

1 **The role of plant secondary metabolites in shaping regional and local plant community**
2 **assembly:** María-José Endara^{a,b}, Abrianna J. Soule^a, Dale L. Forrister^a, Kyle G. Dexter^c, R. Toby
3 Pennington^d, James A. Nicholls^e, Oriane Loiseau^c, Thomas A. Kursar^a, Phyllis D. Coley^a

4

5 ^aDepartment of Biology, University of Utah, Aline W. Skaggs Biology Building, 257 S 1400 E,
6 Salt Lake City, UT 84112-0840, USA.

7 ^bCentro de Investigación de la Biodiversidad y Cambio Climático (BioCamb) e Ingeniería en
8 Biodiversidad y Recursos Genéticos, Facultad de Ciencias de Medio Ambiente, Universidad
9 Tecnológica Indoamérica, Quito, Ecuador.

10 ^cSchool of Geosciences, University of Edinburgh, Old College, South Bridge, Edinburgh EH8
11 9YL, UK.

12 ^dDepartment of Geography, University of Exeter, Laver Building, North Park Road, Exeter, EX4
13 4QE, UK and Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR.

14 ^eThe Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australian
15 National Insect Collection (ANIC), Building 101, Clunies Ross Street, Black Mountain ACT 2601,
16 AU.

17 **Corresponding Author:**

18 María-José Endara

19 Address: Universidad Tecnológica Indoamérica, Centro de Investigación de la Biodiversidad y
20 Cambio Climático (BioCamb), Machala y Sabanilla, Quito, Ecuador

21 Email: majo.endara@utah.edu

22 **Acknowledgments**

23 We thank the Ministry of Environment and Water of Ecuador and the Ministry of Agriculture of
24 Peru for granting the research and exportation permits. Valuable field assistance was provided by
25 Wilder Hidalgo, Zachary Benavidez, Allison Thompson, Yamara Serrano and Mayra Ninazunta.
26 This work was supported by grants from the National Science Foundation (DEB-0640630 and
27 DIMENSIONS of Biodiversity DEB-1135733), and Nouragues Travel Grants Program, CNRS,
28 France, to P.D.C., and the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e
29 Innovación del Ecuador (SENESCYT) to M-J.E.

30

31 **Author contributions**

32 M-J.E., D.L.F., and P.D.C designed and conducted the research. M-J.E. designed and performed
33 the data analysis. D.L.F. and A.J.S. contributed to the metabolomic analysis. J.A.N., R.T.P.,
34 K.G.D., and O.L. contributed the phylogeny of *Inga*. M-J.E., D.L.F., A.J.S. and P.D.C. wrote the
35 manuscript, with input from K.G.D., O.L., and R.T.P.

36

37 **Data availability**

38 Chemical data and scripts to estimate chemical similarity are deposited in a github repository
39 (Forrister & Soule, 2020; https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry).

40 **Abstract**

41 **1.** The outstanding diversity of Amazonian forests is predicted to be the result of several processes.
42 While tree lineages have dispersed repeatedly across the Amazon, interactions between plants and
43 insects may be the principal mechanism structuring the communities at local scales.

44 **2.** Using metabolomic and phylogenetic approaches, we investigated the patterns of historical
45 assembly of plant communities across the Amazon based on the Neotropical genus of trees *Inga*
46 (Leguminosae) at four, widely separated sites.

47 **3.** Our results show a low degree of phylogenetic structure and a mixing of chemotypes across the
48 whole Amazon basin, suggesting that although biogeography may play a role, the metacommunity
49 for any local community in the Amazon is the entire basin. Yet, local communities are assembled
50 by ecological processes, with the suite of *Inga* at a given site more divergent in chemical defenses
51 than expected by chance

52 **4. *Synthesis.*** This is the first study to present metabolomics data for nearly 100 species in a diverse
53 Neotropical plant clade across the whole Amazonia. Our results demonstrate a role for plant-
54 herbivore interactions in shaping the clade's community assembly at a local scale, and suggest that
55 the high alpha diversity in Amazonian tree communities must be due in part to the interactions of
56 diverse tree lineages with their natural enemies providing a high number of niche dimensions.

57

58 **Key-words:** Amazon, chemical defenses, community assembly, local scale, metabolomics, *Inga*,
59 plant-herbivore interactions, regional scale, tropical rain forests.

60

61 **Resumen**

62 **1.** La increíble diversidad de los bosques Amazónicos se cree es el resultado de varios procesos.
63 Aunque los linajes de *Inga* se han dispersado repetidamente a lo largo de la Amazonía, las
64 interacciones entre plantas e insectos podrían ser el mecanismo más importante en el ensamblaje
65 de comunidades a escala local.

66 **2.** Usando métodos metabolómicos y filogenéticos, investigamos los patrones históricos de
67 ensamblaje de comunidades de plantas a lo largo de la Amazonía basándonos en el género
68 neotropical de árboles *Inga* (Leguminosae) en cuatro sitios, ampliamente separados.

69 **3.** Nuestros resultados demuestran un grado de estructura filogenética y una mezcla de chemotipos
70 a lo largo de la Amazonía, sugiriendo que aunque la biogeografía juegue un rol, la metacomunidad
71 para cualquier comunidad regional en la Amazonía es toda la cuenca Amazónica. Comunidades
72 locales son ensambladas por procesos ecológicos, donde todas las especies de *Inga* coexistiendo
73 en un mismo sitio son más divergentes en defensas químicas que al azar.

74 **4. Síntesis.** Este es el primer estudio que presenta datos metabolómicos para casi 100 especies de
75 árboles pertenecientes a un grupo Neotropical diverso a lo largo de su rango de distribución.
76 Nuestros resultados demuestran un rol para las interacciones entre plantas y herbívoros en el
77 ensamblaje de la comunidad de este clado a escala local, y sugiere que la alta diversidad alfa en
78 las comunidades de árboles Amazónicos puede deberse en parte a las interacciones de grupos de
79 árboles diversos con sus enemigos naturales.

80 **Palabras clave:** Amazonía, defensas químicas, ensamblaje de comunidades, escala local,
81 metabolómica, *Inga*, interacciones planta-herbívoro, escala regional, bosques tropicales lluviosos.

82

83 **1. Introduction**

84 Amazonian forests are considered one of world's richest plant assemblages, with an estimated
85 16,000 species of trees for the whole region (ter Steege et al., 2020), and more than 650 woody
86 species in a single hectare (Valencia et al., 2004). At a regional scale, recent studies have
87 highlighted the role of dispersal across the Amazon in assembling tree communities (Dexter et al.,
88 2017; Fine et al., 2014). At a local scale, there is still much debate regarding the ecological and
89 evolutionary mechanisms that determine the co-occurrence of large numbers of species at a site,
90 many of which are congeners. Some studies argue that niche differentiation may arise through
91 competition for resources or adaptation to abiotic niches (Chesson, 2000; Kraft, Adler, et al., 2015;
92 Kraft, Godoy, et al., 2015), while others claim that biotic factors such as natural enemy damage
93 may facilitate coexistence (Coley & Kursar, 2014). The central premise of the latter is that the
94 myriad of defenses against herbivores may generate key additional niche axes that allow
95 coexistence of a greater diversity of species (Levi et al., 2019).

96 The idea that the interactions between plants and their insect herbivores may contribute to
97 the assembly of communities has received considerable recent attention. Specifically, this theory
98 suggests that specialist pests may play a main role in maintaining the high local diversity of
99 rainforests by preventing most plant species from becoming abundant (Janzen, 1970; Connell,
100 1971; Comita et al., 2014). Species do not share herbivores with their nearby neighbors if they
101 have divergent defences (Becerra, 2007; Endara et al., 2017a), which gives a species the advantage
102 of reduced damage or “enemy release” (Yguel et al., 2011). This in turn may promote the
103 coexistence of species that are defensively divergent, increasing local plant species diversity
104 (Janzen, 1970; Becerra, 2007; Fine et al., 2013; Sedio & Ostling, 2013; Coley & Kursar, 2014;
105 Salazar et al., 2016a; Salazar et al., 2016b; Forrister et al., 2019). Kursar et al. (2009) reported that

106 co-occurring species of *Inga* in the Peruvian Amazon were more closely related yet differed more
107 in their defenses than expected by chance. Studies with other genera in the tropics reveal the same
108 patterns (e.g. *Bursera*, *Ficus*, *Piper*, *Protium*, *Psychotria*; Becerra, 1997; Becerra et al., 2009;
109 Coley & Kursar, 2014; Kursar et al., 2009; Salazar et al., 2018; Sedio, 2013; Wills et al., 2016).
110 Because plants have many types of defences that can evolve independently from one another
111 (Endara et al., 2017a), defensive traits may provide a large number of niche dimensions among
112 which a very large number of co-occurring species might sort in ecological time. Thus, plant-
113 herbivore interactions may be key to understanding the high local diversity in tropical forest
114 communities.

115 Relevant progress towards understanding the local and regional processes that underlie the
116 assembly of communities has been made in recent years, though largely focused on the
117 evolutionary attributes of species (phylogenetic history). These studies are based on the premise
118 that historical species interactions and environmental conditions of communities are reflected in
119 phylogenies, and that phylogeny is a good proxy for functional trait data that are difficult to obtain
120 (Mace et al., 2003), especially at the large scale that is necessary for such studies. Yet, if phylogeny
121 is only a proxy for species traits, and some traits may show low or no phylogenetic signal, an ideal
122 approach would be to directly compare the explanatory power of traits and phylogeny (Pearse et
123 al., 2014). Recent advances in analytical techniques have greatly enhanced the potential of
124 researchers to characterize trait diversity at unprecedented scales. One such exciting new
125 development is in the area of metabolomics. Specifically, mass spectrometry-based metabolomics
126 is a powerful tool to characterize the chemical composition of complex biological samples
127 containing tens to hundreds of individual compounds at the community or macroevolutionary scale
128 (Sedio et al., 2017). In particular, tandem mass spectrometry (MS/MS) facilitates the structural

129 comparison of unknown compounds and their comparison to global databases of known chemical
130 structures (Treutler et al., 2016; Wang et al., 2016).

131 Here, we use metabolomic and phylogenetic approaches to investigate the patterns of
132 assembly of plant communities across the whole Amazon basin. We focus our study on the
133 speciose (> 300 species), ecologically important and abundant Neotropical genus of trees, *Inga*
134 (Leguminosae). Our previous studies with *Inga* show that defences diverge rapidly and that
135 divergent defenses may contribute to coexistence in neighborhoods (~ meters; Kursar et al., 2009).
136 In this study, we examine community assembly at the regional (the Amazon basin,) and local scale
137 (within a site, ~ 100 ha) and build on previous work by incorporating a larger number of *Inga*
138 species (37 in Kursar et al. 2009 vs. 91 in this study) collected over their entire geographic range,
139 as well as a more resolved phylogeny and a more comprehensive chemistry dataset. Taken
140 together, we aim to provide a more robust test of the ideas proposed by Kursar et al. (2009) and to
141 extend the spatial scale from meters to kilometers.

142 At four widely separated sites, we characterize the chemical composition of 91 species,
143 which represents roughly 1/3 of known *Inga* species. We follow an untargeted approach that lets
144 us obtain the entire chemical profile of a species rather than quantifying a subset of metabolites.
145 In doing so, we can determine how many compounds are produced by each species and how many
146 compounds are shared between them.

147 A critical component of our analyses is to determine the chemical similarity between all
148 pairwise combinations of *Inga* species. However, this presents an apple/orange comparison
149 challenge as few compounds are shared between species. We therefore have developed methods
150 to account for the fact that two species may have different compounds that are structurally similar
151 (Coley et al., 2019; Endara et al., 2018; Forrister et al., 2019). We join other ecological researchers

152 pioneering metrics to classify chemical structure based on MS/MS spectra in order to quantify
153 differences between species (Sedio et al., 2018). Our untargeted methods provide data on hundreds
154 of compounds per species and we can generate a matrix of MS/MS based structural similarity
155 between every pair of compounds (Wang et al., 2016), which can allow for a calculation of
156 chemical similarity even when no compounds are shared between a pair of species. This in turn
157 allows us to better quantify both the chemical similarity among plant populations and to understand
158 how plant-herbivore interactions may play a role in the assembly of plant communities.
159 Specifically, we expect a lack of phylogenetic and chemical structure in the assembly of *Inga*
160 communities at a regional scale, suggesting that the metacommunity for any regional community
161 in the Amazon is the entire Amazon basin (Dexter et al., 2017). In contrast, the observation that
162 the suite of *Inga* at a given local site are more over-dispersed with respect to defences, would
163 suggest that local communities are assembled by ecological processes.

164

165 **2. Materials and Methods**

166 **2.1 Sampling.**

167 We sampled 91 *Inga* species across the Amazonia between July 2010 and September 2014.
168 Sampling was focused at four sites (~ 100 ha each) that include a wide range of soils along with a
169 large fraction of *Inga* diversity throughout the Amazon (Figure 1). At each site, we sampled all the
170 known *Inga* species: Nouragues, French Guiana, 4°N 53°W, with 46 species; Tiputini in the
171 Yasuní National Park, Ecuador, 0°N, 75°W, 41 species; Los Amigos in Madre de Dios, Peru, 13°S,
172 70°W, 39 species, and Km 41 near Manaus, Brasil, 2°S, 60°W, 29 species. The four sites are
173 lowland moist forests with no pronounced dry season. For simplicity in the text, each site will be
174 referred by the country only.

175 At each site, sampling was performed over six months and at the same time of the year.
176 We focused on expanding leaves of 0.5 – 4 m tall understory saplings, a key stage in the life cycle
177 of a tree (Green et al., 2014). More than 40 km of trails were walked regularly to search for plants,
178 and collections are widely separated. We focused our study on the chemical defenses of young
179 leaves because during this ephemeral stage they receive more than 75% of the herbivore damage
180 accrued during the lifetime of a leaf (Brenes-Arguedas et al., 2008; Coley & Aide, 1991; Kursar
181 & Coley, 2003), and the chemistry of expanding leaves has been shown to be very important for
182 shaping associations between plants and their insect herbivores (Endara et al., 2017a, 2018).

183 **2.2 Phylogenetic reconstruction of *Inga***

184 A phylogenetic tree for 165 *Inga* accessions, including all the taxa sampled at each site, was
185 reconstructed using a newly generated targeted enrichment (HybSeq) dataset of 810 genes. These
186 810 loci include those presented in Nicholls et al. 2015, supplemented with a subset of the loci
187 from Koenen et al. (2020). DNA library preparation, sequencing and the informatics leading to
188 final sequence alignments follow protocols in Nicholls et al. 2015. We used IQtree 2 (Minh et al.,
189 2020) to infer a phylogenetic tree from the complete dataset of 810 genes. We performed a
190 partitioned analysis (Chernomor et al., 2016) after inferring the best-partition scheme for the 810
191 genes and the best substitution model for each partition using ModelFinder module implemented
192 in IQtree 2 (Kalyaanamoorthy et al., 2017). The resulting phylogenetic tree was subsequently time-
193 calibrated using penalized likelihood implemented in the program treePL (Smith & O’Meara,
194 2012). We used cross-validation to estimate the best value of the smoothing parameter. We
195 implemented a secondary calibration point on the crown age of *Inga* with a minimum age of 6 Ma
196 and a maximum age of 10 Ma following previous estimates (Pennington et al., 2006; Richardson,

197 2001). Finally, the complete phylogeny was pruned to include only the 91 species for which
198 chemistry data were available.

199 **2.3 Characterization of leaf defensive chemistry**

200 *Secondary metabolites:*

201 For leaf defence analyses, expanding leaves were dried on silica gel at ambient temperature
202 immediately after collection in the field, and then stored at -20° C. Samples consisted of whole
203 leaves with little or no damage in order to control for potential defense induction, although
204 induction is rare in tropical trees like *Inga* (Bixenmann et al., 2016). The defense metabolome for
205 each species was determined using untargeted metabolomics methods. Defensive compounds were
206 extracted from dried leaf samples in the Coley/Kursar lab at the University of Utah using a solution
207 of (60:40, v/v) ammonium acetate buffered water, pH 4.8:acetonitrile, resulting in 2mL of retained
208 supernatant from 100mg (+/- 2.5 mg) of sample for chromatographic analysis (Wiggins et al.,
209 2016). Small molecules (50-2000 Da) of intermediate polarity were analyzed using
210 ultraperformance liquid chromatography (Waters Acquity I-Class, 2.1 x 150mm BEH C18 and 2.1
211 x 100 mm BEH Amide columns) and mass spectrometry (Waters Xevo G2 QToF) (UPLC-MS) in
212 negative ionization mode. Additionally, MS/MS spectra were acquired for each species by running
213 DDA (Data Dependent Acquisition Mode), whereby MS/MS data were collected for all
214 metabolites that were ionized above a set threshold (Total ion current / TIC of 5000).

215 *L-Tyrosine:*

216 Some *Inga* species invest in the overexpression of the essential amino acid L-tyrosine as an
217 effective chemical defense (Coley et al., 2019). Tyrosine is insoluble in our extraction buffer, so a
218 different protocol was used to determine the percentage of leaf dry weight. Following Lokvam et

219 al. (2006), extractable nitrogenous metabolites were extracted from a 5 mg subsample of each leaf
220 using 1 mL of aqueous acetic acid (pH 3) for 1 h at 85°C. Fifteen microliters of the supernatant
221 were injected on a 4.6 x 250 mm amino-propyl HPLC column (Microsorb 5u, Varian). Metabolites
222 were chromatographed using a linear gradient (17–23%) of aqueous acetic acid (pH 3.0) in
223 acetonitrile over 25 min. Mass of solutes in each injection were measured using an evaporative
224 light scattering detector (SEDERE S.A., Alfortville, France). Tyrosine concentrations were
225 determined by reference to a four-point standard curve (0.2–3.0 mg tyrosine/mL, $r^2=0.99$) prepared
226 from pure tyrosine.

227 **2.4 Data Analysis**

228 We employed a compound based molecular networking approach where we first group related
229 features into compounds and then we generate 1) a species by compound abundance matrix and 2)
230 a compound by compound MS/MS cosine similarity matrix. We combine these data into a pairwise
231 species similarity matrix which accounts for both shared compounds between species and the
232 MS/MS structural similarity of unshared compounds, following a similar approach to one
233 developed Sedio et al. (2017). All scripts from this study are deposited in a github repository
234 (Forrister & Soule, 2020; https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry).

235 ***Creation of species-by-compound matrix:***

236 Raw UPLC-MS data files were converted to mzXML format using the ‘raw2mzML’ package in
237 Python (Schmitt, 2016). Converted files were processed by species within each site (accession)
238 and for MS level 1 peak detection using the XCMS package in R (Smith et al., 2006), which
239 combined chromatographic features into features based on the mass/charge (m/z) ratio and
240 retention time (RT) of individual ions. We then grouped features into putative compounds using

241 CAMERA (Kuhl et al., 2012) which groups features that co-elute and have correlated abundance
242 traces between scans, identifying likely adducts and related features within compounds. Finally,
243 we removed from the analysis known contaminants and surfactants, as well as features with an
244 abundance less than 3x greater than the abundance of that feature in a blank (pure organic solvent).

245 After initial peak detection, features were aligned across accessions based on kernel density
246 clustering of m/z and RT, and putative compounds grouped based on the cosine similarity of
247 aligned feature abundance, resulting in a list of unique compounds across all samples. Here,
248 abundance is considered the intensity or total ion current (TIC) for each feature. Each sample was
249 then re-examined for all compounds to avoid data skewing during peak detection by accession.
250 Finally, in an effort to remove temporal variance in UPLC-MS performance, compound abundance
251 was normalized by the average abundance of a standard retention-time index run the same day.
252 This produced a data frame containing the normalized abundance of each compound within each
253 sample, which was converted to a wide format to create a sample-by-compound matrix where the
254 normalized abundance of each compound was assigned to a unique row (sample) and column
255 (compound). In order to create a species-level comparison of compound abundance, all replicates
256 (minimum of 5) per accession were combined into a single species level chemical profile by
257 averaging the abundance of each compound across all replicates for a given species.

258 It is important to note that while we consider our method of grouping features into putative
259 compounds to be fairly conservative, there remains the possibility of over- or under-splitting
260 features into distinct compounds, with the former being more common. To address this issue in
261 our method, the incorporation of MS/MS structural similarity (see *Creation of compound-by-*
262 *compound matrix*) of distinct compounds allows the overall chemical similarity of samples (see

263 **Chemical similarity of *Inga* species**) containing pseudo-replicated compounds to remain
264 mathematically the same.

265 **Creation of compound-by-compound matrix:**

266 MS compounds (grouped chromatographic features) were matched to their associated MS/MS
267 spectra based on the mz/RT of the parent ion isolated by DDA. A consensus MS/MS spectrum for
268 each compound was generated by averaging all scans matched to that compound. A single MS/MS
269 spectrum for each compound was then submitted to the Global Natural Products Social Molecular
270 Networking in .mgf format (GNPS; <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>;
271 Wang et al., 2016) for molecular networking. In R, the resulting network was used to create a
272 pairwise compound-by-compound similarity matrix based on the similarity of their MS/MS
273 fragmentation spectra. Here, the shortest through-network path between each compound pair was
274 calculated, and a similarity score was assigned using the cosine scores along that path:

275 Eq. 1 $Similarity_{A,B} = (\sum_{i=1}^n \frac{1}{i})^{-1}$

276 where n is the number of edges separating compound A and compound B, and *i* is the cosine score
277 of the current edge. The score ranges from 0 (completely dissimilar) to 1 (identical).

278 **Compound annotation:**

279 Our analysis yielded 6217 compounds from 91 *Inga* species and one species in its sister genus,
280 *Zygia mediana* (156 accessions including the same species from different sites). In order to
281 annotate compounds, we performed MS/MS spectral matching to all publicly available datasets in
282 GNPS as well as *in silico* fragmentation of the Universal Natural Products Database (Allard et al.,
283 2016; Gu et al., 2013) and our own in-house database built from compounds found in *Inga*
284 (Lokvam & Kursar, 2005). We further enumerated the library using *in silico* combinatorial
285 chemistry to generate ~75,000 plausible structures using the “scaffold” and “building block”

286 structures within the CLEVER application (Song et al., 2009). These enumeration structures were
287 chosen based on patterns of biosynthesis that we have observed in *Inga*. All compounds in this *in*
288 *silico* database were uploaded to GNPS as a spectral library after performing *in silico*
289 fragmentation using CFM-ID to predict MS/MS spectra (Allen et al., 2014). We also used
290 Network-Annotation Propagation (da Silva et al., 2018) to further annotate unknown compounds.
291 Library hits and *in silico* prediction suggest that these compounds consist primarily of
292 phenylpropanoids, flavonoid monomers, flavan3ol polymers, and saponin glycosides, which are
293 all classes known for their defensive function. These results confirm previous work done
294 classifying *Inga* chemistry (Kursar et al., 2009).

295 ***Chemical similarity between Inga species:***

296 Following Endara et al. (2018) with some modifications, we estimated chemical similarity between
297 species using the species-by-compound and compound-by-compound matrices. After creating
298 these matrices, compounds were grouped into saponins and phenolics based on m/z, RT, and
299 residual mass defect (RMD), and the species-by-compound matrix was separated based on this
300 grouping. Abundances in each matrix were then normalized such that total abundance of all
301 compounds in any given species was equal to 1.0.

302 Pairwise similarity for each species pair was calculated by quantifying the degree to which
303 two species contain compounds that overlap in the molecular network. This includes the degree to
304 which two species invest in the same compounds (species-by-compound), and the structural
305 similarity of compounds that are not shared between the two species (compound-by-compound).
306 These parameters are calculated as follows:

307 *Total similarity*
308 = *TIC overlap in shared compounds*
309 + *similarity of unshared compounds*

310 To calculate the TIC (abundance) overlap in shared compounds, the minimum TIC values
311 for all compounds that are shared between the two samples are summed. The similarity of unshared
312 compounds is calculated in a similar manner, by pairing the most similar compounds, taking the
313 minimum TIC value for those two compounds, and multiplying by the through-network similarity
314 score. For shared compounds, through-network similarity becomes mathematically obsolete as
315 similarity for the same compound is always equal to 1. Thus, the overall similarity score results as
316 a sum of the investment (TIC) in the same or structurally similar defenses between two samples.

317 The pairwise similarity calculation for each species pair was repeated separately for
318 phenolics and for saponins, resulting in a separate pairwise similarity matrix for each compound
319 class. The similarity matrices from each compound class were combined with tyrosine data to
320 produce an overall chemical similarity score for each sample pair according to the dry weight
321 investment in each of the three compound classes. For further details, please review our gitlab
322 repository (Forrister & Soule, 2020).

323 ***Leaf defensive chemistry and phylogenetic signal:***

324 Phylogenetic signal was estimated for the principal coordinates of the chemical similarity matrix
325 using Blomberg's K (Blomberg et al., 2003). K is close to zero for traits lacking phylogenetic
326 signal, but higher than 1 when close relatives are more similar than expected under the Brownian
327 motion model of character evolution. We used the function *phylosignal* in the R package *picante*
328 v.1.8.2 (Kembel et al., 2020).

329 ***Analysis of community assembly:***

330 We analyzed the assembly of *Inga* communities both at the local scale and at the level of the
331 Amazon basin (regional scale, including the whole Amazon basin). Using incidence data
332 (presence/absence), through a Bayesian approach with generalized linear mixed-effects model
333 (GLMM) in the R package *MCMCglmm* v.2.29 (Hadfield, 2019), we determined patterns of
334 phylogenetic/chemical structure across all the assemblages simultaneously. We partitioned
335 variance in the *Inga* species-by-site matrix into the effects of phylogenetic relatedness (termed
336 phylogenetic effect) and chemical similarity between *Inga* species (a chemical effect). The
337 magnitude of the effect of each term is determined by the magnitude of the variance associated
338 with it. The phylogenetic effect determines the contribution of the main effect of the *Inga*
339 phylogeny to the covariance and captures the variation in the *Inga* co-occurrence data explained
340 by pairwise phylogenetic distances between *Inga* species. The chemical effect is the contribution
341 of the main effect of *Inga* defensive chemistry to the covariance and captures the variation in the
342 *Inga* co-occurrence data explained by the chemical similarity between *Inga* species. Thus, if the
343 structuring of the communities is due to phylogenetic sorting, then the phylogenetic effect would
344 show the greatest variance in the model. In contrast, if the assembly of *Inga* is mainly due to the
345 occurrence of species with different chemistry, then the chemical effect would contribute the
346 greatest to the variance in the model. Because the *Inga* occurrence data is collected from several
347 sites across the Amazon basin, rather than consolidate the data across sites, we analyze the site-
348 specific incidence matrices as the geographic region information effect. In the model, this effect is
349 termed Geographical region (see Table S1).

350 Phylogeny and chemistry were incorporated into the model as variance-covariance
351 matrices of relatedness and similarity, respectively, in the random effect structure of the
352 generalized linear mixed effects model. Region effects were also fitted as random in the model.

353 We compared models that included between-site effects (analyses at the level of the whole Amazon
354 basin, as a random factor) with models that ignored between-site effects to assess patterns within
355 sites (hence, analyses at small spatial scales). For the analyses, parameter-expanded priors were
356 used for all variance components. The chain was run for 500,000 iterations with a burn-in of 50,000
357 and a thinning interval of 450. Because the response variable was incidence data, a Bernoulli error
358 distribution was applied.

359 We also used classic dispersion metrics to determine whether a local *Inga* assemblage is a
360 phylogenetically biased subset of the species that could coexist in that assemblage (Pearse et al.,
361 2014). We estimated whether the mean pairwise distance (MPD, mean of the phylogenetic distance
362 between all the members in a community), and the mean nearest taxon index (MNTD, mean of the
363 phylogenetic distance between a species and its closest relative or neighbor in the community),
364 where under- or over-dispersed compared to the null expectation derived from a random assembly
365 of same-size assemblages from the regional pool (Webb et al., 2002). To assess uncertainty, we
366 repeated this process 9999 times using the functions *ses.mpd* and *ses.mntd*, respectively, in the R
367 package *picante* v.1.8.2 (Kembel et al., 2020).

368 Within-site chemical dissimilarity was estimated following Vleminckx et al. (2018).
369 Observed dissimilarities between *Inga* species at each site were compared to the null expectation
370 of a lack of divergence or convergence for trait expression. For this, the species by compound
371 matrix (see above under the Chemical similarity between *Inga* species section) was randomized
372 by reshuffling the compounds and species equiprobably, preserving differences in the abundance
373 and presence/absence of compounds among species (Gotelli, 2000). Departure from the null
374 expectation was estimated as the mean of the difference between the observed and expected
375 dissimilarities between species at each site. This procedure was repeated 1000 times. A p-value

376 was obtained as the proportion of mean values above (over-dispersion) or below (under-
377 dispersion) zero.

378

379 **3. Results**

380 *Leaf defensive chemistry in Inga shows low phylogenetic signal*

381 We sampled young leaves from a minimum of five individual plants per species per site. A
382 compound accumulation curve shows that five plants capture on average ~75% of the compounds
383 encountered if more individuals are sampled (see Figure S1 in Supporting Information).

384 We determined chemical similarity between *Inga* species based on the similarity of
385 chemical structure and relative abundance of compounds. In general, closely related species of
386 *Inga* in the Amazonia tend to have different chemical defenses. Principal coordinates of the
387 chemical similarity matrix show low phylogenetic signal (PCO1 $K = 0.57$, $P = 0.001$; PCO2 $K =$
388 0.28 , $P = 0.06$), with estimates of K that are substantially lower than the expected value of 1 under
389 Brownian motion evolution.

390 *Low geographic signal of phylogeny and chemistry at regional scales*

391 Because phylogeny is a poor predictor for chemistry in *Inga*, it was possible for us to separate the
392 effect of chemistry and phylogeny in the analyses. Thus, we investigated the relative role of
393 phylogeny and chemical defenses against herbivores in the assembly of *Inga* communities at
394 different scales. Our community structure models at the regional and local scales incorporating
395 phylogenetic and chemical effects showed a differential role for both terms. At large spatial scales
396 (models with between-site information) the phylogenetic effect was larger than the chemical effect,
397 with 12% of the variation in the incidence of *Inga* species across the Amazon region attributed
398 solely to phylogeny, versus 6% attributed to chemistry (Table S1, Fig. 2). In fact, there is little

399 regional selectivity based on chemistry, with all sites showing strong overlap in chemical space
400 (Fig. 3). Geographic information showed a large effect in the model (Table S1).

401

402 *Chemistry is more important than phylogeny at structuring local communities*

403 To determine if chemistry or phylogeny influenced the assembly of species co-occurring at a single
404 site, we fitted community-level structure models at small spatial scales (without between-site
405 information). There was some phylogenetic sorting, but the chemical effect contributed the
406 greatest variation, with more than 60% of the *Inga* occurrence data explained by chemistry (Table
407 S1, Fig. 2). Thus, at small spatial scales, coexistence of *Inga* species is mainly due to the
408 occurrence of species with dissimilar chemical defenses.

409 We further evaluated phylogenetic structure within a community by estimating dispersion
410 metrics and compared the observed values with a null expectation generated by randomly
411 assembling same-size assemblages from the regional pool. None of the four Amazonian
412 communities showed phylogenetic structure (Table 1, Fig. 4).

413 In contrast, trait dispersion analyses showed significant chemical overdispersion for *Inga*
414 communities in the Amazonia. When similarity in all chemical classes was considered, the
415 chemical distance among all the *Inga* species within Peru, French Guiana, and Ecuador is
416 significantly larger than the null expectation (Table 2, Fig. 5). This effect was maintained for
417 phenolics and for saponins (except for Peru and French Guiana, Table 2). Brazil showed
418 significantly chemical overdispersion only for saponins (Table 2, Fig. 5).

419

420 **4. Discussion**

421 We have argued that at a regional level, there is essentially no limitation in the dispersal of species
422 across the Amazon such that the metacommunity for any regional community is the entire Amazon
423 basin (Dexter et al., 2017). In contrast, interactions between plant and insects could be a principal
424 mechanism structuring community assembly at a local scale (Coley & Kursar, 2014; Kursar et al.,
425 2009). Results from our analyses are consistent with these hypotheses. At a large scale, we found
426 a lack of chemical structure in the assembly of *Inga* communities, with low, but significant
427 geographic filtering based on ancestry. In contrast, at each of four widely separated sites in the
428 Amazon, co-occurring species of *Inga* are more different in defense chemistry than expected by
429 chance, implying that species with similar defensive traits are less likely to coexist in the same
430 community. Thus, herbivores may have a key role in niche differentiation of their host plants
431 promoting local diversity.

432

433 ***Low geographic signal for phylogeny and chemistry at regional scales.***

434 Consistent with the hypothesis that regional tree communities in the Amazon are influenced by
435 historical processes of widespread dispersal (Dexter et al., 2017), we found a low signal for
436 phylogeny and almost no signal for chemistry in the assembly of *Inga* communities across the
437 Amazon when between-region information was included (Table S1, Fig. 2). Nevertheless,
438 geographic region had a large effect in the model, implying that biogeography might play a role in
439 *Inga* community assembly at regional level (Table S1). Thus, although *Inga* lineages have
440 dispersed repeatedly across the Amazon (Dexter et al., 2017), the detected signal of regional
441 phylogenetic structure together with the geographic region term effect imply that closely related
442 species might be co-occurring within some regions, and that there are some differences in the
443 lineage composition between regions. These differences could be mediated by environmental

444 filtering at regional scale, such as the gradient in soils observed across the Amazon Basin
445 (Tuomisto et al., 2019). For chemistry, the extremely low signal in the assembly of *Inga*
446 communities at the regional level (Fig. 2) suggests that local assemblages are drawn from a
447 metacommunity representing the full chemical space exhibited by the genus (Fig. 3).

448

449 ***Chemistry not phylogeny structures local communities***

450 In contrast to regional patterns, analyses of community structure at a local scale showed that
451 chemistry better explained variation in the incidence of *Inga* at a single site than plant relatedness
452 (Fig. 2). Thus, defensive chemistry plays a key role in determining which plant species can coexist
453 in each community at small spatial scales. Analyses with phylogenetic dispersion metrics and
454 within-site functional similarity agreed with this hypothesis. Although our community
455 composition models suggest a degree of phylogenetic sorting in species composition (Table S1),
456 dispersion-trait analyses for the four *Inga* communities sampled showed no significant
457 phylogenetic clustering (Table 1, Fig. 4). Meanwhile, the species of *Inga* that are co-occurring in
458 Peru, French Guiana, Ecuador and Brazil are more different in their defensive chemistry than
459 expected by chance (Table 2, Fig. 5). Except for Brazil, this effect was more pronounced for
460 phenolics than for saponins (Table 2, Fig. 5). Phenolics are the most structurally diverse and
461 common compound class for the genus *Inga* (D. Forrister unpubl. results), which is the most
462 divergent among close relatives (Endara et al., 2015). This suggests that phenolics might be under
463 stronger selective pressure to diverge among co-occurring species than other defense classes or
464 that phenol biosynthesis is more easily modified. Given that for *Inga*, each defense class varies
465 independently of the others (Endara et al., 2017), defensive chemistry may represent many axes of
466 trait divergence.

467 Interactions of plant species with their enemies are likely the mechanism responsible for
468 the co-occurrence of species with divergent chemotypes. Specialist herbivores might be foraging
469 on species with similar defensive chemotypes. Within a site, this would allow defensively distinct
470 species to coexist and increase local plant diversity (Sedio & Ostling, 2013). In contrast, species
471 with similar defenses may share herbivores and suffer greater attack, making it more difficult for
472 them to colonize or to coexist in the same community. Thus, herbivores might be regulating the
473 structure of communities through negative-density dependence interactions at scales ranging from
474 meters to kilometers (Agrawal 2007, Becerra 2007, Lau & Strauss 2007, Forrister *et al.* 2019),
475 linking local systems to regional processes (Ricklefs, 2007).

476 An essential component of this proposition is that plant defenses influence host choice.
477 Previously, we found that at a given site, lepidopteran herbivores preferentially feed on subsets of
478 *Inga* species with similar defensive profiles and that different families of herbivores chose hosts
479 based on different defensive traits (Endara et al., 2017). In addition, we have shown that high
480 chemical similarity and shared herbivore communities are associated with a decrease in survival
481 and growth for neighboring plants at the 5-10 meter scale (Forrister et al., 2019). In this study, we
482 provide evidence that the antagonistic interactions with enemies are playing out across the entire
483 community, not just spatially proximal neighbors. Thus, the composition of plant species within a
484 community appears to respond to the entire community of herbivores that could potentially attack
485 them.

486 Because phylogeny is a synthetic measure for phylogenetically conserved traits, the low
487 phylogenetic structure in *Inga* at four widely separated communities suggest that other
488 mechanisms than herbivore pressure might not be contributing as much to their assembly. For
489 example, phylogenetically conserved traits associated with resource use, pollination and dispersal

490 are quite similar across *Inga* species (Endara et al., 2015; Kursar et al., 2009; Pennington et al.,
491 1997). Thus, it is hard to see how they would provide sufficient niche differentiation to explain the
492 coexistence of so many species. Alternatively, if we consider the almost infinite number of
493 possible defense profiles, there could be an enormous number of niches with respect to herbivores
494 (Coley & Kursar, 2014; Levi et al., 2019; Singer & Stireman, 2005). For *Inga*, anti-herbivore
495 defenses fall into at least six different independent axes of defense variation (Endara et al., 2017).
496 It clearly provides a multidimensional, if not hyperdimensional niche space for coexistence
497 (Hutchinson, 1957).

498 Are there parallels in other tropical regions? Several studies have shown that neighbors
499 growing within meters of each other differ in defenses, including the genera *Eugenia*, *Ocotea*
500 and *Psychotria* in Panama (Sedio et al., 2017), *Bursera* in Mexico (Becerra, 2007), *Piper* in
501 Costa Rica (Salazar et al., 2016a,b) and *Protium* in Peru (Vleminckx et al., 2018). Here we
502 extend this concept and show that these patterns of defense divergence hold true across a much
503 larger community of plants, not just immediate neighbors. It is quite striking that these patterns
504 are consistent even when we included in our analyses the *Inga* community in Panama, a site with
505 a different biogeographic history that is isolated from the Amazonian study sites (data not
506 shown). Similarly, community structure and trait dispersion analyses showed significant
507 overdispersion of defensive chemistry at the local scale (Fig S2 and S3). Thus, the similarity of
508 secondary metabolite profiles among species may play a large role in shaping community
509 assembly beyond the tropical forest in Amazonia.

510 **5. Conclusions**

511 A number of recent, independent studies suggest that herbivore pressure contributes to the high
512 local plant diversity, or coexistence, that is typical of plant communities in tropical rainforests

513 (Becerra, 2007; Forrister et al., 2019; Kursar et al., 2009; Salazar et al., 2016a,b; Sedio et al.,
514 2017; Vleminckx et al., 2018). Our phylogenetic and metabolomic approach provides evidence
515 for the key role that natural enemies play in the assembly of these local communities. Although
516 *Inga* species have dispersed freely across the Amazon, with some recent regional in-situ
517 speciation events, what seems to determine which species are allowed to coexist within a single
518 community are natural enemies.

519 Our results expand the spatial scale over which negative-density dependence mechanisms
520 mediate community assembly and bring into play processes related to ecological interactions
521 between populations at larger spatial scales. The fact that coexistence of closely related species is
522 allowed by divergence in defensive traits on scales ranging from meters to kilometers brings the
523 timescale of species sorting and species diversification close to each other (Ricklefs, 2007). This
524 leads us to hypothesize that herbivore pressure might be one of the drivers of species
525 diversification. Thus, divergent selection by herbivores could potentially be one of the main
526 factors behind both the maintenance and the origin of diversity in tropical forests.

527

528 **Acknowledgments**

529 We thank the Ministry of Environment and Water of Ecuador and the Ministry of Agriculture of
530 Peru for granting the research and exportation permits. Valuable field assistance was provided by
531 Wilder Hidalgo, Zachary Benavidez, Allison Thompson, Yamara Serrano and Mayra Ninazunta.
532 This work was supported by grants from the National Science Foundation (DEB-0640630 and
533 DIMENSIONS of Biodiversity DEB-1135733), and Nouragues Travel Grants Program, CNRS,

534 France, to T.A.K. and P.D.C., and the Secretaría Nacional de Educación Superior, Ciencia,
535 Tecnología e Innovación del Ecuador (SENESCYT) to M-J.E.

536

537 **Author contributions**

538 M-J.E., D.L.F., T.A.K. and P.D.C designed and conducted the research. M-J.E. designed and
539 performed the data analysis. D.L.F. and A.J.S. contributed to the metabolomic analysis. J.A.N.,
540 R.T.P., K.G.D., and O.L. contributed the phylogeny of *Inga*. M-J.E., D.L.F., A.J.S. and P.D.C.
541 wrote the manuscript, with input from K.G.D., O.L. and R.T.P.

542 **Data availability**

543 Chemical data and scripts to estimate chemical similarity are deposited in a github repository
544 (Forrister & Soule, 2020; https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry).

545

546 **References**

- 547 Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., Roussi, F.,
548 Litaudon, M., & Wolfender, J.-L. (2016). Integration of molecular networking and *in*
549 silico MS/MS fragmentation for natural products dereplication. *Analytical Chemistry*,
550 88(6), 3317–3323. <https://doi.org/10.1021/acs.analchem.5b04804>
- 551 Allen, F., Pon, A., Wilson, M., Greiner, R., & Wishart, D. (2014). CFM-ID: A web server for
552 annotation, spectrum prediction and metabolite identification from tandem mass spectra.
553 *Nucleic Acids Research*, 42(W1), W94–W99. <https://doi.org/10.1093/nar/gku436>
- 554 Becerra, J. X. (1997). Insects on Plants: Macroevolutionary Chemical trends in host use. *Science*,
555 276(5310), 253–256. <https://doi.org/10.1126/science.276.5310.253>
- 556 Becerra, J. X. (2007). The impact of herbivore–plant coevolution on plant community structure.
557 *Proceedings of the National Academy of Sciences*, 104(18), 7483–7488.
558 <https://doi.org/10.1073/pnas.0608253104>

- 559 Becerra, J. X., Noge, K., & Venable, D. L. (2009). Macroevolutionary chemical escalation in an
560 ancient plant-herbivore arms race. *Proceedings of the National Academy of Sciences*,
561 *106*(43), 18062–18066. <https://doi.org/10.1073/pnas.0904456106>
- 562 Bixenmann, R. J., Coley, P. D., Weinhold, A., & Kursar, T. A. (2016). High herbivore pressure
563 favors constitutive over induced defense. *Ecology and Evolution*, *6*(17), 6037–6049.
564 <https://doi.org/10.1002/ece3.2208>
- 565 Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in
566 comparative data: behavioral traits are more labile. *Evolution*, *57*(4), 717–745.
567 <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- 568 Brenes-Arguedas, T., Rios, M., Rivas-Torres, G., Blundo, C., Coley, P. D., & Kursar, T. A.
569 (2008). The effect of soil on the growth performance of tropical species with contrasting
570 distributions. *Oikos*, *117*, 1453–1460. <https://doi.org/10.1111/j.2008.0030-1299.16903.x>
- 571 Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for
572 phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997–1008.
573 <https://doi.org/10.1093/sysbio/syw037>
- 574 Chesson, P. (2000). General Theory of Competitive Coexistence in spatially-varying
575 environments. *Theoretical Population Biology*, *58*(3), 211–237.
576 <https://doi.org/10.1006/tpbi.2000.1486>
- 577 Coley, P. D., & Aide, T. M. (1991). Comparison of herbivory and plant defenses in temperate
578 and tropical broad-leaved forests. In P. W. Price, T. M. Lewinsohn, G. W. Fernandes, &
579 W. W. Benson (Eds.), *Plant-Animal Interactions: Evolutionary Ecology in Tropical and*
580 *Temperate Regions* (pp. 25–49). Wiley & Sons.
- 581 Coley, P. D., Endara, M., Ghabash, G., Kidner, C. A., Nicholls, J. A., Pennington, R. T., Mills,
582 A. G., Soule, A. J., Lemes, M. R., Stone, G. N., & Kursar, T. A. (2019).
583 Macroevolutionary patterns in overexpression of tyrosine: An anti- herbivore defence in
584 a speciose tropical tree genus, *Inga* (Fabaceae). *Journal of Ecology*, *107*(4), 1620–1632.
585 <https://doi.org/10.1111/1365-2745.13208>
- 586 Coley, Phyllis D., & Kursar, T. A. (2014). On tropical forests and their pests. *Science*,
587 *343*(6166), 35–36. <https://doi.org/10.1126/science.1248110>
- 588 Comita, L. S., Queenborough, S. A., Murphy, S. J., Eck, J. L., Xu, K., Krishnadas, M., Beckman,
589 N., & Zhu, Y. (2014). Testing predictions of the Janzen-Connell hypothesis: A meta-
590 analysis of experimental evidence for distance- and density-dependent seed and seedling
591 survival. *Journal of Ecology*, *102*(4), 845–856. <https://doi.org/10.1111/1365-2745.12232>
- 592 Connell, J. H. (1971). On the role of natural enemies in preventing competitive exclusion in
593 some marine animals and in rainforest trees. In P. J. Den Boer & G. R. Gradwell (Eds.),
594 *Dynamics of Populations* (pp. 298–312). