represents the putative habitat corridor that facilitated gene flow between Sumatra and Borneo.

Table 1. Summary statistics of datasets used for phylogenomics and species delimitation analyses. EC=exons combined; PIS50=top 50% loci with highest parsimony-informative-sites. Branch lengths are in coalescent units.

Dataset	No. loci	Mean length	Total sites	Total var. sites	Total PIS
Intron-unfiltered	11,935	452	5,395,834	2,676,967	950,103
Exon-unfiltered	11,978	212	2,543,793	578,939	243,378
EC-unfiltered	2,186	617	1,349,664	286,927	121,681
UCE-unfiltered	625	713	445,346	103,021	37,368
Intron-PIS50	5,968	513	3,063,129	1,652,988	652,822
Exon-PIS50	5,989	378	1,673,499	428,542	190,302
EC-PIS50	1,093	870	950,907	212,555	92,220

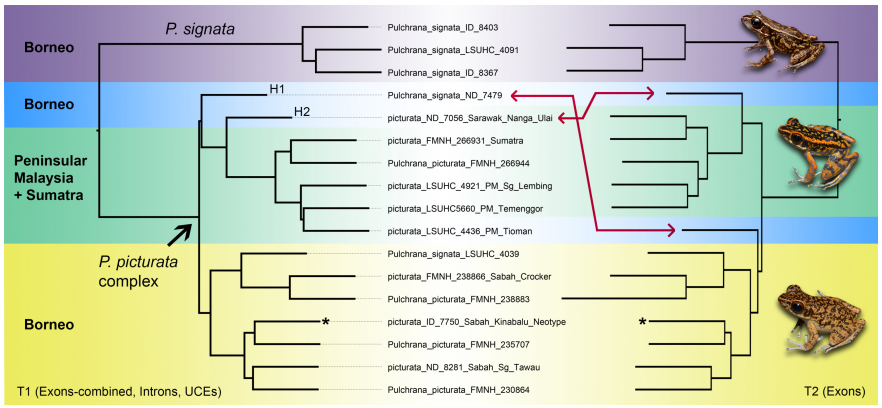
Table 2. Results of HyDe analysis at population and individual levels. P-values <0.05 indicate significant levels of hybridization. Population names follow putative species assignments. Pic=true *P. picturata* , Hyb= (H1 and H2).

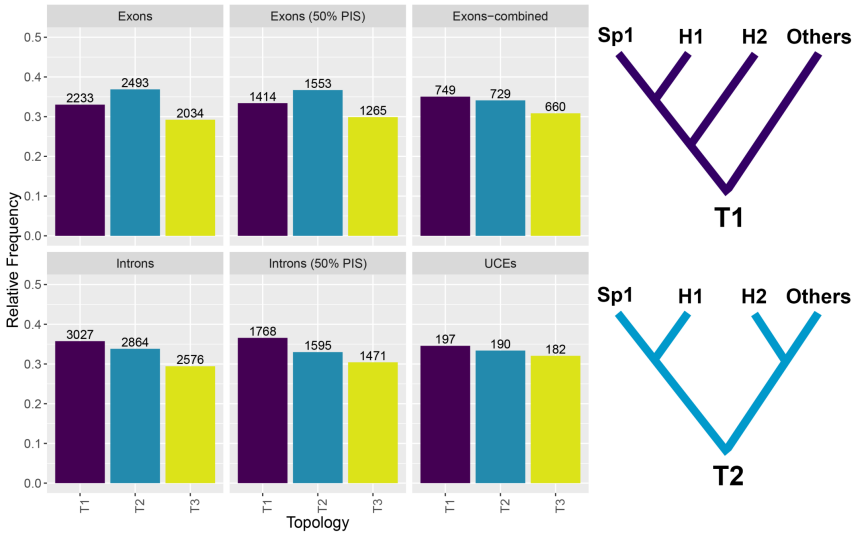
P1	Hybrid	P2	Zscore	Pvalue	Gamma
Population level	**Population level**				
Pic	Hyb	Sp1	6.9922	0.0000	0.2023
Hyb	Sp3	Pic	7.8896	0.0000	0.3450
Sp1	Hyb	Sp2	8.3316	0.0000	0.7772
Sp1	Hyb	Sp3	6.6488	0.0000	0.7480
Hyb	Sp3	Sp2	7.1718	0.0000	0.2626
Pic	Sp3	Sp1	9.1459	0.0000	0.7102
Pic	Sp2	Sp3	4.0127	0.0000	0.9072
Sp1	Sp3	Sp2	8.8885	0.0000	0.2360
Individual level	**Individual level**				
Pic	ND_7056_Sarawak_Nanga_Ulai	Sp1	6.8507	0.0000	0.1177
Pic	ND_7479	Sp1	6.4388	0.0000	0.4719
Hyb	FMNH_238866_Sabah_Crocker	Pic	8.4826	0.0000	0.3501
Hyb	LSUHC_4039	Pic	6.6791	0.0000	0.3159
Hyb	FMNH_238883	Pic	8.4596	0.0000	0.3658
Sp1	ND_7056_Sarawak_Nanga_Ulai	Sp2	7.6828	0.0000	0.8783

