Inferring the history of hybridization: A case study in Iochrominae (Solanaceae)

Daniel J. Gates[1,2], Diana Pilson[1], and Stacey D. Smith[3]
[1]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588
[2]Checkerspot, Inc., Alameda, CA 94501
[3]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, 80309

ABSTRACT

The role of hybridization in evolution and speciation is one of the most active areas of evolutionary biology research. One problem that has continually vexed research in hybridization is accurate and robust detection of hybridization across large samples of species. Advances in phylogenetic data collection and theory have offered both powerful new datasets and methods with which to test for hybridization across large phylogenies. In this study, we compare patterns of hybridization in experimental phylogenetic datasets in order to compare existing methods for detecting hybridization. These analyses include both natural hybrids and an artificial hybrid as a phylogenetic 'positive control'. Although our analyses recover strong indicators of hybridization in datasets with hybrid taxa, the results frequently disagree across methods or fail to identify known relationships (e.g., between the artificial hybrid and its parental taxa). These conflicting outcomes suggest that continued theoretical and empirical studies are needed to identify the factors that influence statistical power to detect past hybridization events in phylogenomic analyses.

INTRODUCTION

The impact of hybridization on evolution and speciation has been well established in evolutionary biology (Dobzhansky, 1940; Mayr, 1942b, a; Anderson, 1953; Stebbins, 1959; Grant and Grant, 1971; Rieseberg, 1995; Barton, 2001; Mallet, 2007; Abbott et al., 2013). Despite the historical notion that hybridization between distant relatives is rare and typically detrimental (Mayr, 1942a, b), numerous studies have indicated that hybridization occurs frequently between both closely and distantly related species (Mallet, 2005; Mallet, Besansky, and Hahn, 2016). Moreover, the introgression of genes via hybridization can be selectively advantageous (Grant et al., 2004; Consortium, 2012; Kirkpatrick and Barrett, 2015). The recognition that hybridization is a widespread and important mechanism of adaptation and speciation has placed a premium on the development of methods to detect signatures of reticulate evolution and identify the lineages and genomic regions involved (Nakhleh, 2013; Hejase and Liu, 2016)

Inferring a history of hybridization, whether among close or distant relatives, has typically relied upon molecular markers. The earliest approaches were developed and applied in a population genetic framework, focusing on pairs or 'swarms' of putatively hybridizing species (Moran, Bell, and Matheson, 1980; Avise and Saunders, 1984; Arnold, Hamrick, and Bennett, 1990). Although population genetic methods are powerful at shallow timescales (Durand et al., 2011), they do not easily scale to cases involving dozens of taxa. The need to test for hybridization across entire clades has driven the development of phylogenetic methods that build on population genetic principles (e.g., Kubatko, 2009; Yu, Degnan, and Nakhleh, 2012). These methods are often computationally intensive and require large amounts of data to distinguish the signal of introgression from the background noise generated by the stochasticity of the coalescent process (Yu et al., 2011; Eaton and Ree, 2013). Nevertheless, advances in sequencing and bioinformatics have made it increasingly feasible to build genome-wide datasets outside of model taxa (e.g., Faircloth et al., 2012; Lemmon, Emme, and Lemmon, 2012; Rubin, Ree, and Moreau, 2012; Jarvis et al., 2014). Thus, future improvements in estimating evolutionary histories involving hybridization are likely to come from theoretical and computational advances (e.g., Kubatko and Chifman, 2015; Solis-Lemus and Ane, 2016).

Current approaches for studying hybridization in a phylogenetic framework can be divided into two broad categories, global and local. Global methods aim to detect the signal of hybridization across the phylogeny without explicitly determining the specific taxa involved (Table 1). These approaches often combine coalescent simulations with distance metrics or other indices to exclude the hypothesis that gene tree discordance can be explained by incomplete lineage sorting alone. In comparison to the global tests, local methods attempt to resolve specific hybridization relationships (i.e., identify the lineages involved in hybridization events) (Table 1). The more recent methods in this category estimate the proportion of genes inherited by a hybrid lineage

from each parent, in addition to inferring the topology of the species network (Kubatko and Chifman, 2015; Solis-Lemus and Ane, 2016; Blischak et al., 2018). After finding a global signal across the tree, local approaches can be used to pinpoint hybridization events. Still, the inference of the exact hybrid relationships remains challenging, and in some scenarios, it may be impossible to recover the full evolutionary history, regardless of the amount of available data (Pardi and Scornavacca, 2015).

Here we explore the ability of local and global methods to detect historical hybridization in a natural and experimental context, using the Iochrominae as a model clade. This group of roughly 40 Andean species in the tomato family has been traditionally divided into six genera, although many of the species and genera remain crossable and interfertile, even after millions of years of divergence (Smith and Baum, 2007). Previous phylogenetic studies with nuclear markers have demonstrated substantial gene tree conflict and suggested at least three reticulation events (Smith and Baum, 2006; Smith et al., 2008). In this study, we apply phylogenomic approaches to quantify the signal of hybridization and test specific hybrid relationships based on morphology and biogeography. Given the interfertility of many Iochrominae, we also created an artificial hybrid to serve as a 'positive control'. Our approach is inspired by the pioneering work of Lucinda McDade (1990, 1992), who used natural hybrids to quantify the impact of hybridization on phylogenetic inference. Following her lead, we similarly manipulate the number of hybrids in our dataset to test the impact on the inference of species trees and hybridization events. We predict that increasing the number of hybrids in the dataset will increase discordance among gene trees and reduce the confidence in estimated species relationships. We also expect that artificial hybrids will be easier to detect as they have not had the opportunity to backcross or accumulate new mutations as in natural hybrids. In addition to improving our understanding of the role of hybridization in the Iochrominae radiation, these results will shed light on the strengths and limitations of current approaches and provide direction for future development of computational methods.

METHODS

Study System

The Iochrominae has been well sampled in previous systematic studies (e.g., Smith and Baum, 2006; Muchhala, Johnsen, and Smith, 2014; Deanna et al., 2018; Gates, Pilson, and Smith, 2018), providing a strong framework for focused analyses of hybridization. This clade is sister to the large Physalinae (tomatillos and their relatives) (Deanna et al., 2019) and is largely restricted to the Andes of South America (Smith and Baum, 2006). Species of Iochrominae often grow in sympatry (Smith et al., 2008) and, in some cases, form hybrid zones (Smith, Kolberg, and Baum, 2008). Phylogenetic studies have sampled several putative hybrids and recovered strongly conflicting relationships in nuclear gene trees (Smith and Baum, 2006). The strongest support for

hybridization from genetic, morphological and biogeographic data corresponds to two taxa (*I. ayabacense* S. Leiva and *I. stenanthum* S. Leiva, Quip. & N. W. Sawyer) (Fig. 1). Native to northern Peru, *I. ayabacense* grows in a zone of sympatry with several species (*I. cyaneum* (Lindl.) M. L. Green*, I. confertiflorum* (Miers) Hunz., *I. arborescens* (Schltdl.) J. M. H. Shaw*,* and *I. lehmannii* Bitter (syn*. I. squamosum* S. Leiva &Quip.) (Smith, Kolberg, and Baum, 2008). Alleles from this species fall either in a clade with *I. lehmannii* or with *I. cyaneum* (Smith and Baum, 2006). Such a biphyletic pattern (with alleles from the same locus in distant clades on the phylogeny) is a signature consistent with recent hybridization (Barkman and Simpson, 2002). The chromosome number of *I. ayabacense* is 2n=24, like other *Iochroma* species, suggesting it is a homoploid hybrid (Smith, Kolberg, and Baum, 2008). Also, its morphology is intermediate, with sturdy funnel-shaped flowers like *I. lehmannii* but flushes of purple floral pigmentation and long corolla tube like *I. cyaneum* (Smith, Kolberg, and Baum, 2008).

