Tracking of Host Defenses and Phylogeny During the Radiation of Neotropical *Inga*-Feeding Sawflies (Hymenoptera; Argidae)

María-José Endara[1,2*†], James A. Nicholls[3†], Phyllis D. Coley[1,4], Dale L. Forrister[1], Gordon C. Younkin[1], Kyle G. Dexter[5,6], Catherine A. Kidner[3,5], R. T. Pennington[5], Graham N. Stone[3‡] and Thomas A. Kursar[1,4‡]

[1] Department of Biology, The University of Utah, Salt Lake City, UT, United States, [2] Centro de Investigación de la Biodiversidad y Cambio Climático (BioCamb) e Ingeniería en Biodiversidad y Recursos Genéticos, Facultad de Ciencias de Medio Ambiente, Universidad Tecnológica Indoamérica, Quito, Ecuador, [3] Ashworth Laboratories, Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh, United Kingdom, [4] Smithsonian Tropical Research Institute, Panama City, Panama, [5] Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, [6] School of GeoSciences, The University of Edinburgh, Edinburgh, United Kingdom

Coevolutionary theory has long predicted that the arms race between plants and herbivores is a major driver of host selection and diversification. At a local scale, plant defenses contribute significantly to the structure of herbivore assemblages and the high alpha diversity of plants in tropical rain forests. However, the general importance of plant defenses in host associations and divergence at regional scales remains unclear. Here, we examine the role of plant defensive traits and phylogeny in the evolution of host range and species divergence in leaf-feeding sawflies of the family Argidae associated with Neotropical trees in the genus *Inga* throughout the Amazon, the Guiana Shield and Panama. Our analyses show that the phylogenies of both the sawfly herbivores and their *Inga* hosts are congruent, and that sawflies radiated at approximately the same time, or more recently than their *Inga* hosts. Analyses controlling for phylogenetic effects show that the evolution of host use in the sawflies associated with *Inga* is better correlated with *Inga* chemistry than with *Inga* phylogeny, suggesting a pattern of delayed host tracking closely tied to host chemistry. Finally, phylogenetic analyses show that sister species of *Inga*-sawflies are dispersed across the Neotropics, suggesting a role for allopatric divergence and vicariance in *Inga* diversification. These results are consistent with the idea that host defensive traits play a key role not only in structuring the herbivore assemblages at a single site, but also in the processes shaping host association and species divergence at a regional scale.

Keywords: coevolution, defense traits, herbivores, host tracking, *Inga*, plant–insect interactions, sawflies, tropical rain forests

INTRODUCTION

Insect herbivores and their plant hosts dominate terrestrial biodiversity (Hunt et al., 2007), and the processes that drive their interaction and diversification remain an enduring focus of research in ecology and evolution (Futuyma and Agrawal, 2009; Janz, 2011; Hembry et al., 2014; Forbes et al., 2017; Nakadai, 2017). This is especially true in the tropics where most of the species occur.

A central paradigm is that insect-plant associations have been shaped by arms race coevolution between plant defenses and insect countermeasures (Becerra, 1997; Becerra et al., 2009; Volf et al., 2018). Ehrlich and Raven (1964) observed that closely related plants are often attacked by closely related herbivores, a pattern they attributed to an 'escape and radiate' model, in which plant lineages diversify following evolutionary innovation of a key defense trait, and specialist herbivore lineages diversify across the plant radiation through evolution of a key countermeasure (Wheat et al., 2007). Where these traits are phylogenetically conserved in each lineage, we expect some degree of phylogenetic concordance between plant and herbivore lineages, resulting either from simultaneous co-diversification (Cruaud et al., 2012), or delayed herbivore colonization of an existing plant radiation (tracking of host resources; Janz, 2011). Thus, plant defenses play a prominent role in the evolution of host associations (Thompson, 1988), yet they are often not considered in studies of plant-herbivore diversification. Robust analyses require not only phylogenetic histories of both plants and herbivores, but also data on ecologically important traits such as plant defenses.

Although insect herbivores are expected to show evolutionary conservatism in host use (Ehrlich and Raven, 1964; Brooks and McLennan, 2002), many studies show herbivore shifts between distantly related hosts that disrupt any signature of codiversification. Some shifts are between hosts with similar chemical defenses for which herbivore countermeasures are to some extent preadapted, implying a process of host-resource tracking (Janz, 2011; Endara et al., 2017) or ecological fitting (Agosta and Klemens, 2008). Insect herbivores can also radiate across hosts with contrasting defensive traits through diversification of specialist host races, leading to ecological speciation (Nyman, 2010; Hardy and Otto, 2014). These alternative mechanisms of divergence without codiversification do not happen in isolation, and their impacts are expected to reflect the distributions of interacting lineages through time and space (Hoberg and Brooks, 2008; Züst et al., 2012; Calatayud et al., 2016). Assessing the contribution of these alternative mechanisms to observed patterns of interaction and diversity is thus a major challenge (Hembry et al., 2014; Russo et al., 2017). In particular, and with notable exceptions (e.g., Kursar et al., 2009; Wilson et al., 2012; Fine et al., 2013; Marquis et al., 2016; Salazar et al., 2016; Endara et al., 2017; Volf et al., 2018), we know little about the processes driving plant-herbivore diversification in the tropical rainforest areas that harbor most of terrestrial biodiversity (López-Carretero et al., 2018).

Here we explore the factors structuring associations between insect herbivores and neotropical trees in the genus *Inga,* a species-rich radiation that shows high local species richness and abundance in many habitats across the Neotropics, and which is characterized by high diversity of chemical, physical and developmental defenses against insect herbivores (Kursar et al., 2009). Previous analyses support a key role for *Inga* defensive chemistry in structuring lepidopteran herbivore assemblages at a single site (Endara et al., 2017), and non-random combinations of defensive traits across sites imply a role for herbivore avoidance in *Inga* community assembly (Kursar et al., 2009). It remains unclear, however, whether the same *Inga* traits structure herbivore associations in widely separated communities. Previous analyses have also found little phylogenetic pattern in *Inga* defenses (Kursar et al., 2009; Endara et al., 2015, 2017), and related lepidopteran herbivores attack *Inga*s with similar defenses, rather than those that are closely related (Endara et al., 2017). Our hypothesis is that herbivores have driven rapid diversification of defensive traits in *Inga*, with herbivore associations resulting from evolutionary tracking of similar defensive phenotypes (i.e., host-resource tracking) rather than cospeciation (Coley et al., 2018). Here we test this hypothesis using data for four regional communities that span the Amazon Basin, in Panama, Peru, Ecuador, and French Guiana. We focus on sawflies (Hymenoptera; Symphyta) in the superfamily Tenthredinoidea, shown in previous work in other regions of the world to be highly sensitive to (and often dependent on) toxic host plant chemistry (Petre et al., 2007; Boevé et al., 2013; Naya et al., 2016). Thus, they are an excellent candidate taxon in which to explore the impact of diversification in this key aspect of *Inga* defenses. We use novel data on *Inga*-sawfly associations and an analytical approach incorporating phylogenies for both lineages (Hadfield et al., 2014) and defense trait data to address the following questions: (i) Is there phylogenetic patterning in *Inga* defenses? (ii) Does *Inga* phylogeny (cospeciation) or defenses (resource tracking) best predict *Inga*-sawfly associations? (iii) Over what geographic scale have *Inga*-sawfly associations evolved? Are sister sawfly or *Inga* species commonly members of the same regional community, implying local, sympatric diversification? Or are sister taxa dispersed across the Neotropics, suggesting a role for allopatric divergence and vicariance in one or both trophic levels?

MATERIALS AND METHODS

Sampling and Quantification of *Inga* Defensive Traits

We sampled 81 *Inga* species and 3 *Zygia* species (a sister clade of *Inga*) at four sites throughout the Amazon and Panama between July 2010 and September 2014: Panama (January-February 2010; Smithsonian Tropical Research Institute on Barro Colorado Island, 9.150°N, 79.850°W), French Guiana (July–August 2011 and 2012; Nouragues Station, 4.08°N, 52.683°W) Peru (July–October 2010 and 2011; Los Amigos Biological Station, Madre de Dios, 12.567°S, 70.100W) and Ecuador (July–September 2013 and 2014; Tiputini Biodiversity Station, 0.638°S. 76.150°W). In each location, we sampled expanding leaves of 0.5–4 m tall understory saplings. Host associations were recorded on c. 60 young leaf flushes per tree species. Sawfly larvae were found on 34 *Inga* and 2 *Zygia* species comprising from 1 to many gregarious larvae on a specific individual host plant (**Supplementary Table S2**).

We measured multiple defensive traits that capture the entire defensive profile of each species. These include developmental defenses (leaf expansion rate and chlorophyll content), biotic defenses (mean number of ants visiting the leaves and extra-floral nectary size) and chemical defenses. This set of defense traits was measured only on expanding leaves because more than 80% of

the damage accrued during the leaf's lifetime happens during the short period (1–3 weeks) of leaf expansion (Coley et al., 2018).

Developmental Defenses

Young leaves can expand rapidly, which shortens the window of vulnerability to herbivores, and they can delay chloroplast development, which reduces the impact of a given amount of damage (Kursar and Coley, 1992). Leaf expansion rate was determined as the percent increase in area per day for c. 13 individuals per species. Chloroplast development was measured as the chlorophyll content (mg dm^{-2}) of leaves between 30 and 80% of full expansion for c. 30 individuals per species (Endara et al., 2017). Since these two traits are correlated, we treat them as a single defense.

Biotic Defenses

Inga leaves have extra-floral nectaries that produce nectar and attract protective ants only during the short period of leaf expansion. We quantified the diameter of these nectaries and the abundance of ants visiting them (# of ants per nectary) in c. 30 individuals per species.

Chemical Defenses

For chemical analyses, expanding leaves were dried in the field over silica at ambient temperature. Although *Inga* has little quantitative or qualitative induction of young leaf defenses (Bixenmann et al., 2016), samples used for chemical analyses were from plants without sawflies. The chemical defensive profile for each species was determined using metabolomics. Metabolites were extracted at the Coley/Kursar laboratory in the University of Utah in 44.3 mmol L^{-1} ammonium acetate, pH 4.8:acetonitrile (60:40, v/v) and analyzed following the protocol of Wiggins et al. (2016). Metabolites with intermediate polarity were analyzed by ultraperformance C18 liquid chromatography coupled to mass spectrometry (UPLC-MS) in negative mode. Raw data from the UPLC-MS analysis in MassLynx were converted to mzXML format using mzConvert (Chambers et al., 2012) and then processed for peak detection, peak alignment and peak filtering using the R package XCMS (Smith et al., 2006; Tautenhahn et al., 2008; Benton et al., 2010). These results were post-processed in the R package CAMERA to assign the various ions derived from one compound (termed 'features') to that compound (Kuhl et al., 2012), as detailed in **Appendix SII**. This analysis yields 2621 compounds from the 36 plant species. Purification and structure determination by 2-D NMR of several dozen compounds, as well as matching MS-MS spectra from our in-house database to the GNPS databases (Global Natural Products Social Molecular Networking)[1] suggest that, for *Inga,* these compounds are mainly phenolics, saponins and amines. None are primary metabolites (**Supplementary Table S1**). All scripts from this study are deposited in github[2].

Overexpression of the essential amino acid, L-tyrosine, ranges from 5 to 20% leaf DW (Dry Weight) in certain species of *Inga.* At these concentrations, it is highly toxic to non-adapted herbivores, and therefore functions as an important chemical

defense (Lokvam et al., 2006). Because tyrosine is insoluble in our extraction buffer, tyrosine concentration as percent of leaf dry weight was determined separately following Lokvam et al. (2006, **Appendix SII**).

Sawfly DNA Barcoding

Taxonomic resources are limited even for adult sawflies (Schmidt et al., 2017), and very few exist for morphological identification of neotropical sawfly larvae to species. We therefore adopted a DNA barcoding approach using sequences for a 645 base pair (bp) fragment of the mitochondrial gene cytochrome oxidase I (*COI*) (For DNA methods, see **Appendix SI**). Every individual was barcoded. For gregarious species, we sequenced a minimum of three in a group and in all cases these belonged to the same MOTU. Sequences were allocated to MOTUs (molecular operational taxonomic units) using two approaches: jMOTU v1.0.8 (Jones et al., 2011) and ABGD (Automatic Barcode Gap Discovery, Puillandre et al., 2012). jMOTU clusters sequences into MOTUs that differ by pre-defined numbers of bases; we examined divergence distances amongst sequences ranging from 1 to 65 bp, with a low BLAST identity filter of 97%. In the presence of a barcoding gap (Puillandre et al., 2012), a plot showing numbers of MOTUs as a function of sequences divergence should form a plateau, with no change in MOTU number across the divergence levels corresponding to the gap (Acs et al., 2010). ABGD defines MOTUs based upon prior values of within-species divergence, and assesses how MOTU number changes as within-species divergence increases. We used prior within-species divergence limits ranging from 0.3 to 6%, split into 30 steps. We used the K2P distance measure, with a transition to transversion ratio of 1.47, as estimated by jModeltest v2.1.7 (Darriba et al., 2012), and the default value of 1.5 for slope increase. Output from the recursive partitioning scheme was used, with the final number of MOTUs chosen at the point where the plot of MOTU versus intraspecific divergence leveled off. Both approaches gave highly concordant results.

Because mitochondrial haplotypes can be shared among species, and hence give misleading indications of species membership in sawflies (Prous et al., 2011; Schmidt et al., 2017) and more widely (Funk and Omland, 2003; Nicholls et al., 2012), we sequenced our candidate *COI* MOTUs for two nuclear loci, *wingless* (coding, 327 bp; $n = 75$ sequences) and *ITS2* (non-coding, 609 bp; $n = 80$; for molecular methods, see **Appendix SI**). Sampling incorporated all singleton MOTUs and 2–4 individuals of MOTUs with more extensive sampling. *COI* sequence data were highly effective in resolving relationships between sawfly samples with high posterior probability (**Supplementary Figure S1**). The first barcoding gap using jMOTU was apparently at 7–10 bp (1–1.5% divergence), identifying 42 MOTUs. These were highly concordant with 40 MOTUs identified by ABGD for sequence divergence from 1.03 to 1.41% (**Supplementary Figure S2**), the only difference being that jMOTU split two of the 40 ABGD MOTUs into two. Relationships between MOTUs identified using *COI* data were highly concordant with those based on nuclear *ITS2* and *wingless* (**Supplementary Figures S3, S4**). Our final sawfly MOTU definitions ($n = 41$) incorporated information from both mitochondrial and nuclear data, with

---

[1] https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp
[2] https://github.com/ColeyKursarLab/endara_sawflies_2018

COI MOTUs retained if at last one nuclear gene showed the same clustering of individuals. Sawfly MOTUs were allocated to candidate taxonomic families by querying each against voucher sequences in the Barcoding of Life BOLDSYSTEMS database[3].

Gene trees for each of the three loci were generated using MrBayes *v3.2.2* (Ronquist et al., 2012). Based on relative numbers of variable sites at each codon position, *wingless* was treated as a single partition while *COI* was partitioned between codon positions 1 + 2, and position 3. As a non-coding locus, *ITS2* was treated as a single partition. We used the closest available substitution model in MrBayes as per the recommendation provided by *jModeltest* (Guindon and Gascuel, 2003; Darriba et al., 2012), as follows: *COI*(1,2), GTR+I+G; *COI*(3), GTR+G; *wingless*, GTR+I+G; *ITS2*, GTR+G. MrBayes analyses were run for 20 million generations for *ITS2* and *wingless*, sampling every 2500 generations, with a burn-in of 16 million generations. The analysis for *COI* was run for 40 million generations to achieve convergence, sampling every 5000 generations, with a burn-in of 32 million generations. Likelihood comparisons showed a relaxed IGR clock model to be better supported than either no clock or a strict clock for all loci.

Phylogenetic Relationships Among Sawfly MOTUs

We determined phylogenetic relationships for our MOTUs at two levels. To place our *Inga*-feeding sawfly MOTUs in a wider phylogenetic context, we carried out additional phylogenetic analyses using data for *COI* and an additional coding nuclear locus, *PGD* (496 bp). *PGD* data provided high resolution in a previous wide-ranging phylogenetic analysis of sawflies (Malm and Nyman, 2015), and allow us to place our MOTUs within this taxonomic framework. Analyses for each gene incorporated data on sawfly species from recent phylogenetic (Schulmeister et al., 2002; Malm and Nyman, 2015) and barcoding (Hartsough et al., 2007; Schmidt et al., 2017) surveys of sawflies. These analyses identified related taxa on the basis of nearest matches identified from BOLD. The taxa from the surveys that we added to our analysis comprised 11 species in the family Tenthredinidae, 9 in the family Pergidae and 34 in the family Argidae, none of which are neotropical. We also included a similar number of sequences for neotropical taxa, and the only available *COI* voucher sequence for an *Inga*-feeding sawfly, a specimen of *Ptenos leucopoda* (Argidae) sampled from *Inga oerstediana* (and also recorded from *I. vera*) in Costa Rica (Smith et al., 2013). Metadata and Genbank accession numbers for these reference sequences are provided in **Supplementary Table S3**. We constructed gene trees for each locus using MrBayes, using the closest available substitution model to that identified as appropriate using *jModeltest* (Guindon and Gascuel, 2003; Nylander, 2004; Darriba et al., 2012). Based on relative numbers of variable sites at each codon position, *PGD* data were modeled in two partitions, 1 + 2, and 3, each with a GTR+I+G model, while *COI* was divided into three partitions by codon, each with a GTR+I+G model. For each gene we assumed a relaxed clock, with a birth-death speciation model. To provide an order of magnitude age for *Inga*-associated

sawfly lineages, we calibrated the *COI* tree using two alternative estimates: the Brower rate estimate of 0.0115 substitutions per million years (Brower, 1994) and the higher rate of 0.0177 derived by Papadopoulou et al. (2010).

For analysis of evolutionary dynamics in sawfly *Inga* trophic associations, we generated an overall species (MOTU) tree using data for all four loci (*COI, ITS2, PDG, wingless;* 2077 bp) for the 39 MOTUs identified as putative Argidae using the Bayesian *BEAST algorithm (Heled and Drummond, 2010) within BEAST v2.4.1 (Drummond et al., 2012). The *BEAST model used 5 partitions with the following substitution models: *COI* (codon positions 1,2), TN+I+G; *COI* (codon position 3), TN+G; *wingless*, GTR+I+G; *ITS2*, GTR+G; *PGD*, GTR+I+G. We used a Yule speciation model, and compared likelihood support for each combination of relaxed *versus* strict clock models and constant *versus* linearly changing population size. This approach supported a constant population size and an independent relaxed lognormal clock for each partition. We carried out two runs of the *BEAST analysis, with outputs combined in Logcombiner, part of the BEAST suite (Drummond et al., 2012). Each run was for 500 million generations, sampling every 62500 generations, with a burn-in of 300 million generations. Analysis of run diagnostics in Tracer v1.6 (Rambaut and Drummond, 2007) showed all parameters to have an effective sample size of > 100.

Generation of an *Inga* Species Tree

We constructed a species tree for 77 *Inga* accessions representing the taxa from which sawflies were collected, using data for ten coding nuclear loci previously identified as being phylogenetically informative in a wider study of *Inga* phylogenomics (Nicholls et al., 2015) (**Supplementary Table S4**). Aligned sequences for each locus in all *Inga* specimens are available from the Dryad Digital Repository[4]. The ten loci ranged in length from 272 to 2767 bp, with 9–14.7% of sites variable, and spanned a total of 16,125 bp (**Supplementary Table S4**). All ten loci were sequenced in all 77 *Inga* accessions. We co-estimated gene tree topologies and an overall species tree topology using *BEAST, as described above. We used the substitution model previously identified for each locus by Nicholls et al. (2015) (**Supplementary Table S4**). We specified a Yule speciation model and assumed a constant population size. We selected a relaxed lognormal clock over a strict clock model based on very high Bayes factor support (574, estimated as 2Ln harmonic mean likelihood) following criteria in Kass and Raftery (1995). Our analyses ran for 500 million generations, sampled every 62,500 generations, with a burn-in of 50 million generations. Analysis of run diagnostics in Tracer v1.6 (Rambaut and Drummond, 2007) showed all parameters to have an effective sample size of > 100.

Data Analysis
Estimation of Sampling Effort

Sawfly MOTU accumulation curves were generated in the Vegan R package using sampling over *Inga* species [specaccum(data, "random")] and sampling over sawfly individuals [specaccum(data, method = "rarefaction")]. The

---

[3]http://boldsystems.org

[4]doi: 10.5061/dryad.8403km4

"random" method finds the mean accumulation curve and its standard deviation from random permutations of the data. The "rarefaction" method finds the expected species richness and its standard deviation by sampling individuals instead of sites. It achieves this by applying function "rarefy" with number of individuals corresponding to average number of individuals per *Inga* species – which for our data is 1286 sawflies/34 plant taxa = 38 individuals.

Chemical Similarity Between Species of *Inga*

We analyzed data for phenolics and saponins separately. Saponins were defined as all compounds with chromatographic retention time > 18 min and m/z > 580 for the precursor ion, with the remainder classified as phenolics. For several *Inga* species, early eluting compounds have been purified and their structures elucidated by 2D-NMR (J. Lokvam, unpublished). This shows that the bulk of early eluting compounds are phenolics. For about 10 species, the late-eluting fraction was separated from phenolics, hydrolyzed to remove sugars, the triterpene aglycons isolated and their structures elucidated by 2D-NMR (J. Lokvam, unpublished). This work indicates that the bulk of the late-eluting compounds are saponins. Certainly, we cannot rule out that some peaks may belong to other classes. Compounds that are shared across species were matched based on m/z (mass to charge ratio) and retention time. Because many compounds, 1097 out of 2621, are found in only one species, we also quantified species similarity based on the structural similarity of unshared compounds. This matters because unshared compounds are typically treated as having zero relationship even though they may have significant structural similarity. In metabolomics, molecules can be identified based on whether the MS fragmentation pattern (MS/MS spectrum) of an unknown matches spectra in curated databases. A limitation is that these databases include few secondary metabolites, providing little opportunity to quantify the structural relatedness of similar molecules. A recent advance is to quantify the similarity of the MS/MS spectra of a large number of molecules. These data generate a network using the online workflow at the Global Natural Products Social Molecular Networking site (GNPS)[5]. In the resulting network, each node or circle represents a unique compound, with edges (lines) connecting nodes based on structural similarity. Each pair of compounds is assigned a structural similarity score ranging from 0 (completely dissimilar) to 1 (identical) based on the similarity of their MS/MS fragmentation spectra (Watrous et al., 2012). To accomplish this, we obtained as many MS/MS spectra as possible, for 1925 out of our 2621 study compounds. See **Appendix SII** for MS/MS methods and calculation of the chemical similarity of species from molecular networks.