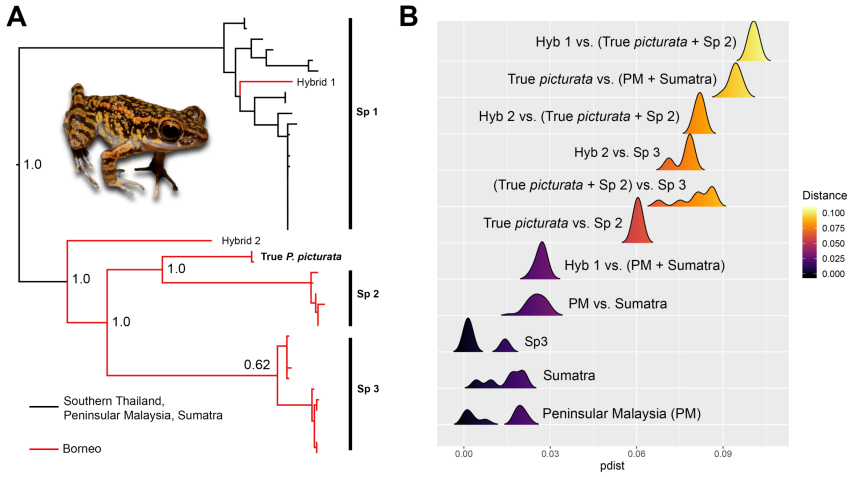
P1	Hybrid	P2	Zscore	Pvalue	Gamma
Sp1	ND_7479	Sp2	6.8121	0.0000	0.4450
Sp1	ND_7056_Sarawak_Nanga_Ulai	Sp3	9.0282	0.0000	0.8301
Sp1	ND_7479	Sp3	-9999.0000	1.0000	-0.3707
Hyb	FMNH_238866_Sabah_Crocker	Sp2	8.4319	0.0000	0.2869
Hyb	LSUHC_4039	Sp2	4.9586	0.0000	0.2127
Hyb	FMNH_238883	Sp2	8.0984	0.0000	0.2790
Pic	FMNH_238866_Sabah_Crocker	Sp1	8.7067	0.0000	0.7282
Pic	LSUHC_4039	Sp1	10.0259	0.0000	0.6779
Pic	FMNH_238883	Sp1	8.6311	0.0000	0.7252
Pic	FMNH_230864	Sp3	3.2790	0.0005	0.9175
Pic	ND_8281_Sabah_Sg_Tawau	Sp3	4.7139	0.0000	0.8988
Sp1	FMNH_238866_Sabah_Crocker	Sp2	9.1245	0.0000	0.2333
Sp1	LSUHC_4039	Sp2	9.4740	0.0000	0.2589
Sp1	FMNH_238883	Sp2	8.0379	0.0000	0.2157

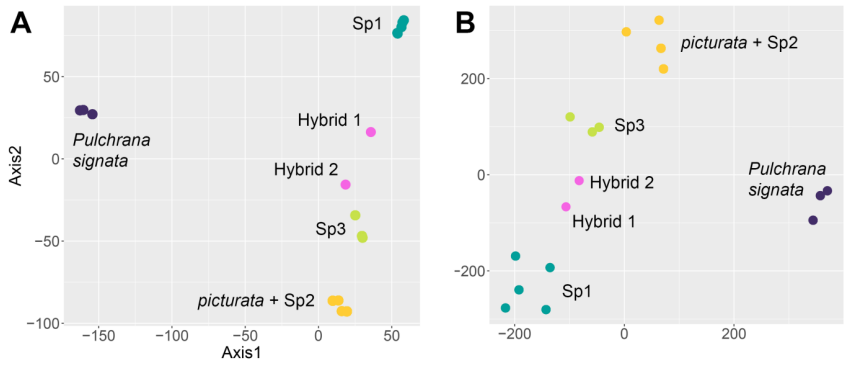
Supplementary Material

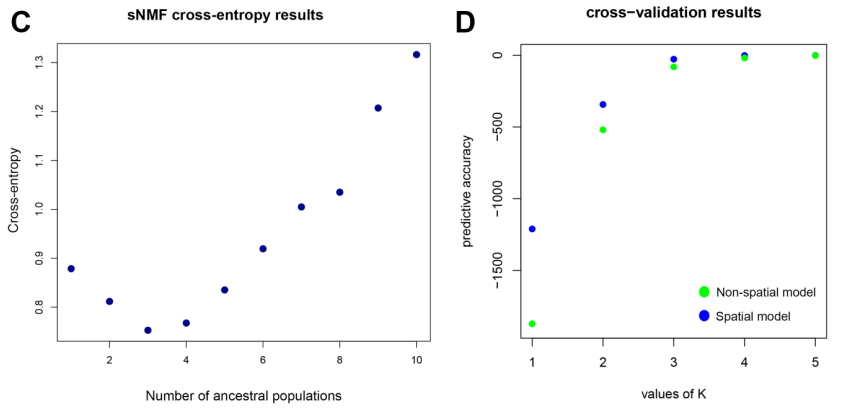
Table S1. List of samples used and sequenced in this study (see also Brown & Siler, 2014, for additional details).