The other putative hybrid taxon, *I. stenanthum*, has similarly intermediate coloration, being creamy at the base and mauve towards the mouth of the flower. It occurs in another zone of sympatry in northern Peru and has been proposed to be a hybrid of the lineage comprising the purple-flowered *I. cornifolium* (Kunth.) Miers and *I. cyaneum* and a white-flowered lineage from the clade containing *I. confertiflorum* and *I. arborescens* (Smith and Baum, 2006; Smith and Leiva, 2011). *I. stenanthum* presents an intermediate morphology, being purple at the mouth of the corolla and fading to cream towards the base. It has long tubular flowers as in *I. cornifolium* and *I. cyaneum* but more triangular corolla lobes as in *I. confertiflorum* and *I. arborescens*. In individual gene tree analyses, alleles from *I. stenanthum* show affinity towards both the white- and purple-flowered lineages (Gates et al., unpublished data). In combined analyses, the phylogenetic position of *I. stenanthum* is unstable, but it often appears as the sister lineage to the clade containing the putative parental taxa (Smith and Baum, 2006; De-Silva et al., 2017). The latter phylogenetic pattern was proposed by McDade (1990) as a common outcome for hybrid taxa.

In addition to these two putative natural hybrids, we included one artificial hybrid, generated in the greenhouse. This accession is a cross between *I. arborescens* (Bohs 2428 (UT), grown from seed collected from Las Cruces, Costa Rica) and *I. cyaneum* (Smith 265 (WIS), grown from seed collected from plants cultivated by W. G. D'Arcy at the Missouri Botanical Garden; likely provenance, Ecuador). These parental species are thought to be 2 to 4 million years diverged (De-Silva et al., 2017), and some F_1 seeds produce vigorous offspring. The hybrid flower has an intermediate phenotype (funnel-shaped flowers, white with purple markings). Its fertility has not been thoroughly examined, although it seems to be capable of backcrossing to the parental taxa (S. D. Smith, unpublished data).

Experimental design

In order to assess our ability to detect hybridization with various phylogenomic methods, we created a series of datasets with different combinations of the two putative natural hybrids (*I. stenanthum* and *I. ayabacense*), their putative parental lineages (*I. arborescens*, *I. lehmannii*, *I. cyaneum*), the artificial F_1 hybrid (*I. arborescens* x *I. cyaneum*), plus an outgroup (*Physalis peruviana* L.). For two of the three putative parents (*I. arborescens* and *I. lehmannii*), we have two sampled individuals (Table 2). All of the experimental datasets contain the same number of tips (six) and all have at least one individual of each putative parent plus the outgroup (Fig. 2). Dataset 1 has no putative hybrids, while dataset 2 has the F_1 hybrid substituting for one of the two *I. arborescens* individuals. Dataset 3 has the F_1 and *I. ayabacense* replacing one of each of the *I. arborescens* and *I. lehmannii* individuals. As *I. ayabacense* tends to fall with *I. lehmannii* in species trees (Gates, unpublished) estimated with STAR (Liu et al., 2009), the topology for dataset 3 is expected to mirror that for dataset 2 (Fig. 2), albeit with less support due to the expected hybrid ancestry. The fourth dataset has the two putative natural hybrids (*I. stenanthum* and *I. ayabacense*) in the place of one of each of the *I. arborescens* and *I. lehmannii* individuals. The putative parents for *I. stenanthum* are the same as for the F_1 (*I. arborescens* x *I. cyaneum*) although STAR species tree analyses tend to infer a closer relationship with *I. arborescens* (Gates, unpublished). Thus, we expect a similar species tree topology to the previous datasets, but a weaker signal of hybridization since *I. stenanthum* is likely a late generation hybrid compared to the F_1 (Fig. 2).

Target capture and assembly

We conducted targeted sequence capture to build a molecular dataset for phylogenomic analyses following Gates et al. (2018). Briefly, we used genomic resources for tomato (Sato et al., 2012) and *Iochroma cyaneum* (Gates et al., 2016; Strickler et al., unpublished data) to design probes for 242 single copy genes, most of which were pulled from the conserved orthologous set (Wu et al., 2006). These loci totaled approximately 1 MB and were used as the template for custom bait synthesis of 80 nucleotide targets at 2X tiling from MycroArray (MycroArray Inc; Ann Arbor, MI). Genomic DNA was extracted from silica-dried leaf tissue using a 2XCTAB protocol (Doyle and Doyle, 1987) and sheared DNA to approximately 500 bases with sonication. For all libraries except for the F_1 hybrid, we used the TruSeq library preparation kits for sequence adapter ligation and Mycroarray capture protocol V1. These libraries were sequenced on a single lane of an Illumina HiSeq2500 on 100bp pair-end rapid mode. For the F_1 hybrid, we used the NEBNext Ultra II kit (New England Biolabs, Ipswich, MA) for library preparation and Mycroarray capture protocol V4. This library was sequenced on an Illumina HiSeq 4000, also in 100bp pair-end mode. We filtered raw reads by using Trimmomatic (Bolger, Lohse, and Usadel, 2014) and assembled filtered raw reads with the iterative read-mapping program YASRA (Ratan, 2009). We created a master alignment for all individuals by orienting assemblies to homologous bases of the reference using custom Blat (Kent, 2002) and samtools (Li et al., 2009) scripts. We then used this master alignment to subset individuals into each of the four experimental datasets. We

conducted gene tree searches on each gene in each of the four datasets using RAxML 8 (Stamatakis, 2014) with 100 bootstrap replicates to assess support. These alignments and gene tree estimates were used for all downstream analyses (available at https://github.com/danjgates/HybrData).

Detecting patterns of hybridization from gene tree distributions

To quantify the level of gene tree conflict associated with varying numbers of hybrids in our four datasets, we first applied the internode and tree certainty (IC/TC) measures of Salichos, Stamatakis, and Rokas (2014) as implemented in RAxML 8 (Stamatakis, 2014). These information theory-based metrics can be applied to trees generated from any method and optimality criterion (e.g., parsimony, likelihood) and are computationally tractable with large, genome-scale data matrices. We computed IC for each of the three internodes in our experimental datasets (Fig. 2). These values represent the degree of conflict between relationships inferred with STAR (Liu et al., 2009) and the most common conflicting partition. When IC approaches 1, there is little conflict at the node, and when it is near zero, there is strong conflict, with similar numbers of gene trees showing the bipartition in the species tree and the next most common bipartition. For each of our four datasets, we also computed the tree credibility (TC) score, which is the sum of the IC scores over all ingroup bipartitions (Salichos, Stamatakis, and Rokas, 2014). We predicted that the addition of tips with known or inferred hybrid ancestry would increase gene tree conflict and thus decrease the certainty in relationships across the phylogeny.

We further characterized the distribution of the gene trees in each dataset by using "Tree Incongruence Checking in R" or "TICR" (Stenz et al., 2015; Solis-Lemus, this volume). This program uses pseudo-loglikelihoods (PLL) calculated from rooted quartet distributions to test the fit of gene tree distributions to different hypotheses of panmixia or species tree structure. While rejection of a population tree may be consistent with a history of hybridization and may be used as evidence of hybridization (Stenz et al., 2015), the program does not directly test a model of hybridization against the other models of bifurcation or panmixia. We used TICR to compute the PLL and associated $\chi 2$ statistic for both the panmixia (star-like ingroup tree) and fully bifurcating species tree in each of the four datasets. In comparing datasets with hybridization to those without (e.g. one F_1 dataset vs. the no putative hybrid dataset), we expect that there should be less support for a bifurcation model as indicated by a greater $\chi 2$ value and more support for a panmixia model as indicated by a lowered $\chi 2$ value in the dataset with greater amounts of hybridization. The TICR analyses rely on concordance factors (CFs) for each quartet (Ane et al., 2007), estimated with BUCKY (Larget et al., 2010). We extracted these CFs for each clade as an additional window into patterns of gene tree discord associated with adding hybrid taxa.