We constructed a dendrogram of chemical similarity between species by fitting a hierarchical clustering model to the equally weighted chemical similarity matrix with 10,000 permutations using the R package PCVLUST (Suzuki and Shimodaira, 2014). For more details see **Appendix SII**.

Because there are many possible equations and data transformations for calculating species similarity scores, we

---

[5]https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp

compared several of these alternatives to lepidopteran dietary preferences following Endara et al. (2017). These analyses validated our method (**Appendix SII**).

Relationship Between Plant Traits and Phylogenetic Signal

Phylogenetic signal was evaluated for continuous host defensive trait data (developmental and biotic defenses), and for the principal coordinates of the chemistry similarity matrix using Blomberg's K (Blomberg et al., 2003). K is close to zero for traits lacking phylogenetic signal, but close to one for traits whose values through the phylogeny match expectations under a Brownian model of evolution. We used the function *phylosig* in the R package *phytools* v.0.6-44 (Revell, 2017).

Analysis of Herbivore–Host Plant Associations

Due to the gregarious habit of sawflies, we use incidence data (presence–absence) for analyses of host associations. Thus, if a specific MOTU was associated with a specific *Inga* host plant in several sampling events on the same plant, it would have been counted only once. To determine the extent to which host phylogeny and/or host defenses structure the associations between sawflies and their hosts, we used maximum likelihood to model the probability of sawfly occurrence (p) using a binomial distribution with the number of trials equal to the total number of herbivore species associated with each *Inga* species. These analyses included all *Inga* species, even those on which sawflies were never found, so that we could determine which *Inga* traits predict an association with any sawfly MOTU. We fitted models that incorporated only the intercept, and the effects of one or more *Inga* defensive traits and the principal coordinates of the phylogenetic distance matrix and the chemical similarity matrix using the R packages *bbmle* v.1.0.20 (Bolker, 2017) and *emdbook* v.1.3.9 (Bolker, 2016). For these analyses, we used the whole *Inga* phylogeny (unpublished *Inga* phylogeny, Nicholls et al., unpublished). The models were run using sampling effort as a covariate (number of leaf flushes searched per *Inga* species). We performed model comparison based on Akaike Information Criterion for small sample sizes (AICc).

Evolutionary interactions between sawflies and *Inga* hosts were determined using a Bayesian approach with generalized linear mixed-effects models (GLMM) in the R library *MCMCglmm* (Hadfield and Nakagawa, 2010; Hadfield, 2017). We performed these analyses only with those *Inga* species that are associated with sawflies. Following Hadfield et al. (2014), we partitioned variance in the sawfly incidence data per *Inga* host into the effects of the phylogenetic histories of plants and herbivores, whether in isolation (termed evolutionary effects by Hadfield et al., 2014) or as interactions (a coevolutionary effect), and chemical similarity between *Inga* hosts (a defense effect). This model approach also allows the estimation of other factors, where interactions have evolved independently of the phylogenies and *Inga* chemistry similarity. The magnitude of the effect for each term is determined by the magnitude of the variance. Following Hadfield et al. (2014), the first term in the model captures the effect of the geographic region information

(here termed Geographical region). The second term determines the contribution of the main effect of the sawfly phylogeny to the covariance and captures the variation in host range explained by the phylogeny (Phylogenetic main effect for sawflies). The third term is the contribution of the main effect of *Inga* chemistry to the covariance and captures the variation in sawfly species richness explained by chemical similarity between *Inga* hosts (Defensive main effect for *Inga* hosts). The fourth term is the contribution of the main effect of *Inga* phylogeny to the covariance and captures the variation in sawfly species richness explained by the phylogeny of the *Inga* hosts (Phylogenetic main effect for *Inga* hosts). The fifth term captures the degree to which related *Inga* have similar sawfly assemblages irrespective of sawfly phylogeny (Phylogenetic *Inga* evolutionary effect). The sixth term captures the degree to which species that are similar in chemistry have similar sawfly assemblages irrespective of sawfly phylogeny (*Inga* defense interaction). The seventh term captures the degree to which related sawflies have similar *Inga* hosts assemblages irrespective of *Inga* phylogeny (Phylogenetic parasite evolutionary effect). The eighth term is the contribution of the coevolutionary interaction to the covariance and captures the degree to which related sawflies feed on related *Inga* (Coevolutionary effect). The ninth term is the contribution of the interaction between *Inga* chemistry and sawfly phylogeny and captures the degree to which related sawflies feed on *Inga* that are similar in chemistry (Defense tracking effect). The last three terms capture interspecific variation in host range (Main effect for sawflies), interspecific variation in sawfly species richness (Main effect for *Inga* hosts) and associations between specific *Inga* hosts and sawflies species (Interaction effect) not due to phylogeny or chemistry.

Phylogeny and chemistry were incorporated into the model as variance-covariance matrices of relatedness and similarity, respectively, in the random effect structure of the generalized linear mixed effect model. We compared models that included site effects (analyses at large spatial scales, as a random factor) and which controlled for sampling effort (as a fixed factor), with models that ignored between-site patterns (hence, analyses at small spatial scales) and sampling effort completely. For the analyses, parameter-expanded priors were used for all variance components following Hadfield et al. (2014). The chain was run for 500,000 iterations with a burn-in of 50,000 and a thinning interval of 450. Because the response variable was incidence data, a Bernoulli error distribution was applied. Models were fitted using the R package *MCMCglmm* v.2.23 (Hadfield, 2017).

Correlations between sawfly phylogenetic relationships with host plant phylogenetic relationships and with host plant chemistry were explored using the function *parafit* (Legendre et al., 2002) in the R package *Ape* v.5.0 (Paradis et al., 2004). We used the global test in *parafit* to test the null hypotheses that (i) the evolution of sawflies and *Inga*, as revealed by the two phylogenetic trees and their trophic associations, has been independent; and (ii) by substituting the *Inga* chemogram for the *Inga* phylogeny that sawfly diversification has been independent of host plant chemistry. Pairwise patristic distances were extracted between sawfly MOTUs from the 4-locus Argidae species tree, and between their corresponding *Inga* host plants

from the 10-locus species tree and *Inga* chemogram using the *cophenetic.phylo* command in *Ape*. *Parafit* analyses used 9999 permutations. Matches between the sawfly phylogeny and each of the *Inga* phylogeny and chemogram were optimized using the function *cophylo* in the R package *phytools* (Revell, 2017).

Visualization of the *Inga*-sawfly associations in phylogenetic space was performed using a Principal Component Analysis. Using the function *phylomorphospace* in the R package *phytools* (Revell, 2017), phylogenetic relationships between sawfly MOTUs was mapped onto *Inga* phylospace. For this analysis, we use the whole *Inga* phylogeny (unpublished *Inga* phylogeny, Nicholls et al., unpublished).

RESULTS

Inga Sawflies Are a Diverse Monophyletic Radiation of Specialist Herbivores Within the Family Argidae

Our *COI* barcoding approach identified 41 MOTUs for sawflies feeding on *Inga* and *Zygia* host plants (**Supplementary Figure S1**), differing by 7–10 bp (1–1.5% divergence). Each sawfly MOTU attacked a very narrow range of 1–2 host *Inga* species, and each *Inga* species only hosted a small number of sawfly MOTUs. This pattern is consistent with the MOTU accumulation curve across sampled sawfly individuals, which suggested that adding more *Inga* taxa to the sampling would only add more specialist sawflies (e.g., sawfly MOTU accumulation curve across *Inga* species rise sharply, **Supplementary Figure S5**). In addition, because the sawfly MOTU accumulation curve across sampled sawfly individuals is asymptotic, this indicates that a more extensive sampling would not yield many additional *Inga*-sawfly interactions (**Supplementary Figure S5**). Thirty-nine MOTUs were identified by BOLD query as likely members of the family Argidae, while the remaining two were most similar to sequences for species in the family Tenthredinidae (**Supplementary Table S2**). Phylogenetic analysis showed that the 39 putative Argidae comprise a well-supported monophyletic clade within this family for the nuclear PGD locus (**Figure 1**; clade posterior probability = 1.0) and also for the more extensive taxon set sequenced for mitochondrial *COI* (**Supplementary Figure S6**). The remaining two MOTUs were placed within a strongly supported clade of voucher sequences for the family Tenthredinidae (**Figure 1**; PP = 0.99). Calibrations of the mutation rate for *CO1* estimate the median age of the common ancestor of this Argidae clade at 6.27 (95% confidence interval 4.78–7.93) million years using the Brower (1994) estimate and 5.31 (4.05–6.72) million years using the Papadopoulou et al. (2010) estimate.
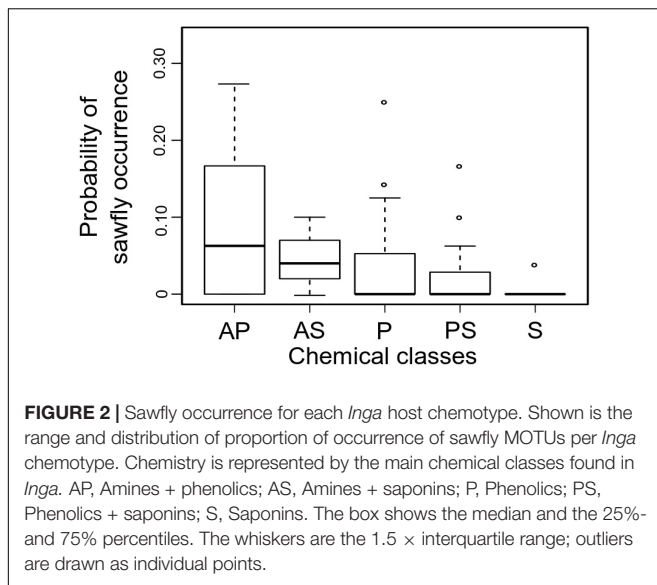
Sawflies Feed on a Chemically Distinct Subset of Available *Inga* Hosts

We investigated the specific role of each host trait in predicting sawfly *Inga* associations in analyses including joint-absence information (i.e., analyses including observations where sawflies were never collected on certain species of *Inga*). We found
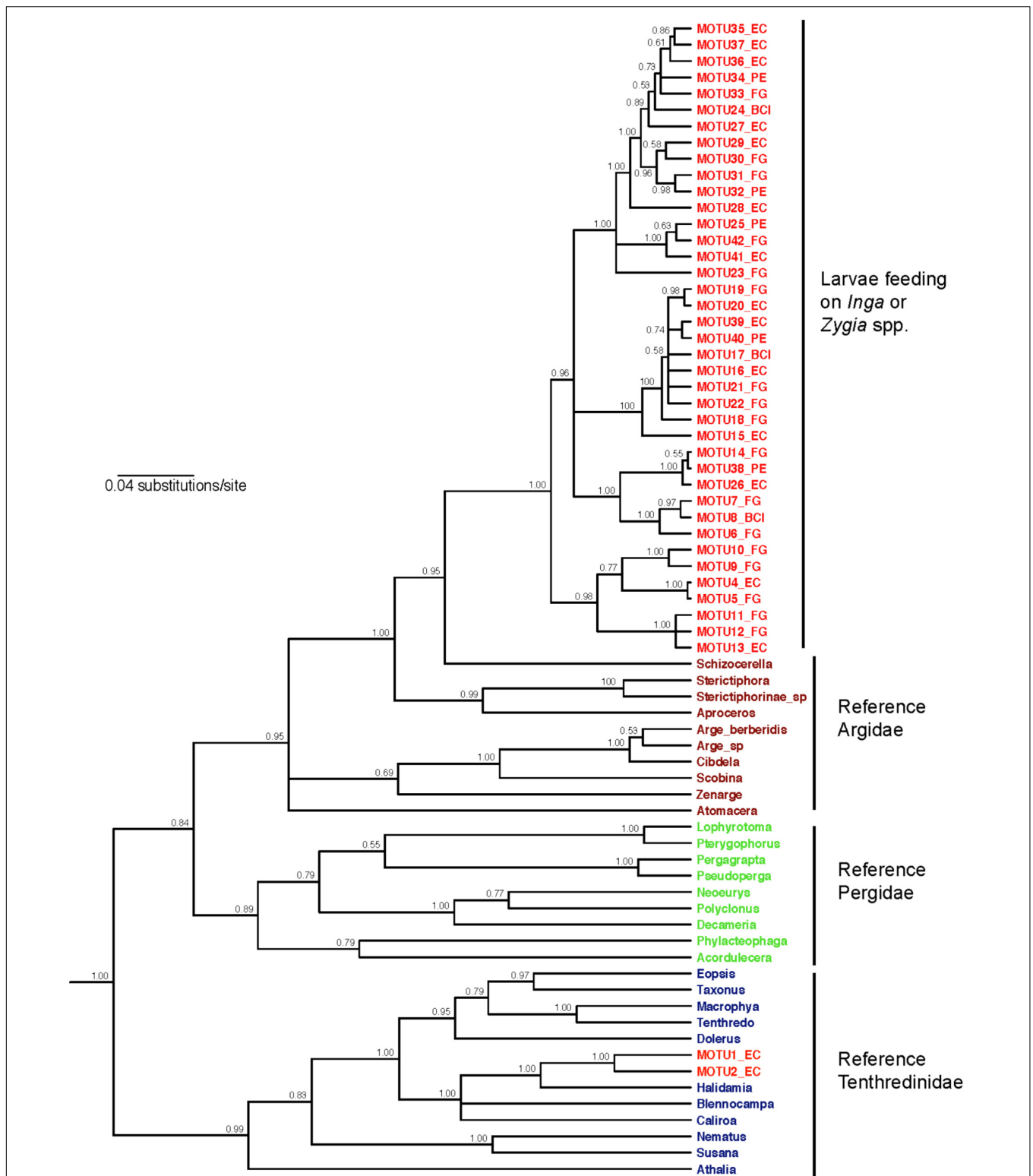
FIGURE 1 | Phylogenetic relationships for the gene *PGD* among the *Inga*-feeding sawfly MOTUs and a panel of voucher sequences for sawflies in the families Argidae, Pergidae (sister group to Argidae; Malm and Nyman, 2015) and Tenthredinidae. The tree shown is a majority-rule consensus tree constructed in MrBayes, using substitutions modeled as GTR+I+G for 1st and 2nd codon positions combined, and GTR+I+G for 3rd positions. We used a relaxed clock, with a birth-death speciation model. Numbers at nodes indicate posterior probability. Taxon labels are colored by sampling source: red MOTU numbers are larvae found feeding on *Inga* or *Zygia*, while other colors indicate reference sequences for adult Argidae, Pergidae, and Tenthredinidae.

FIGURE 2 | Sawfly occurrence for each *Inga* host chemotype. Shown is the range and distribution of proportion of occurrence of sawfly MOTUs per *Inga* chemotype. Chemistry is represented by the main chemical classes found in *Inga*. AP, Amines + phenolics; AS, Amines + saponins; P, Phenolics; PS, Phenolics + saponins; S, Saponins. The box shows the median and the 25%- and 75% percentiles. The whiskers are the 1.5 × interquartile range; outliers are drawn as individual points.

that similarity in chemical defenses among *Inga* hosts was the most important predictor for the occurrence of sawflies in general [proportional odds estimate for PCO1 = 0.26, (95% CI = 1.3 – 0.04), proportional odds estimate for PCO2 = 0.13, (95% CI = 0.95 – 0.02)]. Specifically, sawflies as a group prefer hosts that are defended by amine metabolites [proportional odds estimate for the presence of amines = 1.52, 95% CI (9.89 to 0.41), **Figure 2**], while the probability of occurrence of sawflies decreases with the presence of saponins [proportional odds estimate for the presence of saponins = 0.18, 95% CI (1.99 to 0.008), **Figure 2**].

Closely Related *Inga* Hosts Fed on by Sawflies Are Similar in Chemical and Developmental Defenses

For the *Inga* that were fed upon by sawflies, we quantified chemical similarity between species based on the similarity of chemical structure and relative abundance of compounds. We found that closely related *Inga* species and geographically separated populations of the same *Inga* species tend to have similar chemical defenses. Principal coordinates of the

TABLE 1 | Measure of phylogenetic signal for each *Inga* defensive trait and the principal coordinates of the chemistry similarity matrix (PCO) using Blomberg's *K*.

Defensive traits	K statistic	P (reps = 9999)
Chemistry PCO1 (39%)	0.71	0.0002
Chemistry PCO2 (17%)	1	0.0001
Leaf expansion rate	0.37	0.05
Chlorophyll content	0.49	0.01
Ant number	0.12	0.58
Extra-floral nectary size	0.09	0.8

For PCO components, values in parentheses represent the percentage of variation explained by each component.

chemistry similarity matrix showed phylogenetic signal (PCO1 $K = 0.71$, p = 0.0002; PCO2 $K = 1$, $p = 0.0001$, **Table 1**). For example, lineages from the *Inga capitata* species complex (**Figure 3A**, left-hand phylogeny) share a series of tyramine gallates and quinic acid gallates. Similarly, the clade containing *Inga edulis, Inga poeppigiana, Inga ruiziana* and *Inga thibaudiana* share similar chemistry based on gallocatechin/epigallocatechin gallates. However, we find examples of closely related taxa with contrasting chemistry, a typical pattern for the genus as a whole (Kursar et al., 2009). For instance, *Inga umbellifera_no_Y* in French Guiana lacks overexpression of tyrosine in expanding leaves, whereas its sister species, *I. umbellifera* from Panama, contains 10.1% of leaf dry mass as tyrosine.

Developmental defenses of *Inga* species fed on by sawflies showed a similar pattern to chemistry. Leaf expansion rate and chlorophyll content showed weak phylogenetic signal (leaf expansion rate $K = 0.37$, $p = 0.05$, chlorophyll content $K = 0.49$, $p = 0.001$). In contrast, biotic defenses were divergent among close relatives in *Inga* that are sawfly hosts, with no evidence for phylogenetic conservatism in ant visitation and extra-floral nectary size (**Table 1**).
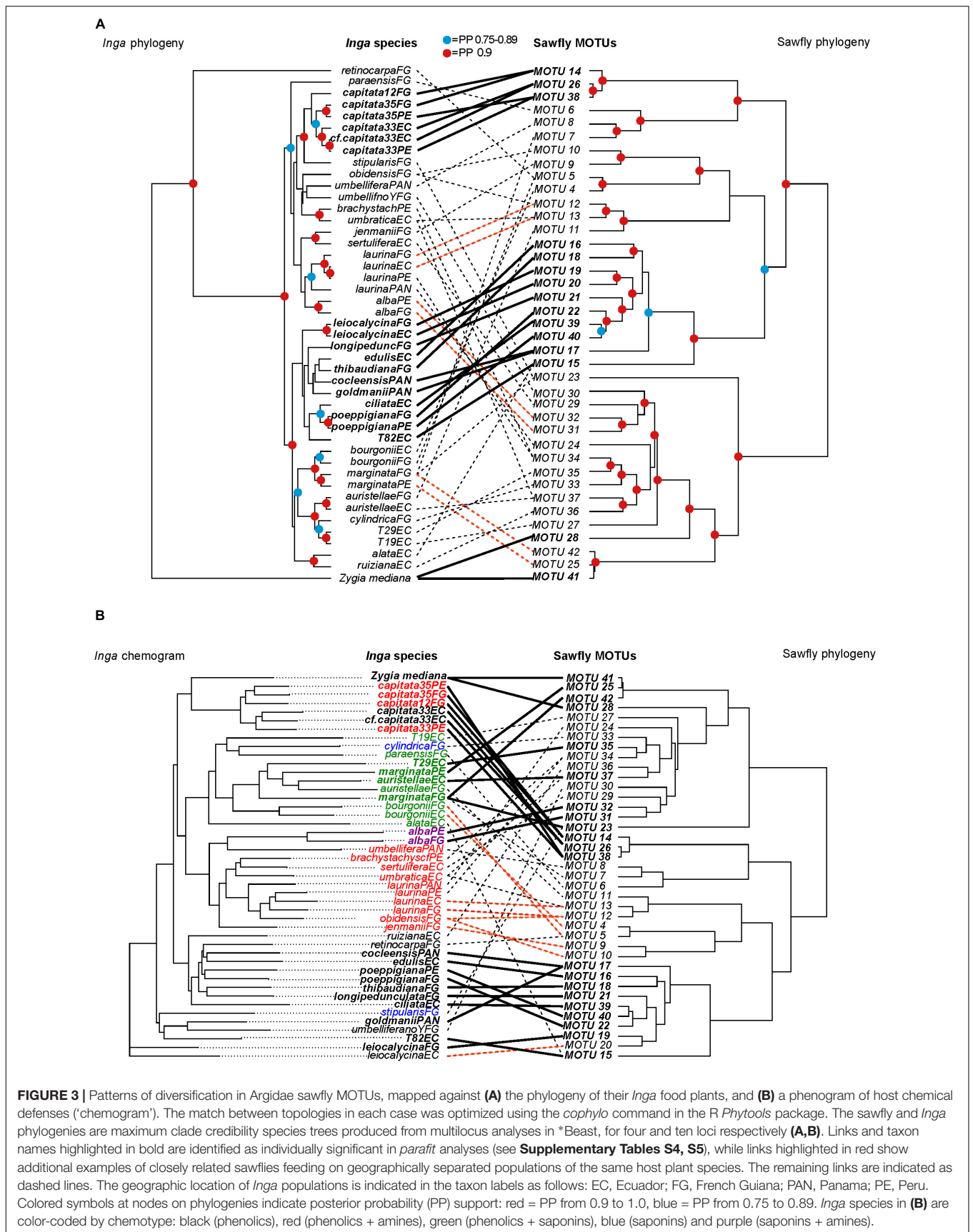
Chemically Similar *Inga* Hosts Are Attacked by Similar Sets of Sawflies

Evolutionary interactions between sawflies and their *Inga* hosts were tested using a four-locus phylogeny for Argidae sawfly MOTUs and a ten-locus phylogeny for their *Inga* food plants. Because only chemistry was selected as an important predictor for sawfly *Inga* associations, the following analyses were performed without the other host defensive traits. Phylogenies for both groups were well resolved, with strong posterior support at many nodes (**Figure 3A**).

Parafit analysis revealed a significant signature of codiversification between these two groups (global correlation, $p = 0.015$). The 19 sawfly-*Inga* interactions contributing most strongly to this pattern are concentrated in two sawfly and *Inga* clades (**Figure 3A**), and include closely related sawfly MOTUs that feed on geographically separated populations of the same species of *Inga* (see highlighted links in **Figure 3A** for sawflies feeding on *Inga alba, Inga capitata, Inga laurina, Inga leiocalycina, Inga marginata,* and *Inga poeppigiana*). However, there are also multiple examples of a single sawfly MOTU that feeds on phylogenetically divergent host plants (e.g., *Inga auristellae* and *Inga umbratica* attacked by MOTU 37 in Ecuador, *Inga stipularis* and *Inga marginata* attacked by MOTU23 in French Guiana, *Inga retinocarpaFG* and *Inga bourgoniiFG* attacked by MOTU 5), and divergent *Inga* hosts attacked by closely related sawfly MOTUs (e.g., *Inga umbraticaEC* and *Inga auristellaeFG* attacked by MOTUs 11 and 13). A single *Inga* species can also be attacked by phylogenetically divergent sawfly MOTUs (e.g., *Inga marginata*FG attacked by sawfly MOTUs 7, 42 and 23, and *Inga umbratica* attacked by MOTUs 13 and 37).

There is a much stronger correlation between the sawfly phylogeny and *Inga* chemistry (global correlation, $p = 0.001$) (**Figure 3B**). Many of the links contributing to this pattern (18 of 25 interactions, **Supplementary Table S5**) are the same as
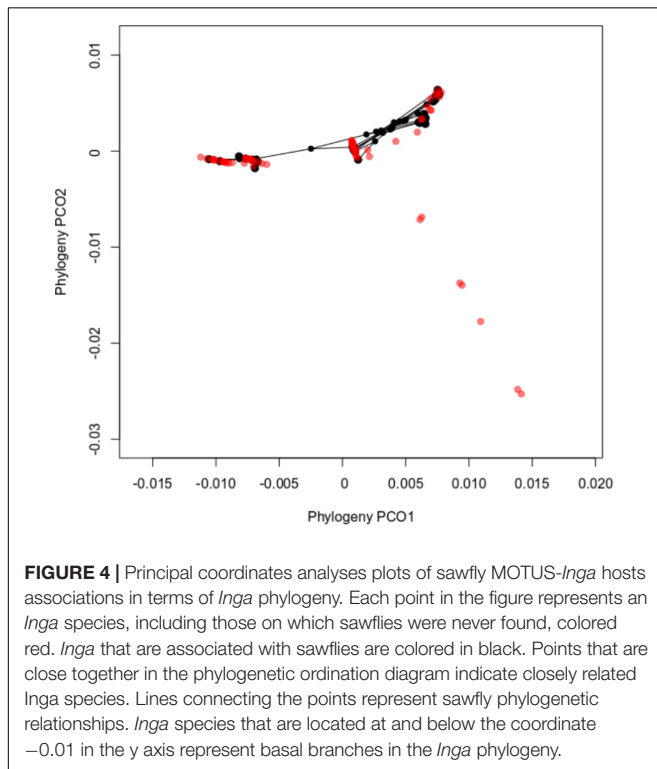
FIGURE 3 | Patterns of diversification in Argidae sawfly MOTUs, mapped against **(A)** the phylogeny of their *Inga* food plants, and **(B)** a phenogram of host chemical defenses ('chemogram'). The match between topologies in each case was optimized using the *cophylo* command in the R *Phytools* package. The sawfly and *Inga* phylogenies are maximum clade credibility species trees produced from multilocus analyses in *Beast, for four and ten loci respectively **(A,B)**. Links and taxon names highlighted in bold are identified as individually significant in *parafit* analyses (see **Supplementary Tables S4, S5**), while links highlighted in red show additional examples of closely related sawflies feeding on geographically separated populations of the same host plant species. The remaining links are indicated as dashed lines. The geographic location of *Inga* populations is indicated in the taxon labels as follows: EC, Ecuador; FG, French Guiana; PAN, Panama; PE, Peru. Colored symbols at nodes on phylogenies indicate posterior probability (PP) support: red = PP from 0.9 to 1.0, blue = PP from 0.75 to 0.89. *Inga* species in **(B)** are color-coded by chemotype: black (phenolics), red (phenolics + amines), green (phenolics + saponins), blue (saponins) and purple (saponins + amines).

TABLE 2 | Proportion of variation in sawfly incidence data attributed to phylogenetic and defensive terms.

	Including geographical region information		Without geographical region information	
	Controlling for sampling effort	Not controlling for sampling effort	Controlling for sampling effort	Not controlling for sampling effort
Geographical region	0.163 (0.000–0.591)	0.165 (0.000–0.596)	0.214 (0.000–0.783)	0.236 (0.000–0.955)
Phylogenetic main effect for sawflies	0.009 (0.000–0.042)	0.003 (0.000–0.015)	0.010 (0.000–0.04)	0.010 (0.000–0.041)
Phylogenetic main effect for *Inga* hosts	0.005 (0.000–0.021)	0.012 (0.000–0.005)	0.006 (0.000–0.029)	0.006 (0.000–0.024)
Defense main effect for *Inga* hosts	0.018 (0.000–0.074)	0.015 (0.000–0.055)	0.011 (0.000–0.449)	0.009 (0.000–0.036)
Inga hosts evolutionary interaction	0.009 (0.000–0.038)	0.007 (0.000–0.03)	0.016 (0.000–0.063)	0.020 (0.000–0.065)
Inga defense interaction	0.537 (0.091–1.009)	0.546 (0.104–0.967)	0.663 (0.188–1.238)	0.650 (0.236–1.039)
Sawfly evolutionary interaction	0.019 (0.000–0.251)	0.019 (0.000–0.071)	0.016 (0.000–0.066)	0.015 (0.000–0.057)
Coevolutionary interaction	0.065 (0.000–0.251)	0.058 (0.000–0.293)	0.015 (0.000–0.10)	0.010 (0.000–0.045)
Defense tracking interaction	0.065 (0.000–0.250)	0.051 (0.000–0.161)	0.026 (0.000–0.10)	0.020 (0.000–0.078)
Main effect for sawflies	0.005 (0.000–0.021)	0.007 (0.000–0.02)	0.005 (0.000–0.023)	0.008 (0.000–0.034)
Main effect for *Inga* hosts	0.006 (0.000–0.026)	0.007 (0.000–0.028)	0.006 (0.000–0.028)	0.004 (0.000–0.015)
Interaction effect	0.063 (0.000–0.241)	0.069 (0.000–0.264)		

Columns contain the posterior modes (with 95% confidence intervals in parentheses) for the estimates. See Materials and Methods for a description of each term.

FIGURE 4 | Principal coordinates analyses plots of sawfly MOTUS-*Inga* hosts associations in terms of *Inga* phylogeny. Each point in the figure represents an *Inga* species, including those on which sawflies were never found, colored red. *Inga* that are associated with sawflies are colored in black. Points that are close together in the phylogenetic ordination diagram indicate closely related Inga species. Lines connecting the points represent sawfly phylogenetic relationships. *Inga* species that are located at and below the coordinate −0.01 in the y axis represent basal branches in the *Inga* phylogeny.

those contributing to the correlation between the *Inga* and sawfly phylogenies. There are many examples of closely related sawfly MOTUs attacking chemically similar *Inga* taxa (**Figure 3B**). In some cases, the two chemically similar *Inga* species are not closely related phylogenetically. For example, sawfly MOTU 12 attacks both *Inga laurina* and *Inga obidensis* in French Guiana. These two host plants have similar chemistry (**Figure 3B**), but are quite divergent phylogenetically (**Figure 3A**). There are four examples of the same sawfly MOTU attacking two hosts that are very divergent chemically (sawfly MOTU 15 attacking both *Inga* T82, *Inga alata* and MOTU 37 attacking *Inga auristellae, Inga umbratica* in Ecuador; MOTU 5 attacking *Inga retinocarpa* and *Inga bourgoniii* and MOTU 23 attacking *Inga marginata* and *Inga stipularis* in French Guiana) (**Figure 3B**).

In agreement with *Parafit* analyses, our *MCMCglmm* evolutionary models incorporating phylogenetic and chemical effects showed that the defense interaction term contributed the greatest variation to the sawfly incidence data, suggesting that the association between sawflies and *Inga* hosts is mainly due to chemistry (*Inga* defense interaction term in **Table 2**). The defense interaction term is the only term whose lower confidence limits exclude zero in any model, and this is true for all four models in **Table 2**. Chemically similar *Inga* species are attacked by related sets of sawfly MOTUs, having taken sawfly phylogeny into account. This was true in models with and without between-site information and sampling effort (**Table 2**). At large spatial scales (models with between-site information), coevolutionary and defense tracking effects were moderately large indicating that closely related sawflies are feeding on closely related *Inga*, which also are similar in chemistry (**Table 2**). However, when the models were fitted without controlling for sampling size and at small spatial scales (without between-site information), both the coevolutionary effect and the defense tracking effect decreased.

Geographic region has a large effect in all models (**Table 2**). In some cases, closely related species of sawflies are separated by geography but feed on the same species of *Inga*. For example, MOTU 31 attacks *Inga alba* in Peru, and its sister species MOTU 32 is associated with *Inga alba* in French Guiana (**Figures 3A,B**). MOTU 19 is associated with *Inga leiocalycina* in French Guiana and the sister lineage, MOTU 20, is associated with *Inga leiocalycina* in Ecuador (**Figures 3A,B**). These observations are most consistent with an allopatric mode of sawfly speciation, suggesting that biogeography is an important component in sawfly *Inga* associations.

The ordination diagram of the sawfly *Inga* associations in phylogenetic space (**Figure 4**) supported these findings by clustering sawfly MOTUS associated with *Inga* hosts that are closely related. This graph also shows the level of specialization for sawflies. The portion of *Inga* phylogenetic space towards the bottom right has seven species upon which we did not find any sawflies. These belong to early-diverging lineages of *Inga*. In fact, the sampled sawfly species feed entirely on one clade of *Inga*, albeit a clade that encompasses the large majority of *Inga* species.

DISCUSSION

Sawfly Barcoding

Work on tropical plant-herbivore associations has long been hampered by lack of taxonomic resources. DNA barcoding is well established as a major tool in circumventing this taxonomic impediment in species-rich tropical ecosystems (Janzen et al., 2005; Miller et al., 2016). Our barcoding of sawfly larvae has generated host plant association data for 41 *Inga* or *Zygia*-feeding MOTUs, and represents a substantial extension to what is known for neotropical sawflies. Forty of the full set of 41 MOTUs (38 of the 39 putative Argidae MOTUs) are novel. Two putative Argidae specimens from Barro Colorado, Panama, showed a 99% match to a voucher sequence for the argid species *Ptenos leucopoda*, described from Guanacaste, Costa Rica, and are probably members of this species. Twenty-two other individuals in eight MOTUs showed ≥ 90% sequence similarity to voucher sequence for species in the Argidae genus *Ptenos*, and are also probably members of this genus.

Sawfly faunas in many tropical regions of the world remain relatively understudied, and even where adults have been sampled the larval foodplants of most species remain unknown. As an example, the genus *Ptenos*, to which some of our *Inga*-sampled sawflies certainly belong, contains around 31 species from the southwestern United States to Argentina, but to our knowledge, published food plant associations are only known for one species, *P. leucopoda* (Smith et al., 2013). Pairing of adults and larval stages is a major benefit of DNA barcoding (e.g., Stone et al., 2008) – but few voucher barcode sequences for identified adults exist for many groups of sawflies. For example, Schmidt et al. (2017) reported BOLD reference barcode sequences for only 49 of the 918 known Argidae species worldwide. Only one of our specimens showed a high match to an identified voucher, for *Ptenos leucopoda* from Costa Rica. While sequence match places the other 40 MOTUs confidently within the families Argidae

($n = 38$) and Tenthredinidae ($n = 2$), their species status remains to be determined. The sequence divergence threshold we have used, at 1.5%, is slightly lower than the 2% applied by Schmidt et al. (2017) for the same sequence region in their Europe-focused barcode study of sawflies. However, Schmidt et al. (2017) found sequences for 13 of 49 Argidae voucher taxa to differ by less than 2%, suggesting that our empirically determined lower threshold is appropriate for this group.

Inga-Sawfly Evolutionary Associations

Our results extend Ehrlich and Raven's main prediction that closely related plants are associated with closely related herbivores (Ehrlich and Raven, 1964). Colonization of *Inga* by sawflies seems to have been restricted to two events: (1) once by the ancestor of the *Inga*-associated Argidae clade, and (2) once by the common ancestor of the two *Inga*-associated Tenthredinidae MOTUs (**Figure 1**). Here we focus on Argidae. Given the high phylogenetic conservatism for chemical defenses in the species of *Inga* associated with sawflies (**Table 1**), we would predict high topological congruence between *Inga* and sawfly phylogenies. Evolutionary analysis suggested a significant congruence between both topologies (**Figure 3A**). This result is further supported by the monophyly of the argid sawflies associated with *Inga*. Most *Inga* and *Zygia*-associated sawflies belong to a single clade that can be confidently placed in the family Argidae with reference to identified reference material – including a sequence match with Costa Rican sequences for the species *Ptenos leucopoda*. It is possible, however, that the monophyly of the Argidae group of 39 MOTUs could be an artifact resulting from undersampling of alternative host plant groups in the Neotropics. Nevertheless, the fact that related sawflies have not been found on other hosts in Guanacaste, Costa Rica[6] despite many years of sampling, suggests that this sawfly clade is genuinely restricted to *Inga* and close relatives.

The genus *Inga* is thought to represent a geologically young radiation, with a common ancestor between 4 and 10 million years ago (Richardson et al., 2001). If associated sawflies have co-diversified with their *Inga* hosts, we expect the ages of the two radiations to be similar. Because there are no fossil records for the *Inga*-associated Argidae clade, we used independent estimates for beetles and butterflies in order to calibrate the *Inga*-associated sawfly phylogeny. Comparisons with fossil-calibrated phylogenies for other sawfly taxa suggests that these calibrations are broadly applicable to sawflies (Nyman et al., 2006; Malm and Nyman, 2015). Based on these data, the estimate we obtained for the age of the common ancestor of the *Inga*-associated Argidae clade suggests that this group diversified at broadly the same time, or more recently, than their plant hosts [mean of 6.27 (between 4.78 and 7.93) million years ago using the Brower (1994) estimate and a mean of 5.31 (between 4.05 and 6.72) million years ago using the Papadopoulou et al. (2010) estimate]. Given the uncertainty in the date of the *Inga* radiation, these results are consistent with *Inga*-associated sawflies having diversified alongside their hosts, a conservative pattern of host plant use also found in other sawfly clades (Nyman et al., 2010;

---

[6]janzen.sas.upenn.edu/index.html

Schmidt and Walter, 2014) and in leaf-feeding beetles, seed predators and many other insect herbivore groups (Farrel and Mitter, 1998; Janz and Nylin, 1998; Winkler and Mitter, 2008; Edger et al., 2015). Alternatively, the radiation of Argidae might be younger than *Inga*, a pattern consistent with host-resource tracking or ecological fitting.

Ehrlich and Raven (1964) hypothesized that any taxonomic correspondence between plants and herbivores was the result of herbivore tracking of phylogenetically conserved host plant traits. Several lines of evidence suggest that defensive chemistry plays the key role in structuring sawfly associations with *Inga*. First, among all host traits, chemistry was identified as the most important predictor in sawfly *Inga* associations, with sawflies preferring *Inga* hosts that express amines (**Figure 2**). Second, after controlling for phylogenetic effects, we find that host associations in sawflies are more strongly correlated with *Inga* chemistry than *Inga* phylogeny (**Table 2** and **Figures 3A,B**). The significant concordance between the topologies of *Inga* and sawfly phylogenies could thus be explained as the result of phylogenetic conservatism in *Inga* chemistry for the set of species attacked by sawflies. Chemistry is better able to explain *Inga*-sawfly associations than the *Inga* phylogeny alone because some sawfly sister taxa are associated with hosts that are chemically similar but not closely related (**Figures 3A,B**), while there are very few cases of sawfly sister MOTUs associated with chemically divergent hosts.

Phylogenetic concordance between plants and herbivores could represent either a signature of codiversification or a radiation onto existing *Inga* (delayed resource tracking). The facts that host-shifting in sawflies is more strongly determined by *Inga* defenses than by *Inga* phylogeny (**Table 2** and **Figures 3A,B**), and that most examples of shifts between *Inga* hosts include species that are similar in defensive chemistry, regardless of relatedness (**Figures 3A,B**), support delayed host tracking. Nevertheless, it is striking that none of the more basal species in the *Inga* phylogeny are attacked by sawflies (**Figure 4**). This strongly implies cospeciation, that the ancestors of both the argid and tenthredinid sawflies now associated with *Inga* colonized, and then codiversified alongside an already ongoing radiation of *Inga*. In the end, which hypothesis is correct depends on the relative ages of the *Inga* and sawfly radiations. Our best estimate of the age of the common ancestor of the *Inga*-associated Argidae is fairly constrained (4.02–7.93 million years). In contrast, our estimate for the age of the common ancestor of *Inga* ranges from 4 to 10 million years, with the further caveat that the more derived *Inga* that are sawfly hosts are younger by an unknown extent. While the dates used here are consistent with codiversification, delayed resource tracking cannot be ruled out until the dates of origin for both crown groups, particularly *Inga*, are known with more certainty.

Although the significance of defensive traits in plant-herbivore diversification has been recognized (Futuyma and Agrawal, 2009), it is often not included in coevolutionary studies. Most studies compare the congruence between the ages and topologies of insect and host-plant phylogenies with the expectation that closely related hosts will share closely related herbivores (reviewed in Suchan and Alvarez, 2015). Alternative hypotheses,

such as tracking of host defenses, cannot be tested. We argue that in order to understand the process and factors that influence the evolution of herbivore host ranges, characterization of relevant host traits is essential.

Inga-Sawfly Patterns of Diversification
Previous work suggests that modes of speciation vary among sawfly lineages with different life history strategies. Analyses of temperate nematine sawflies suggest that lineages with externally feeding larvae tend to feed on multiple host plant species (Nyman et al., 2006, 2010), and, as a result are more likely to diversify through allopatric speciation than via host shifts. In contrast, gall-inducing sawfly lineages, which are more intimately associated metabolically with their hosts, are both more likely to feed on a narrow host range and to diversify by shifts among willow host species (Nyman et al., 2006).

Although they are external feeders, the narrow host ranges observed for *Inga*- and *Zygia*- feeding sawflies (1-2 hosts per MOTU) more closely match patterns seen in specialist gall-inducing sawflies than the wider host associations seen in externally feeding sawflies on willow. This high host specificity could result from constraints or adaptations related to host use, such as host-finding capabilities, avoidance of larval predators, and avoidance or sequestration of host toxins (Brooks and McLennan, 2002). For the sawflies associated with *Inga*, a control choice experiment in a previous study suggested that host preference is primarily driven by leaf secondary metabolites and possibly nutrition (Endara et al., 2015). Although much of the available literature concerns the superfamily Tenthredinidae in the northern hemisphere, and the families Pergidae and Argidae in Australia, many sawflies show adaptations for dealing with, and using the host plant chemistry. Many can sequester and modify toxic host compounds for use in their own anti-predator defense [e.g., Diprionidae (Eisner et al., 1974); Tenthredinidae (Boevé et al., 2013); Argidae (Petre et al., 2007)], particularly against ants (Boevé and Schaffner, 2003; Petre et al., 2007; Boevé et al., 2013). This is particularly relevant in *Inga,* many species of which recruit ant guards through secretion of extrafloral nectar. The lack of any significant association between the presence of ants and sawflies on *Inga* suggests that sawflies may not be highly sensitive to ants that provide some defense against other herbivores (Endara et al., 2017). In *Inga,* we observed that when contacted by ants, sawfly larvae raised their abdomen, and ants generally retreated immediately (MJ Endara, personal observation). In addition, most of the sawfly MOTUs found on *Inga* are gregarious, a characteristic often considered a sign of chemical defense (Boevé et al., 2013). Thus, sawflies associated with *Inga* may have an intimate relationship with their host chemistry.

Although the specialized relationship between sawflies and *Inga* would suggest a mode of speciation similar to the specialist, gall-inducing sawflies, our phylogenetic analysis reveals that the predominant mode of speciation is allopatric, similar to external sawfly feeders on willow (Nyman et al., 2010). Results from the evolutionary analysis that included phylogenetic and chemical effects show that the coevolutionary effect best explained variation in sawfly incidence when between-region information

was included (**Table 2**). This suggests that pairs of sister *Inga* host populations and sawfly MOTUs occur in non-overlapping geographic regions (Hadfield et al., 2014). This pattern can be seen throughout the whole sawfly phylogeny, with more than 60% of lineage splits potentially caused by non-ecological factors in allopatry. For example, MOTU 31 attacks *Inga alba* in Peru, and its sister species MOTU 32 is associated with *Inga alba* in French Guiana (**Figures 3A,B**). This is evidence for allopatric speciation between sawfly sister taxa associated with the same *Inga* host (Barraclough and Vogler, 2000). Thus, *Inga*-feeding sawflies could have diverged and speciated in allopatry either directly because of *Inga* speciation or because the same ecological and geographical factors that facilitated *Inga* speciation could have facilitated the speciation of its sawfly herbivores. Alternatively, although species accumulation curves show that further sampling would not yield many additional *Inga*-sawfly interactions, we may have missed collecting sister sawfly species at the same site, meaning that speciation in sympatry cannot be totally ruled out.

The finding that the speciation process in the *Inga*-sawflies is largely non-ecological in allopatry does not exclude the possibility that some diversification events may have an ecological basis (i.e., host shifts). Along the phylogeny, four instances of lineage splits can potentially be ecologically based, with two host shifts to novel hosts in sympatry (MOTU 36 is associated with *Inga ruiziana* in Ecuador which produces phenolics, whereas the sister species MOTU 37 is associated with *Inga auristellae* which produces saponins, **Figures 3A,B**) and in allopatry (MOTU 7 is associated with *Inga marginata* in French Guiana which produces saponins, and the sister lineage MOTU 8 attacks *Inga umbellifera* in Panama which produces amines, **Figures 3A,B**). The other two host shifts simply involved range expansion (i.e., switch to a different host but with a similar chemistry), with one example in sympatry (in French Guiana, MOTU 7 is associated with *Inga obidensis* and MOTU 8 is attacking *Inga jenmanii*, both hosts produce amines, **Figures 3A,B**) and the other in allopatry (MOTU 16 is associated with *Inga edulis* in Ecuador and MOTU 18 is associated with *Inga thibaudiana* in French Guiana, with both hosts producing phenolics, **Figures 3A,B**). Excluding few exceptions, none of these switches involved phylogenetically closely related hosts, but rather chemically similar ones (**Figures 3A,B**), highlighting the importance of plant chemistry in ecological speciation.

CONCLUSION

Our phylogeny- and trait-based analysis of the interactions between *Inga* and Argidae sawflies indicates the importance of including ecologically relevant traits for host selection in studies of herbivore-host plant coevolution. For example, closely related sawfly species often shift to *Inga* that are similar chemically but not closely related phylogenetically. Our results suggest a major role for host chemistry in explaining both the observed concordance between *Inga* and sawfly phylogenies, and in explaining the deviations from this pattern resulting from evolutionary tracking of defensive traits by sawflies.

Our analyses suggest two modes of diversification of sawflies: (i) allopatric divergence between sawfly sister taxa associated with the same *Inga* food plant and (ii) niche shifts. The vast majority of lineage splits in these sawflies seem to have occurred non-ecologically in allopatry, a pattern that may well be true for other groups of insect herbivores (Nyman et al., 2010). Thus, sawflies primarily speciate allopatrically, but descendent species are constrained to use the same host species or others with similar chemistry. Closely related sawflies very rarely attack chemically dissimilar *Inga* species, implying that, for the most part, these herbivores have not experienced the niche shifts thought to promote diversification in other insect herbivores, and particularly in highly specialized taxa (Rundle and Nosil, 2005; Dyer et al., 2007; Futuyma and Agrawal, 2009).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01237/full#supplementary-material

Sawfly sequence data can be found in GenBank, accessions MH206848 – MH207017 for COI, MH206768 – MH206847 for ITS2, MH206617 – MH206691 for *wg* and MH206692 – MH206767 for *PGD*.

FIGURE S1 | MrBayes majority-rule consensus tree for the mitochondrial *COI* DNA barcode fragment. Numbers above nodes indicate posterior probabilities. Taxon label colors indicate membership of 1.5% sequence divergence jMOTU taxa, indicated by the labels at right.

FIGURE S2 | Results of MOTU identification analyses of *Inga*- and *Zygia*-feeding sawflies, using a 645 bp fragment of the mitochondrial *COI* DNA barcoding region for **(a)** jMOTU and **(b)** ABGD.

FIGURE S3 | MrBayes majority-rule consensus tree for the nuclear locus *ITS2,* sequenced for exemplars of each of the selected 41 jMOTU 1.5% *COI* MOTUs. Numbers above nodes indicate posterior probabilities. Taxon labels are colored to indicate membership of different MOTUs.

FIGURE S4 | MrBayes majority-rule consensus tree for the nuclear locus *wingless,* sequenced for exemplars of each of the selected 41 jMOTU 1.5% *COI* MOTUs. Numbers above nodes indicate posterior probabilities. Taxon labels are colored to indicate membership of different MOTUs.

FIGURE S5 | Sawfly MOTU accumulation curves when sampling over *Inga* host plant taxa, and when sampling over individuals. For each curve, the mean estimate is shown as a dark blue line and the standard deviation as a pale blue shaded region either side. The total numbers of *Inga* taxa and sawfly specimens in these analyses were 34 and 1286, respectively.

FIGURE S6 | Phylogenetic relationships for the gene *CO1* among the *Inga*-feeding sawfly MOTUs and a panel of voucher sequences for sawflies in the families Argidae, Pergidae (sister group to Argidae; Malm and Nyman, 2015) and Tenthredinidae. The tree shown is a majority-rule consensus tree constructed in MrBayes, using substitutions modeled as GTR+I+G for each of 1st and 2nd codon positions, and GTR+G for 3rd positions. We used a relaxed clock, with a birth-death speciation model. Numbers at nodes indicate posterior probability.

Taxon labels are colored by sampling source: red MOTU numbers are larvae found feeding on *Inga* or *Zygia*, while other colors indicate reference sequences for adult Argidae, Pergidae and Tenthredinidae. The taxon label MOTU17_BCI marked with two asterisks is a voucher sequence for a specimen of *Ptenos leucoopoda* (Argidae) sampled from *Inga oerstediana* (and also recorded from *I. vera*) in Costa Rica (Smith et al., 2013).

TABLE S1 | List of compounds putatively identified through matches to reference MSMS spectra on the Global Natural Products Social Molecular Networking database (https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp). The cosine score is a measure of the similarity of MS/MS-derived fragments between two compounds.

TABLE S2 | Metadata for all sawfly specimens collected in this study, including host plant and collection location, MOTU allocation (1.5% jMOTU taxa), and Genbank accession numbers for all sequenced gene fragments. Note that in our sampling system, each study site has independent collection numbers. Thus, it is possible for two *Inga* plants to have the same host plant number, but only because they were sampled at different sites.

TABLE S3 | Metadata for additional reference sawfly sequences, with species name, country of origin, Genbank accession numbers for *COI* and *PGD* gene fragments, and source reference.

TABLE S4 | Information on the ten sequence loci used for construction of the *Inga* species tree. Locus number, reference transcript, functional annotation and the substitution model used in phylogenetic analyses all refer to Nicholls et al. (2015).

TABLE S5 | (A) *Parafit* analysis output for sawfly and *Inga* phylogenies, for sawfly MOTUs in the family Argidae. **(B)** *Parafit* analysis of concordance between sawfly phylogeny and Inga chemogram. In **(A)** and **(B)** herbivore-*Inga* associations that are identified as individually significant are highlighted in yellow.

APPENDIX SI | Molecular methods for PCR amplification of sawfly sequences.

APPENDIX SII | Detailed chemical methods for construction of a chemical similarity matrix.

REFERENCES

Acs, Z., Challis, R., Bihari, P., Blaxter, M., Hayward, A., Melika, G., et al. (2010). Phylogeny and DNA barcoding of inquiline oak gallwasps (Hymenoptera: Cynipidae) of the Western Palaearctic. *Mol. Phylogenet. Evol.* 55, 210–225. doi: 10.1016/j.ympev.2009.12.004

Agosta, S. J., and Klemens, J. A. (2008). Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecol. Lett.* 11, 1123–1134. doi: 10.1111/j.1461-0248.2008.01237.x

Barraclough, T. G., and Vogler, A. P. (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* 155, 419–434. doi: 10.1086/303332

Becerra, J. X. (1997). Insects on plants: macroevolutionary chemical trends in host use. *Science* 276, 253–256. doi: 10.1126/science.276.5310.253

Becerra, J. X., Noge, K., and Venable, D. L. (2009). Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18062–18066. doi: 10.1073/pnas.0904456106

Benton, P., Want, E. J., and Ebbels, T. M. D. (2010). Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics* 26, 2488–2489. doi: 10.1093/bioinformatics/btq441

Bixenmann, R. J., Coley, P. D., Weinhold, A., and Kursar, T. A. (2016). High herbivore pressure favors constitutive over induced defense. *Ecol. Evol.* 6, 6037–6049. doi: 10.1002/ece3.2208

Blomberg, S. P., Garland, T., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

Boevé, J.-L., and Schaffner, U. (2003). Why does the larval integument of some sawfly species disrupt so easily? The harmful hemolymph hypothesis. *Oecologia* 134, 104–111. doi: 10.1007/s00442-002-1092-4

Boevé, J.-L., Blank, S. M., Meijer, G., and Nyman, T. (2013). Invertebrate and avian predators as drivers of chemical defensive strategies in tenthredinid sawflies. *BMC Evol. Biol.* 13:198. doi: 10.1186/1471-2148-13-198

Bolker, B. (2016). *Emdbook: Ecological Models and Data (Book Support).* Available at: http://cran.r-project.org/web/packages/emdbook/emdbook.pdf

Bolker, B. (2017). *Bbmle: Tools for General Maximum Likelihood Estimation.* Available at: http://cran.r-project.org/web/packages/bbmle/bbmle.pdf

Brooks, D. R., and McLennan, D. A. (2002). *The Nature of Diversity: An Evolutionary Voyage of Discovery.* Chicago, IL: The University of Chicago Press. doi: 10.7208/chicago/9780226922478.001.0001

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6491–6495. doi: 10.1073/pnas.91.14.6491

Calatayud, J., Horreo, J. L., Madrigal-Gonzalez, J., Migeon, A., Rodriguez, M. A., Magalhaes, S., et al. (2016). Geography and major host evolutionary transitions shape the resource use of plant parasites. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9840–9845. doi: 10.1073/pnas.1608381113

Chambers, M. C., MacLean, B., Burke, B., Amode, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotech.* 30, 918–920. doi: 10.1038/nbt.2377

Coley, P. D., Endara, M. J., and Kursar, T. A. (2018). Consequences of interspecific variation in defenses and herbivore host choice for the ecology and evolution of Inga, a speciose rainforest tree. *Oecologia* 187, 361–376. doi: 10.1007/s00442-018-4080-z

Cruaud, A., Ronsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., et al. (2012). An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Syst. Biol.* 61, 1029–1047. doi: 10.1093/sysbio/sys068

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Dyer, L. A., Singer, M. S., Lill, J. T., Stireman, J. O., Gentry, R. J., Marquis, R. G., et al. (2007). Host specificity of Lepidoptera in tropical and temperate forests. *Nature* 448, 696–699. doi: 10.1038/nature05884

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants: a study in coevolution. *Evolution* 18, 586–608. doi: 10.1111/j.1558-5646.1964.tb01674.x

Eisner, T., Johnessee, J. S., Carrel, J., and Meinwald, J. (1974). Defensive use by an insect of a plant resin. *Science* 184, 996–999. doi: 10.1126/science.184.4140.996

Endara, M. J., Coley, P. D., Ghabash, G., Nicholls, J. A., Dexter, K. G., Stone, G., et al. (2017). Herbivores and plants: coevolutionary arms race or evolutionary defense chase? *Proc. Natl. Acad. Sci. U.S.A.* 114, E7499–E7505. doi: 10.1073/pnas.1707727114

Endara, M. J., Weinhold, A., Cox, J. E., Wiggins, N. L., Coley, P. D., and Kursar, T. A. (2015). Divergent evolution in antiherbivore defenses within species complexes at a single Amazonian site. *J. Ecol.* 103, 107–111. doi: 10.1111/1365-2745.12431

Farrel, B. D., and Mitter, C. (1998). The timing of insect/plant diversification: might Tetraopes (Coleoptera: Cerambycidae) and Asclepias (Asclepiadaceae) have co-evolved? *Biol. J. Linn. Soc.* 63, 553–577. doi: 10.1111/j.1095-8312.1998.tb00329.x

Fine, P. V. A., Metz, M. R., Lokvam, J., Mesones, I., Zuniga, J. M. A., Lamarre, G. P. A., et al. (2013). Insect herbivores, chemical innovation, and the evolution of habitat specialization in Amazonian trees. *Ecology* 94, 1764–1775. doi: 10.1890/12-1920.1

Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K., Widmayer, H. A., et al. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* 71, 1126–1137. doi: 10.1111/evo.13164

Funk, D. J., and Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Ann. Rev. Ecol. Syst.* 34, 397–423. doi: 10.1016/j.ympev.2009.08.024

Futuyma, D. J., and Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18054–18061. doi: 10.1073/pnas.0904106106

Guindon, S., and Gascuel, O. (2003). A simple, fast and accurate method to estimate large phylo- genies by maximum-likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hadfield, J. D. (2017). *MCMCglmm: MCMC Generalized Linear Mixed Models*. Available at: https://cran.r-project.org/web/packages/MCMCglmm/MCMCglmm.pdf.

Hadfield, J. D., Krasnov, B. R., Poulin, R., and Nakagawa, S. (2014). A tale of two phylogenies: comparative analyses of ecological interactions. *Am. Nat.* 183, 174–187. doi: 10.1086/674445

Hadfield, J. D., and Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.* 23, 494–508. doi: 10.1111/j.1420-9101.2009.01915.x

Hardy, N. B., and Otto, S. P. (2014). Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proc. Biol. Sci..* 281:20132960. doi: 10.1098/rspb.2013.2960

Hartsough, C. D., Connor, E. F., Smith, D. R., and Spicer, G. S. (2007). Systematics of two feeding morphs of *Schizocerella pilicornis* (Hymenoptera: Argidae) and recognition of two species. *Ann. Entomol. Soc. Am.* 100, 375–380. doi: 10.1603/0013-8746(2007)100[375:SOTFMO]2.0.CO;2

Heled, J., and Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27, 570–580. doi: 10.1093/molbev/msp274

Hembry, D. H., Yoder, J. B., and Goodman, K. R. (2014). Coevolution and the diversification of life. *Am. Nat.* 184, 425–438. doi: 10.1086/677928

Hoberg, E. P., and Brooks, D. R. (2008). A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. *J. Biogeogr.* 35, 1533–1550. doi: 10.1111/j.1365-2699.2008.01951.x

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R., et al. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318, 1913–1916. doi: 10.1126/science.1146954

Janz, N. (2011). Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. *Annu. Rev. Ecol. Evol. Syst.* 42, 71–89. doi: 10.1146/annurev-ecolsys-102710-145024

Janz, N., and Nylin, S. (1998). Butterflies and plants: a phylogenetic study. *Evolution* 52, 486–502. doi: 10.2307/2411084

Janzen, D. H., Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., and Hebert, P. D. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1835–1845. doi: 10.1098/rstb.2005.1715

Jones, M., Ghoorah, A., and Blaxter, M. (2011). jMOTU and taxonator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS One* 6:e19259. doi: 10.1371/journal.pone.0019259

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R., and Neumann, S. (2012). CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* 84, 283–289. doi: 10.1021/ac202450g

Kursar, T. A., and Coley, P. D. (1992). The consequences of delayed greening during leaf development for light absorption and light use efficiency. *Plant Cell Environ.* 15, 901–909. doi: 10.1111/j.1365-3040.1992.tb01022.x

Kursar, T. A., Dexter, K. G., Lokvam, J., Pennington, R. T., Richardson, J. E., Weber, M., et al. (2009). The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus Inga. *Proc. Natl. Acad. Sci. U.S.A..* 106, 18073–11807. doi: 10.1073/pnas.0904786106

Legendre, P., Desdevises, Y., and Bazin, E. (2002). A statistical test for host–parasite coevolution. *Syst. Biol.* 51, 217–234. doi: 10.1080/10635150252899734

Lokvam, J., Brenes-Arguedas, T., Lee, J. S., Coley, P. D., and Kursar, T. A. (2006). Allelochemic function for a primary metabolite: the case of L-tyrosine hyper-production in Inga umbellifera (Fabaceae). *Am. J. Bot.* 93, 1109–1115. doi: 10.3732/ajb.93.8.1109

López-Carretero, A., del-Val, E., and Boege, K. (2018). "Plant-Herbivore Networks in the Tropics," in *Ecological Networks in the Tropics: An Integrative Overview of Species Interactions from Some of the Most Species-Rich Habitats on Earth*, eds W. Dáttilo and V. Rico-Gray (New York, NY: Springer International), 73–91. doi: 10.1007/978-3-319-68228-0

Malm, T., and Nyman, T. (2015). Phylogeny of the symphytan grade of Hymenoptera: new pieces into the old jigsaw(fly) puzzle. *Cladistics* 31, 1–17. doi: 10.1111/cla.12069

Marquis, R. J., Salazar, D., Baer, C., Reinhardt, J., Priest, G., and Barnett, K. (2016). Ode to Ehrlich and Raven or how herbivorous insects might drive plant speciation. *Ecology* 97, 2939–2951. doi: 10.1002/ecy.1534

Miller, S. E., Hausmann, A., Hallwachs, W., and Janzen, D. H. (2016). Advancing taxonomy and bioinventories with DNA barcodes. *Philos. Transl. R. Soc. Lond. B Biol. Sci.* 371:20150339. doi: 10.1098/rstb.2015.0339

Nakadai, R. (2017). Species diversity of herbivorous insects: a brief review to bridge the gap between theories focusing on the generation and maintenance of diversity. *Ecol. Res.* 32, 811–819. doi: 10.1007/s11284-017-1500-1

Naya, M., Avila-Núñez, J. L., and Calcagno-Pissarelli, M. P. (2016). Haemolymph defense capacity of the Neotropical sawfly Aneugmenus merida against ant predation. *J. Insect. Behav.* 29, 459–472. doi: 10.1007/s10905-016-9573-1

Nicholls, J. A., Challis, R. J., Mutun, S., and Stone, G. N. (2012). Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridising oak gallwasps. *Mol. Ecol.* 21, 4051–4062. doi: 10.1111/j.1365-294X.2012.05683.x

Nicholls, J. A., Pennington, R. T., Jozef, E., Koenen, M., Hughes, C. E., Hearn, J., et al. (2015). Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus Inga (Leguminosae: Mimosoideae). *Front. Plant Sci.* 6:710. doi: 10.3389/fpls.2015.00710

Nylander, J. A. A. (2004). *MrModeltest v2. Program Distributed by the Author.* Uppsala: Uppsala University.

Nyman, T. (2010). To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant feeding insects. *Biol. Rev.* 85, 393–411. doi: 10.1111/j.1469-185X.2009.00109.x

Nyman, T., Farrell, B. D., Zinovjev, A. G., and Vikberg, V. (2006). Larval habits, host-plant associations, and speciation in nematine sawflies (Hymenoptera: Tenthredinidae). *Evolution* 60, 1622–1637. doi: 10.1111/j.0014-3820.2006.tb00507.x

Nyman, T., Vikberg, V., Smith, D. R., and Boevé, J.-L. (2010). How common is ecological speciation in plant-feeding insects? A 'Higher' Nematinae perspective. *BMC Evol. Biol.* 10:266. doi: 10.1186/1471-2148-10-266

Papadopoulou, A., Anastasiou, I., and Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* 27, 1659–1672. doi: 10.1093/molbev/msq051

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

Petre, C. A., Detrain, C., and Boevé, J. L. (2007). Anti-predator defence mechanisms in sawfly larvae of Arge (Hymenoptera, Argidae). *J. Insect Physiol.* 53, 668–675. doi: 10.1016/j.jinsphys.2007.04.007

Prous, M., Heidemaa, M., and Soon, V. (2011). *Empria longicornis* species group: taxonomic revision with notes on phylogeny and ecology (Hymenoptera, Tenthredinidae). *Zootaxa* 2756, 1–39.

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* 21, 1864–1877. doi: 10.1111/j.1365-294X.2011.05239.x

Rambaut, A., and Drummond, A. J. (2007). *Tracer v1.4.* Available at: http://beast.bio.ed.ac.uk/Tracer

Revell, L. J. (2017). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Richardson, J. E., Pennington, R. T., Pennington, T. D., and Hollingsworth, P. M. (2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293, 2242–2245. doi: 10.1126/science.1061421

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rundle, H. D., and Nosil, P. (2005). Ecological Speciation. *Ecol. Lett.* 8, 336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Russo, L., Miller, A. D., Tooker, J., Bjornstad, O. N., and Shea, K. (2017). Quantitative evolutionary patterns in bipartite networks: vicariance, phylogenetic tracking or diffuse co-evolution? *Methods Ecol. Evol.* 9, 761–772. doi: 10.1111/2041-210X.12914

Salazar, D., Jaramillo, M. A., and Marquis, R. J. (2016). Chemical similarity and local community assembly in the species rich tropical genus Piper. *Ecology* 97, 3176–3183. doi: 10.1002/ecy.1530

Schmidt, S., Taeger, A., Moriniere, J., Liston, A., Blank, S. M., Kramp, K., et al. (2017). Identification of sawflies and horntails (Hymenoptera, 'Symphyta') through DNA barcodes: successes and caveats. *Mol. Ecol. Res.* 17, 670–685. doi: 10.1111/1755-0998.12614

Schmidt, S., and Walter, G. H. (2014). Young clades in an old family: major evolutionary transitions and diversification of the eucalypt-feeding pergid sawflies in Australia (Insecta, Hymenoptera, Pergidae). *Mol. Phyl. Evol.* 74, 111–121. doi: 10.1016/j.ympev.2014.02.002

Schulmeister, S., Wheeler, W. C., and Carpenter, J. M. (2002). Simultaneous analysis of the basal lineages of Hymenoptera (Insecta) using sensitivity analysis. *Cladistics* 18, 455–484. doi: 10.1111/j.1096-0031.2002.tb00287.x

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787. doi: 10.1021/ac051437y

Smith, D. R., Janzen, D. H., and Hallwachs, W. (2013). Food plants and life histories of sawflies of the families Argidae and Tenthredinidae (Hymenoptera) in Costa Rica, a supplement. *J. Hymenopt. Res.* 35, 17–31. doi: 10.3897/jhr.35.5496

Stone, G. N., Atkinson, R. J., Rokas, A., Nieves-Aldrey, J.-L., Melika, G., Ács, Z., et al. (2008). Evidence for widespread cryptic sexual generations in apparently asexual *Andricus* gallwasps. *Mol. Ecol.* 17, 652–665. doi: 10.1111/j.1365-294X.2007.03573.x

Suchan, T., and Alvarez, N. (2015). Fifty years after Ehrlich and Raven, is there support for plant-insect coevolution as a major driver of species diversification? *Entomol. Exp. Appl.* 157, 98–112. doi: 10.1111/eea.12348

Suzuki, R., and Shimodaira, H. (2014). pvclust: hierarchical clustering with P-values via multiscale bootstrap resampling. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117

Tautenhahn, R., B€ottcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9:504. doi: 10.1186/1471-2105-9-504

Thompson, J. N. (1988). Coevolution and alternative hypothesis on insect/plant interactions. *Ecology* 69, 893–895. doi: 10.2307/1941238

Volf, M., Segar, S. T., Miller, S. E., Isua, B., Sisol, M., Aubona, G., et al. (2018). Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in Ficus. *Ecol. Lett.* 21, 83–92. doi: 10.1111/ele.12875

Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., et al. (2012). Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1743–E1752. doi: 10.1073/pnas.1203689109

Wheat, C. W., Vogel, H., Wittstock, U., Braby, M. F., Underwood, D., and Mitchell-Olds, T. (2007). The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20427–20431. doi: 10.1073/pnas.0706229104

Wiggins, N. L., Forrister, D. L., Endara, M. J., Coley, P. D., and Kursar, T. A. (2016). Quantitative and qualitative shifts in defensive metabolites define chemical defense investment during leaf development in Inga, a genus of tropical trees. *Ecol. Evol.* 6, 478–492. doi: 10.1002/ece3.1896

Wilson, J. S., Forister, M. L., Dyer, L. A., O'Connor, J. M., Burls, K., Feldman, C. R., et al. (2012). Host conservatism, host shifts and diversification across three trophic levels in two neotropical forests. *J. Evol. Biol.* 25, 532–546. doi: 10.1111/j.1420-9101.2011.02446.x

Winkler, I. S., and Mitter, C. (2008). "The phylogenetic dimension of insect/plant interactions: a summary of recent evidence," in *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, ed. K. Tillmon (Berkeley, CL: University of California Press), 240–263.

Züst, T., Heichinger, C., Grossniklaus, U., Harrington, R., and Turnbull, D. J. (2012). Natural enemies drive geographic variation in plant defenses. *Science* 338, 116–119. doi: 10.1126/science.1226397