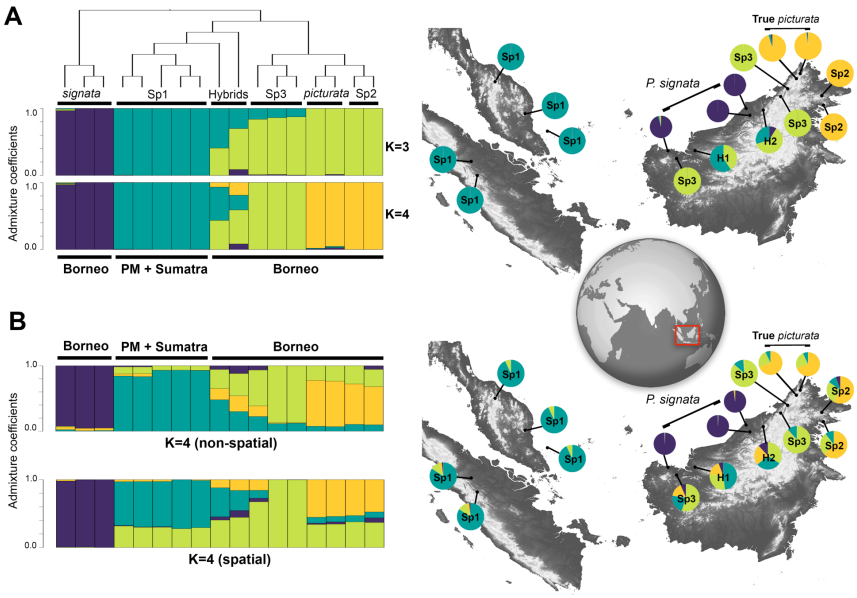


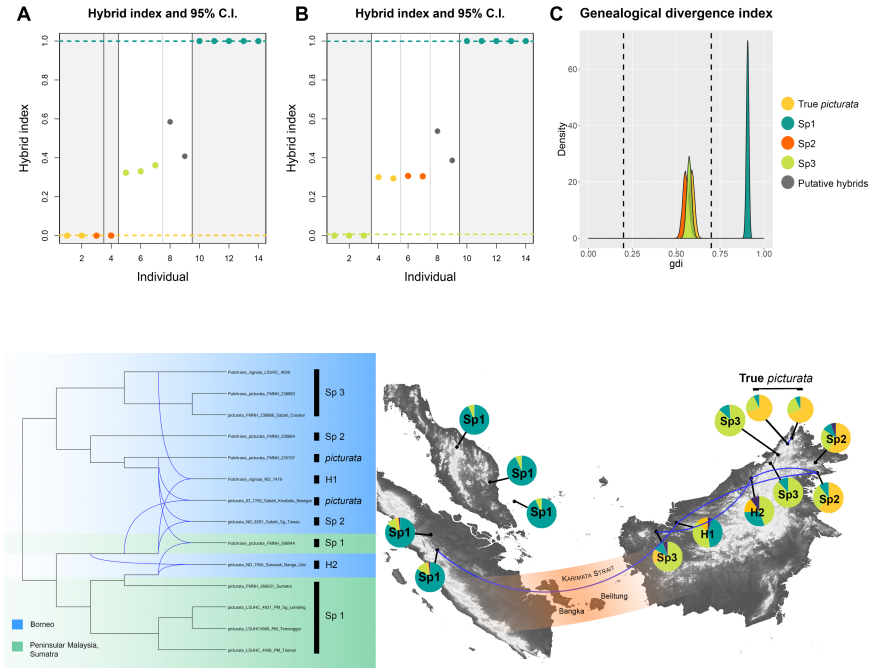












Hosted file

Table 1.xlsx available at https://authorea.com/users/302616/articles/432726-gene-flow-creates-a-mirage-of-cryptic-species-in-a-southeast-asian-spotted-stream-frog-complex

Hosted file

Table 2.xlsx available at https://authorea.com/users/302616/articles/432726-gene-flow-creates-a-mirage-of-cryptic-species-in-a-southeast-asian-spotted-stream-frog-complex

Hosted file

Table S1.xlsx available at https://authorea.com/users/302616/articles/432726-gene-flow-creates-a-mirage-of-cryptic-species-in-a-southeast-asian-spotted-stream-frog-complex