Testing for hybridization in empirical datasets

In addition to examining uncertainty and discordance across the tree, we also implemented three specific tests for hybridization. While the two approaches above (IC/TC and TICR) may offer support for hybridization, these methods do not explicitly test whether a hybridization model is a more appropriate model than a model without hybridization. In addition to reticulation, conflict among gene trees and instability in species tree inference can arise due to other factors, such as incomplete lineage sorting (Degnan and Rosenberg, 2006), recombination (Schierup and Hein, 2000), model misspecification (Buckley, 2002), biased taxon sampling (Hillis, 1998), missing data (Wiens, 2003), sequencing errors (Kuhner and McGill, 2014), and alignment uncertainty (Hossain et al., 2015). Thus, without explicit modeling of processes of hybridization, metrics like IC and CF only offer weak support for hybridization.

In order to directly test for hybridization, we first applied HyDe (Hybrid Detection; Blischak et al., 2018). This program localizes hybridization events (i.e., identifies hybrid taxa and their putative parental taxa) based on patterns in sequence alignments. Detecting these events relies on invariants, i.e., phylogenetically informative site patterns in quartet subtrees, such as those used in the ABBA/BABA patterns used to compute D-statistics (Patterson et al., 2012). Specifically, HyDe uses the ratio of the differences between specific site patterns, as this approach was found to provide the highest statistical power (Kubatko and Chifman, 2015). We also implemented a similar alignment-based method that computes D-statistics across all quartet subtrees to test for introgression (Eaton and Ree, 2013). While HyDe reports statistics for every combination of two parental tips and one hybrid tip, the D-statistic approach reports all possible triplet combinations that are consistent with the species tree. For the purpose of this analysis, we only report the maximum statistic for each method in each dataset as this should be the strongest proposed hybrid relationship.

Along with these alignment based approaches, we used an additional method, SNaQ (Species Networks applying Quartets; Solis-Lemus and Ane, 2016; Solis-Lemus, this volume), to fit reticulation events based upon gene tree distributions. Like the TICR method described above, SNaQ calculates a quartet-based PLL for models with and without hybridization. Unlike TICR, however, SNaQ does directly test support for hybridization. For each dataset, we report the maximum PLL score for the strictly bifurcating species tree as well as the maximum score from the best network produced by a full search with up to six optimal reticulation events. For SNaQ as well as HyDe, we expect that inferred reticulation events will follow the predictions based on morphology and biogeography and that more recent hybridization events (the F_1 hybrid and the biphyletic *I. ayabacense*) will be more consistently and accurately recovered than the older events (those that formed *I. stenanthum*).

RESULTS

Addition of hybrid taxa increases discordance and decreases tree-like signal

Comparisons between the no hybrid and one F_1 dataset offer strong evidence that hybridization has an impact on a number of tree metrics. Values for internode certainty (IC) ranged from 0.19 to 0.60 in the no hybrid tree and dropped to 0 to 0.32 with the substitution of the F1 hybrid for one of the putative parental individuals (Fig. 3a,b). Overall tree certainty (TC) fell from 1.04 to 0.32 (Fig. 3a,b). We also saw decreases in concordance factors, from 0.47 to 0.71 without the F_1 to 0.26 to 0.53 with the F_1. Given that the F_1 is an artificial hybrid between *I. arborescens* and *I. cyaneum* (Fig. 1), we would expect the branches connecting these taxa to be most strongly affected by adding the hybrid and, indeed, these were the only internodes with decreases in IC and CF values. In fact, the unaffected branch, uniting the two *I. lehmannii* individuals, saw a slight increase in IC and CF (Fig. 3a,b). These increases in gene tree conflict with the addition of the F_1 hybrid shifted the $\chi 2$ values from TICR test in line with decreasing tree-like signal. Although both datasets reject the two extremes (panmixia and a fully bifurcating history), we observed weakening of the rejection of complete panmixia ($\chi 2_P$ dropping from 26 to 11) and a strengthening the rejection of a fully bifurcating tree ($\chi 2_B$ rising from 58 to 75) with the inclusion of the F_1 (Fig. 3a,b).

The increase from one hybrid tip to a second hybrid tip brought additional decreases in certainty and concordance. In both datasets 3 and 4, all of the internal branches are expected to be influenced by the hybridization events and indeed, all internodes dropped to IC values less than 0.03 and CFs of less than 0.36 (Fig. 3c,d). Overall tree certainty in both cases was near zero, and panmixia was strongly favored over a fully-bifurcating history (Fig. 3c,d). The combination of the F_1 and *I. ayabacense* in dataset 3 had a slightly stronger effect on these values, which may be due to the recency of their formation. By comparison, *I. stenanthum* appears to be a more ancient hybrid (with less intermediate morphology, no strongly biphyletic alleles, and a geographic distribution more distant from its putative parents) (Smith and Baum, 2006). Thus, some of the conflicting signal originally carried in this lineage may have eroded due to sorting of parental alleles, the fixation of new mutations, and/or backcrossing to one of the parents.

Tests of hybridization support different relationships than expected

All three of the tests used to detect hybridization events (HyDe, SNaQ, and D-statistics) provided strong support for hybrid ancestry but also pointed to some unexpected relationships. First, both HyDe and the D-statistics supported some degree of reticulation in dataset 1 (the no-hybrid datasets) and two taxa, *I. arborescens* and *I. lehmannii*, were consistently implicated (Table 3a,b). With correction for multiple tests, these results are marginally significant. SNaQ recovered the same tips involved, but in this case, the addition of the reticulation did not improve the psuedo-loglikelihood of the tree (Table 3c). Although the implication of some reticulation involving *I. arborescens* and *I. lehmannii* was surprising, it is notable that *I. lehmannii* does

show instability in traditional phylogenetic analyses and variously appears as more closely related to *I. arborescens* (Deanna et al., 2018) or *I. cyaneum* (Deanna et al., 2019), always with weak support (bootstrap < 75%). Species tree analyses favor the former relationship (Gates, Pilson, and Smith, 2018), but these have not previously included tests for hybridization. Still, the strength of support for hybridization was weak compared to the other three datasets (Table 3). Given these mixed results, we consider that *I. lehmannii* does not have a large degree of hybrid ancestry, and that the patterns may reflect a small rate of gene flow between *I. lehmannii* and *I. arborescens* in southern Ecuador and northern Peru, where the two grow in proximity (S.D. Smith, pers. obs.).

The datasets with one or more hybrid taxa consistently supported the presence of hybridization, and in a few cases, recovered the predicted relationship. We expected the best case scenario to be the one F_1 dataset, where the parents of the artificial hybrid (*I. arborescens* and *I. cyaneum*) are known and included in the tree. HyDe identified the F_1 as being involved but placed it as a parent, with *I. lehmannii* as the other parent and *I. arborescens* as the hybrid (Table 3a). The D-statistics pointed to the same three taxa as involved in hybridization (Table 3b). By contrast, SNaQ recovered the correct relationship among the parental taxa and the F_1 and with strong support for dataset 3 (Table 3c).

The results for datasets with two of the six taxa being hybrids were similarly mixed. For dataset 3 (F_1 + *I. ayabacense*), HyDe correctly grouped *I. ayabacense* with *I. arborescens* and *I. cyaneum*, but inferred *I. arborescens* instead of *I. ayabacense* as the hybrid among the three. Neither D-statistics nor SNaQ estimated a predicted grouping of parents with hybrids for dataset 3. For dataset 4 with two natural putative hybrids (*I. ayabacense* + *I. stenanthum*), only the D-statistics recovered a predicted grouping of tips involved in hybridization (*I. ayabacense* and its putative parents, *I. lehmannii* and *I. cyaneum*) (Table 3).

DISCUSSION

Effects of hybridization on patterns of gene-tree discordance

Our manipulations of the taxon sampling show that adding known or putative hybrids increases discordance between gene trees and decreases the signal of tree-like relationships. All of the datasets in which putative or known hybrid taxa were swapped with putative parental taxa showed marked decreases in concordance factors and internode certainty and a strong preference for panmixia over tree-like relationships (Fig. 3). The comparison of the one F_1 dataset compared with the no putative hybrid dataset offers the strongest evidence on this front because the F_1 is an artificial hybrid generated in the greenhouse. In this dataset, the concordance factors for the nodes connecting the two putative parental species were reduced by roughly half and the certainty for those relationships dropped to zero. Doubling the number of hybrid taxa (datasets 3 and 4) further increased gene tree conflict, and with hybrids distributed across the tree, complete

panmixia could not be rejected by TICR (Fig. 3). Altogether, we see patterns consistent with prevalent, detectable hybridization in all trees that we expect to exhibit a signal of reticulate ancestry.

The methods designed to identify hybrid tips and their parental taxa also showed strong evidence of hybridization in datasets with more putative or known hybrids. For example, SNaQ and HyDe did not find significant support for hybridization in dataset 1, with no known hybrids (Table 3). In the remaining three datasets, all of the methods inferred significant reticulation, with the strength of that inference increasing with the number of hybrid taxa. These results indicate that the sharp discordance generated by the addition of hybrid tips, as indicated by decreasing concordance factors and tree certainty (Fig. 3), is generally correctly interpreted by HyDe, D-statistics and SNaQ as evidence of hybridization during the evolutionary history of *Iochroma*.

Challenges in determining the exact hybrid relationships

Our results also indicate that exact hybrid relationships may be difficult to assign. Looking across the three methods and the three datasets with augmented hybrid taxa, only two of these nine combinations estimated a predicted set of three taxa (two hypothesized parental tips along with the hypothesized hybrid tip). These differed across methods, with D-statistics inferring a predicted trio of tips in dataset 4, and SNaQ in dataset 2 (shaded in Table 3). Although both HyDe and SNaQ directly estimate which tip in a trio is the hybrid tip, only SNaQ correctly inferred a predicted relationship, identifying the F_1 as the product of a cross between *I. arborescens* and *I. cyaneum*. This relationship between the artificial hybrid and its parents was intended to serve as our 'positive control' for detecting hybrid ancestry, and it is worrisome that only one of the three methods recovered the correct topology.

The wide range of inferred and seemingly spurious hybrid relationships across methods and datasets may relate to both our experimental design and our study system. We intentionally built a small dataset with two putative natural hybrids and their putative parents with the goal of maximizing the potential for methods to converge upon the correct sets of relationships. While we are not entirely certain of these relationships, the combination of morphological, biogeographic and genetic data provides the strongest possible *a priori* predictions, beyond generating these hybrids ourselves. Still, a larger dataset, with more taxa to break up the branches between the putative parents, might have provided more power to discriminate among possible placements for reticulation events. The fact that hybridization appears to be common in *Iochroma* may also contribute to the challenge of estimating reticulation events. Controlled greenhouse crosses suggest that all *Iochroma* are able to interbreed to some degree, and they can also cross with closely related genera (Smith and Baum, 2007). Apparent hybrid zones are well documented in herbaria as well as in the literature (Smith and Baum, 2006; Smith and Leiva, 2011). Accordingly, all of the *Iochroma* lineages may contain a signal of hybrid ancestry,

consistent with rejection of a fully-bifurcating tree by TICR for the no hybrid dataset (Fig. 3a). This episodic gene flow, combined with incomplete lineage sorting (ILS) along short branches of the backbone of the species tree (Deanna et al., 2019), may result in relatively little information for making robust inferences about which tips have a significantly reticulate history.

It is important to note that the difficulty of making inferences about hybridization is already well documented in empirical and theoretical studies. Simulations attempting to reconstruct networks where there are few tips (<5) and multiple hybridization events have often failed to find the accurate network (Yu et al., 2011; Solis-Lemus and Ane, 2016). Presumably, this is because with few taxa and multiple reticulations, the set of gene trees may be equally well explained by different network topologies (Pardi and Scornavacca, 2015). The fact that multiple hybridizations can confound inference could explain why SNaQ identified the correct relationships in the one-F1 dataset but failed in datasets 3 and 4 with two hybrids (Table 3). Moreover, SNaQ assumes that the edges can only be involved in one reticulation event (Solis-Lemus and Ane, 2016), and in datasets 3 and 4, *I. cyaneum* likely contributed to two events. Although making inferences about multiple reticulations from small trees is challenging, a similar empirical study was able to strongly support up to four reticulations in a dataset with six species (Wen et al., 2016). Given these conflicting notions in the literature about when hybridization can be confidently detected, we suggest that future theoretical studies explore a broad array of scenarios to determine which factors have the strongest influence on the power to correctly infer reticulate relationships. We expect that population size and divergence time would have major impacts because these parameters directly govern the amount of signal and noise (ILS) in multilocus datasets (Degnan and Rosenberg, 2006). Violations of mutation rate assumptions may also have a large effect upon methods like HyDe and D-statistics because increases or decreases in mutation rates will make some lineages more similar to distant relatives without hybridization ever taking place. As empirical studies like ours continue to apply these relatively new methods, we expect that additional biological factors that potentially influence statistical power are likely to emerge.

Hybridization in Iochrominae

One goal of this study was to assess support for the hybrid origins of *I. ayabacense* and *I. stenanthum* with genome-wide markers. While previous studies had assembled evidence for these hypotheses from a handful of nuclear markers along with morphological and biogeographic information (Smith and Baum, 2006), this is the first study to directly test for hybridization with phylogenomic methods. The analyses of gene tree discordance provide some support for hybrid ancestry in that the presence of these taxa in the datasets decreased tree-like signal and certainty in relationships among tips. Also, the analyses with the two natural hybrids (dataset 4) showed similarly low concordance to those with one natural hybrid and one artificial hybrid (dataset 3), suggesting that a natural hybrid (in this case, *I. stenanthum*) disrupted tree-like relationships to a

similar degree as the F_1 (Fig. 3). Nevertheless, our results are far from conclusive in terms of which lineages gave rise to these two hybrid species. The local methods each returned different estimates of relationships among putative parental taxa and putative natural hybrids (Table 3). Moreover, none of these estimated relationships followed those predicted based on previous studies (Fig. 1) despite the fact that the ingroup taxa were comprised of putative parental taxa. Accurate inference of the hybrid ancestry for *I. stenanthum* was expected to be challenging because its phylogenetic position is more suggestive of an ancient hybrid, e.g. between stem lineages that gave rise to *I. arborescens* and *I. cyaneum* (Smith and Baum, 2006). By contrast, we expected the relationships between *I. ayabacense* and its putative parents would be easily detected given that the strong biphyletic patterns (e.g. one *LFY* allele sister to *I. lehmannii* and the other sister to *I. cyaneum*) in gene trees (Smith and Baum, 2006). The inability of multiple approaches to recover the predicted topology for *I. ayabacense,* not to mention the F_1 positive control, suggests that renewed attempts at testing these hybrid origins will require more statistical power (more loci, more individuals, and/or more tips). Ultimately, with a larger dataset, we expect that the majority of local methods should converge on the same set of relationships, especially since they all build on quartet-based patterns.

CONCLUSIONS

The growth of phylogenomic datasets along with the well-established prevalence of hybridization in nature has driven the development of an array of statistical tools for detecting and localizing reticulation in phylogenies. The application of these methods has led to the inference of hybridization at multiple phylogenetic scales, from sister species (e.g., Turissini and Matute 2017) to more distantly related taxa (e.g., Buckley et al. 2006; Eaton et al. 2013). Our exploratory analyses, which included both known and hypothesized hybrid tips, suggest that, while the broad signature of reticulate evolution is relatively easily detected (Reid et al. 2012; Blanco-Pastor et al. 2012; Villiers et al. 2013), the accurate localization of hybridization on the phylogeny is significantly more difficult. While these inferences may be challenging in *Iochroma* because of periodic gene flow across the entire tree, we do not consider this clade to be an outlier. Hybridization has accompanied the diversification of many taxa (e.g., butterflies, birds, fish) and is certainly well-documented in plant clades. Groups in which past and on-going hybridization is suspected are likely to be the primary targets for empirical applications of methods, such as SNaQ and HyDe. Thus, we encourage further empirical or experimental studies manipulating the degree of hybrid ancestry in addition to theoretical work to explore the power to make inferences across a spectrum of biologically realistic scenarios (including recently formed hybrid species in a backdrop of ILS and periodic gene flow).With more robust inferences, we may begin to arrive at a broader understanding of how reticulation affects patterns of genomic variation and how these impacts relate to the timing, duration and scale of hybridization events.

ACKNOWLEDGEMENTS

The authors thank L. L. Knowles and L. S. Kubatko for organizing the Society for Systematic Biologists workshop in which this work was first presented. We are grateful to Segundo Leiva for his collaboration in field collecting trips and for sharing his expertise in *Iochroma* taxonomy. We also thank J. Mower, B. Carstens and S. Russo for feedback on an earlier version of this chapter. This research was supported by NSF DEB-1355518 to S.D.S. and a Graduate Assistance in Areas of National Need fellowship to D.J.G.

REFERENCES

ABBOTT, R., D. ALBACH, S. ANSELL, J. W. ARNTZEN, S. J. E. BAIRD, N. BIERNE, J. W. BOUGHMAN, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229-246.

ANDERSON, E. 1953. Introgressive Hybridization. *Biological Reviews of the Cambridge Philosophical Society* 28: 280-307.

ANE, C., B. LARGET, D. A. BAUM, S. D. SMITH, AND A. ROKAS. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412-426.

ARNOLD, M. L., J. L. HAMRICK, AND B. D. BENNETT. 1990. Allozyme variation in Louisiana Irises - a test for introgression and hybrid speciation. *Heredity* 65: 297-306.

AVISE, J. C., AND N. C. SAUNDERS. 1984. Hybridization and Introgression among Species of Sunfish (Lepomis) - Analysis by Mitochondrial-DNA and Allozyme Markers. *Genetics* 108: 237-255.

BARKMAN, T. J., AND B. B. SIMPSON. 2002. Hybrid origin and parentage of Dendrochilum acuiferum (Orchidaceae) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. *Systematic Botany* 27: 209-220.

BARTON, N. H. 2001. Speciation. *Trends in Ecology & Evolution* 16: 325.

BLANCO-PASTOR, J. L., P. VARGAS, AND B. E. PFEIL. 2012. Coalescent Simulations Reveal Hybridization and Incomplete Lineage Sorting in Mediterranean Linaria. *Plos One* 7.

BLISCHAK, P. D., J. CHIFMAN, A. D. WOLFE, AND L. S. KUBATKO. 2018. HyDe: A python package for genome-scale hybridization detection. *Systematic Biology* 67: 821-829.

BOLGER, A. M., M. LOHSE, AND B. USADEL. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.

BUCKLEY, T. R. 2002. Model misspecification and probabilistic tests of topology: Evidence from empirical data sets. *Systematic Biology* 51: 509-523.

BUCKLEY, T. R., M. CORDEIRO, D. C. MARSHALL, AND C. SIMON. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology* 55: 411-425.

CONSORTIUM, H. G. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94-98.

DE-SILVA, D. L., L. L. MOTA, N. CHAZOT, R. MALLARINO, K. L. SILVA-BRANDAO, L. M. G. PINEREZ, A. V. L. FREITAS, et al. 2017. North Andean origin and diversification of the largest ithomiine butterfly genus. *Scientific Reports* 7: 45966.

508 DE VILLIERS, M. J., M. D. PIRIE, M. HUGHES, M. MOLLER, T. J. EDWARDS, AND D. U. BELLSTEDT. 2013. An
509 approach to identify putative hybrids in the 'coalescent stochasticity zone', as
510 exemplified in the African plant genus *Streptocarpus* (Gesneriaceae). *New Phytologist*
511 198: 284-300.
512 DEANNA, R., S. D. SMITH, T. SARKINEN, AND F. CHIARINI. 2018. Patterns of chromosomal evolution in
513 the florally diverse Andean clade Iochrominae (Solanaceae). *Perspectives in Plant*
514 *Ecology Evolution and Systematics* 35: 31-43.
515 DEANNA, R., M. D. LARTER, G. E. BARBOZA, AND S. D. SMITH. 2019. Repeated evolution of a
516 morphological novelty: a phylogenetic analysis of the inflated fruiting calyx in the
517 Physalideae tribe (Solanaceae). *American Journal of Botany*: in press.
518 DEGNAN, J. H., AND N. A. ROSENBERG. 2006. Discordance of species trees with their most likely
519 gene trees. *Plos Genetics* 2: 762-768.
520 DOBZHANSKY, T. 1940. Speciation as a stage in evolutionary divergence. *American Naturalist* 74:
521 312-321.
522 DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure from small quantities of fresh
523 leaf tissues. *Phytochemical Bulletin* 19: 11-15.
524 DURAND, E. Y., N. PATTERSON, D. REICH, AND M. SLATKIN. 2011. Testing for ancient admixture
525 between closely related populations. *Molecular Biology and Evolution* 28: 2239-2252.
526 EATON, D. A. R., AND R. H. REE. 2013. Inferring phylogeny and introgression using RADseq data:
527 An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62:
528 689-706.
529 FAIRCLOTH, B. C., J. E. MCCORMACK, N. G. CRAWFORD, M. G. HARVEY, R. T. BRUMFIELD, AND T. C. GLENN.
530 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple
531 evolutionary timescales. *Systematic Biology* 61: 717-726.
532 GATES, D. J., D. PILSON, AND S. D. SMITH. 2018. Filtering of target sequence capture individuals
533 facilitates species tree construction in the plant subtribe Iochrominae (Solanaceae).
534 *Molecular Phylogenetics and Evolution* 123: 26-34.
535 GATES, D. J., S. R. STRICKLER, L. A. MUELLER, B. J. S. C. OLSON, AND S. D. SMITH. 2016. Diversification of
536 R2R3-MYB transcription factors in the tomato family Solanaceae. *Journal of Molecular*
537 *Evolution* 83: 26-37.
538 GRANT, P. R., B. R. GRANT, J. A. MARKERT, L. F. KELLER, AND K. PETREN. 2004. Convergent evolution of
539 Darwin's finches caused by introgressive hybridization and selection. *Evolution* 58: 1588-
540 1599.
541 GRANT, V., AND K. A. GRANT. 1971. Natural Hybridization between Cholla Cactus Species Opuntia-
542 Spinosior and Opuntia-Versicolor. *Proceedings of the National Academy of Sciences of*
543 *the United States of America* 68: 1993-&.
544 HEJASE, H. A., AND K. J. LIU. 2016. Mapping the genomic architecture of adaptive traits with
545 interspecific introgressive origin: a coalescent-based approach. *Bmc Genomics* 17: 8.
546 HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic*
547 *Biology* 47: 3-8.
548 HOSSAIN, A. S. M. M., B. P. BLACKBURNE, A. SHAH, AND S. WHELAN. 2015. Evidence of Statistical
549 Inconsistency of Phylogenetic Methods in the Presence of Multiple Sequence Alignment
550 Uncertainty. *Genome Biology and Evolution* 7: 2102-2116.

JARVIS, E. D., S. MIRARAB, A. J. ABERER, B. LI, P. HOUDE, C. LI, S. Y. W. HO, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346: 1320-1331.

JOLY, S., P. A. MCLENACHAN, AND P. J. LOCKHART. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* 174: E54-E70.

KENT, W. J. 2002. BLAT--the BLAST-like alignment tool. *Genome Research* 12: 656-664.

KIRKPATRICK, M., AND B. BARRETT. 2015. Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Molecular Ecology* 24: 2046-2055.

KONOWALIK, K., F. WAGNER, S. TOMASELLO, R. VOGT, AND C. OBERPRIELER. 2015. Detecting reticulate relationships among diploid *Leucanthemum* Mill. (Compositae, Anthemideae) taxa using multilocus species tree reconstruction methods and AFLP fingerprinting. *Molecular Phylogenetics and Evolution* 92: 308-328.

KUBATKO, L., AND J. CHIFMAN. 2015. An invariants-based method for efficient identification of hybrid speciation from large-scale genomic data. https://www.biorxiv.org/content/10.1101/034348v1.

KUBATKO, L. S. 2009. Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology* 58: 478-488.

KUHNER, M. K., AND J. MCGILL. 2014. Correcting for sequencing error in maximum likelihood phylogeny inference. *G3-Genes Genomes Genetics* 4: 2544-2551.

LARGET, B. R., S. K. KOTHA, C. N. DEWEY, AND C. ANE. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26: 2910-2911.

LEMMON, A. R., S. A. EMME, AND E. M. LEMMON. 2012. Anchored hybrid enrichment for massively high-throughput pylogenomics. *Systematic Biology* 61: 727-744.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN, N. HOMER, G. MARTH, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.

LIU, L., L. L. YU, D. K. PEARL, AND S. V. EDWARDS. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58: 468-477.

MALLET, J. 2005. Hybridization as an invasion of the genorne. *Trends in Ecology & Evolution* 20: 229-237.

------. 2007. Hybrid speciation. *Nature* 446: 279-283.

MALLET, J., N. BESANSKY, AND M. W. HAHN. 2016. How reticulated are species? *Bioessays* 38: 140-149.

MAYR, E. 1942a. Grounds for a Statistical Method in Regional Animal Geography. *Geographical Review* 32: 694-695.

------. 1942b. Zoogeographical Studies of the Tsugitaka Mountains of Formosa. *Geographical Review* 32: 693-694.

MCDADE, L. A. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. *Evolution* 44: 1685-1700.

------. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. *Evolution* 46: 1329-1346.

MORAN, G. F., J. C. BELL, AND A. C. MATHESON. 1980. The Genetic-Structure and Levels of Inbreeding in a Pinus-Radiata D Don Seed Orchard. *Silvae Genetica* 29: 190-193.

MUCHHALA, N., S. JOHNSEN, AND S. D. SMITH. 2014. Competition for hummingbird pollination shapes flower color variation in Andean Solanaceae. *Evolution* 68: 2275-2286.

595 NAKHLEH, L. 2013. Computational approaches to species phylogeny inference and gene tree
596 reconciliation. *Trends in Ecology & Evolution* 28: 719-728.
597 PARDI, F., AND C. SCORNAVACCA. 2015. Reconstructible Phylogenetic Networks: Do Not Distinguish
598 the Indistinguishable. *Plos Computational Biology* 11.
599 PATTERSON, N., P. MOORJANI, Y. T. LUO, S. MALLICK, N. ROHLAND, Y. P. ZHAN, T. GENSCHORECK, et al. 2012.
600 Ancient admixture in human history. *Genetics* 192: 1065-+.
601 PEASE, J. B., AND M. W. HAHN. 2015. Detection and polarization of introgression in a five-taxon
602 phylogeny. *Systematic Biology* 64: 651-662.
603 RATAN, A. 2009. Assembly algorithms for next-generation sequence data. PhD Phd Dissertation,
604 The Pennsylvania State University, University Park, PA, USA.
605 REID, N., J. R. DEMBOSKI, AND J. SULLIVAN. 2012. Phylogeny estimation of the radiation of western
606 North American Chipmunks (Tamias) in the face of introgression using reproductive
607 protein genes. *Systematic Biology* 61: 44-62.
608 RIESEBERG, L. H. 1995. The Role of Hybridization in Evolution - Old Wine in New Skins. *American*
609 *Journal of Botany* 82: 944-953.
610 ROSENZWEIG, B. K., J. B. PEASE, N. J. BESANSKY, AND M. W. HAHN. 2016. Powerful methods for
611 detecting introgressed regions from population genomic data. *Molecular Ecology* 25:
612 2387-2397.
613 RUBIN, B. E. R., R. H. REE, AND C. S. MOREAU. 2012. Inferring phylogenies from RAD sequence data.
614 *Plos One* 7.
615 SALICHOS, L., A. STAMATAKIS, AND A. ROKAS. 2014. Novel information theory-based measures for
616 quantifying incongruence among phylogenetic trees. *Molecular Biology and Evolution*
617 31: 1261-1271.
618 SATO, S., S. TABATA, H. HIRAKAWA, E. ASAMIZU, K. SHIRASAWA, S. ISOBE, T. KANEKO, et al. 2012. The
619 tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635-
620 641.
621 SCHIERUP, M. H., AND J. HEIN. 2000. Consequences of recombination on traditional phylogenetic
622 analysis. *Genetics* 156: 879-891.
623 SMITH, S. D., AND D. A. BAUM. 2006. Phylogenetics of the florally diverse Andean clade
624 Iochrominae (Solanaceae). *American Journal of Botany* 93: 1140-1153.
625 ------. 2007. Systematics of Iochrominae (Solanaceae): Patterns in floral diversity and
626 interspecific crossability. *Acta Horticulturae* 745: 241-254.
627 SMITH, S. D., AND S. LEIVA. 2011. A new species of *Iochroma* (Solanaceae) from Ecuador. *Novon*
628 21: 491-495.
629 SMITH, S. D., V. J. KOLBERG, AND D. BAUM. 2008. Morphological and cytological evidence for
630 homoploid hybridization in *Iochroma* (Solanaceae). *Madroño* 55: 280-284.
631 SMITH, S. D., S. J. HALL, P. R. IZQUIERDO, AND D. A. BAUM. 2008. Comparative pollination biology of
632 sympatric and allopatric Andean *Iochroma* (Solanaceae). *Annals of the Missouri*
633 *Botanical Garden* 95: 600-617.
634 SOLIS-LEMUS, C., AND C. ANE. 2016. Inferring phylogenetic networks with maximum
635 pseudolikelihood under incomplete lineage sorting. *Plos Genetics* 12.
636 STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
637 phylogenies. *Bioinformatics* 30: 1312-1313.

638 STEBBINS, G. L. 1959. The Synthetic Approach to Problems of Organic Evolution. *Cold Spring*
639 *Harbor Symposia on Quantitative Biology* 24: 305-311.
640 STENZ, N. W. M., B. LARGET, D. A. BAUM, AND C. ANE. 2015. Exploring Tree-Like and Non-Tree-Like
641 Patterns Using Genome Sequences: An Example Using the Inbreeding Plant Species
642 Arabidopsis thaliana (L.) Heynh. *Systematic Biology* 64: 809-823.
643 WEN, D. Q., Y. YU, M. W. HAHN, AND L. NAKHLEH. 2016. Reticulate evolutionary history and
644 extensive introgression in mosquito species revealed by phylogenetic network analysis.
645 *Molecular Ecology* 25: 2361-2372.
646 WIENS, J. J. 2003. Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a
647 missing data problem? *Journal of Vertebrate Paleontology* 23: 297-310.
648 WU, F. N., L. A. MUELLER, D. CROUZILLAT, V. PETIARD, AND S. D. TANKSLEY. 2006. Combining
649 bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes
650 (COSII) for comparative, evolutionary and systematic studies: A test case in the
651 euasterid plant clade. *Genetics* 174: 1407-1420.
652 YU, Y., J. H. DEGNAN, AND L. NAKHLEH. 2012. The probability of a gene tree topology within a
653 phylogenetic network with applications to hybridization detection. *Plos Genetics* 8: 456-
654 465.
655 YU, Y., T. CUONG, J. H. DEGNAN, AND L. NAKHLEH. 2011. Coalescent Histories on Phylogenetic
656 Networks and Detection of Hybridization Despite Incomplete Lineage Sorting.
657 *Systematic Biology* 60: 138-149.
658

Table 1. Various approaches that have been used to detect hybridization. Global analyses are those which test for hybridization involving two or more branches of the phylogeny, but do not attempt to localize the events. Local methods aim to determine the number and location of the reticulation events, and therefore which lineages have been involved.

Pattern	Type	Reference
Tree distance	Global	(Buckley et al., 2006)
Count deep coalescences	Global	(Reid, Demboski, and Sullivan, 2012)
Tree distance	Global	(Blanco-Pastor, Vargas, and Pfeil, 2012)
Network likelihood index	Global	(Konowalik et al., 2015)
Genealogical sorting index	Global	(de Villiers et al., 2013)
Shallow interspecific coalescences	Local	(Joly, McLenachan, and Lockhart, 2009)
Species network likelihood	Local	(Kubatko, 2009)
Minimize deep coalescences	Local	(Yu et al., 2011)
Species network likelihood	Local	(Yu, Degnan, and Nakhleh, 2012)
Four taxa nucleotide distances	Local	(Durand et al., 2011)
Five taxa nucleotide distances	Local	(Pease and Hahn, 2015)
Minimum pairwise sequence distance	Local	(Rosenzweig et al., 2016)
Sequence invariant patterns	Local	(Kubatko and Chifman, 2015; Blischak et al., 2018)
Species network pseudolikelihood	Local	(Solis-Lemus and Ane, 2016)

Table 2. Accessions used for this study.

Species	Relationship	Voucher	Locality	DNA ID#
I. arborescens	Putative parent	Smith 312 (MO)	Peru: Contumaza. 7.42409°S 78.90111°W	98
I. arborescens	Putative parent	Smith 209 (WIS)	Ecuador: Alluriquin. 0.32145°S 78.99764°W	250
I. arborescens x I. cyaneum	F1 Hybrid	Smith 687 (COLO)	University of Colorado-Boulder Greenhouses	-
I. ayabacense	Putative hybrid	Smith 337 (MO)	Peru: Ayabaca. 4.61462°W 79.71975°S	126
I. cyaneum	Putative parent	Smith 265 (WIS)	Univ. of Wisconsin-Madison Greenhouses	156
I. lehmannii	Putative parent	Smith 487 (MO)	Ecuador: Cañar 2.37168°W 78.96976°S	228
I. lehmannii	Putative parent	Smith 330 (MO)	Peru: Ayabaca. 4.64422°W 79.71975°S	176
I. stenanthum	Putative hybrid	Smith 313 (MO)	Peru: Contumaza. 7.40116°S 78.89658°W	99
P. peruviana	Outgroup	Smith 217 (MO)	Peru: Quito. 0.16761° S 78.48133° W	91

Table 3. Identification of hybrid relationships with HyDe, D-statistics and SNaQ. In (b), columns A, B, and C represent the phylogenetic position in a three tip relationship ((A,B),C) and the p-value indicates hybridization between A and C. For (a) and (b), the asterisks indicate significance after correction for multiple comparisons ($\alpha = 0.05/30 = 0.0017$ for HyDe and $\alpha = 0.05/10 = 0.005$ for D-statistics). For (c), the asterisks indicate significance in a likelihood ratio test with one degree of freedom (corresponding to the additional parameter introduced by the reticulation). Instances in which the method inferred a set of relationships consistent with the hypothesized hybrid ancestry are highlighted in gray.

(a) HyDe

Dataset	Hybrid	Parent 1	Parent 2	Test statistic	*P*
1. No hybrids	*I. lehmannii* (176)	*I. arborescens* (250)	*I. lehmannii* (228)	3.18	0.002
2. One F_1	*I. arborescens* (98)	F_1	*I. lehmannii* (176)	5.08	0*
3. F_1 + *I. ayabacense*	*I. arborescens* (98)	*I. ayabacense*	*I. cyaneum* (156)	8.99	0*
4. *I. ayabacense* + *I. stenanthum*	*I. cyaneum* (156)	*I. lehmannii* (228)	*I. stenanthum*	8.64	0*

(b) D-statistic analysis

Dataset	A	B	C	Test statistic	*P*
1. No hybrids	*I. lehmannii* (176)	*I. arborescens* (250)	*I. lehmannii* (228)	0.13	0.002*
2. One F_1	F_1	*I. arborescens* (98)	*I. lehmannii* (228)	0.42	0*
3. F_1 + *I. ayabacense*	F_1	*I. ayabacense*	*I. arborescens* (98)	0.25	0*
4. *I. ayabacense* + *I. stenanthum*	*I. ayabacense*	*I. lehmannii* (228)	*I. cyaneum* (156)	0.29	0*

(c) SNaQ analysis

Dataset	Hybrid	Parent 1	Parent 2	Reticulation PLL	Tree PLL
1. No hybrids	*I. lehmannii* (228)	*I. arborescens* (98)	*I. lehmannii* (176)	-1.86	-3.26
2. One F_1	F_1	*I. arborescens* (250)	*I. cyaneum* (156)	-2.11*	-26.80
3. F_1 + *I. ayabacense*	*I. lehmannii* (228)	*I. arborescens* (250)	*I. lehmannii* (228)	-31.99*	-91.32
4. *I. ayabacense* + *I. stenanthum*	*I. lehmannii* (228)	*I. arborescens* (250)	*I. lehmannii* (228)	-133.28*	-257.97

Fig. 1. Known and hypothesized hybrid relationships. (a) *Iochroma cyaneum* and *I. arborescens* were crossed in the greenhouse to create the F_1 hybrid on the right. (b) *I. cyaneum* and *I. lehmannii* are the putative parental taxa of *I. ayabacense*, right. (c) Members of the clade containing *I. cyaneum and I. cornifolium* (shown, left) and the clade containing *I. arborescens* and *I. confertiflorum* (middle) are the putative parental taxa of *I. stenanthum*, right.

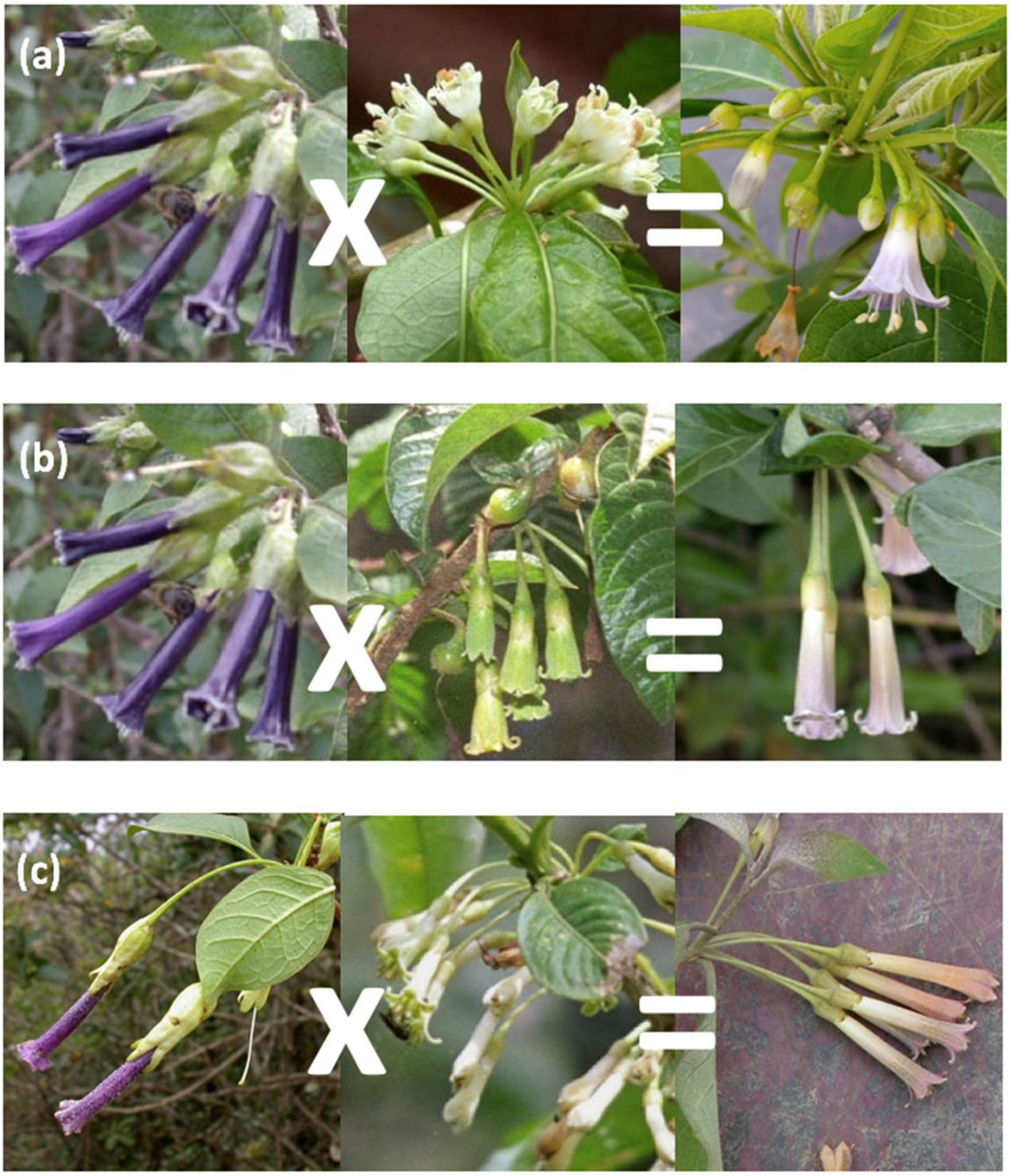

Fig. 2. Hypothesized relationships among sampled tips and experimental design. The top tree shows the relationships among *I. arborescens*, *I. lehmannii*, and *I. cyaneum* based on previous studies (see text). Numbers indicate DNA accession number for reference (Table 2). The lines connect putative or known hybrid taxa (Fig. 1) to their putative parental lineages. The letter scheme indicates the putative parentals (AC denotes an A x C hybrid). Tests for hybridization were carried out on the four six-taxon datasets at the bottom, which range from no putative hybrids to two putative hybrids. See text for complete description.

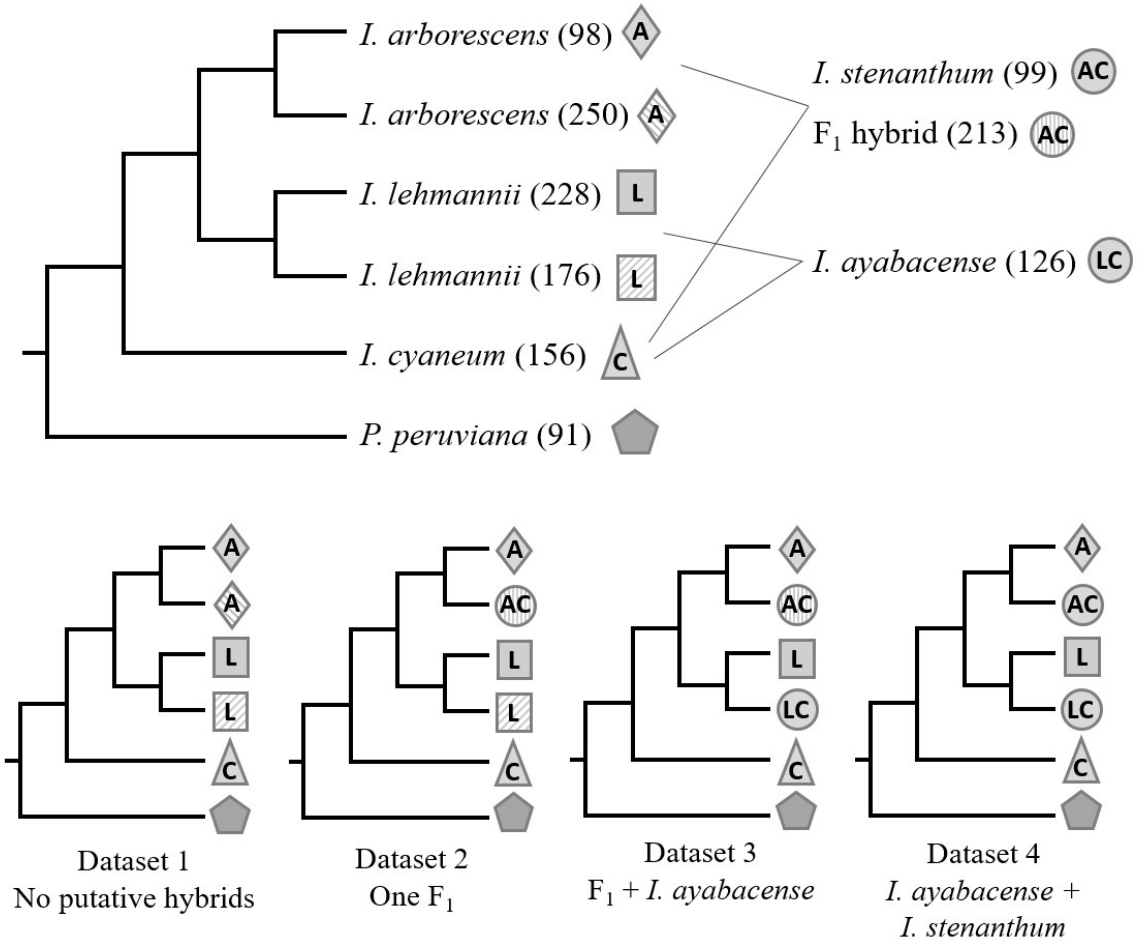

Fig. 3. Gene tree conflict and signal across experimental datasets. Topologies for each dataset follow the species tree inferred by STAR. The symbols for tips follow Fig. 2, with known or putative hybrids indicated with two letters (AC or LC) corresponding to their putative parental lineages. Numbers above the branches are the numbers of gene trees estimated to have this relationship from concordance analysis. Proportions of gene trees (concordance factors) are in parentheses. Numbers below the branches denote internode certainty values, rounded to two decimal places. Tree certainty (TC) are given in the boxes for each tree along with the χ^2 value for two models (panmixia, P, and fully-bifurcating, B) tested by TICR. Values above 8 are significant at P=0.05 (with three degrees of freedom, Stenz et al. 2015), suggesting that the data reject the model. For example, dataset 4 strongly rejects a bifurcating tree (χ^2_B of 94) but does not reject the panmictic model (χ^2_P of 1.7).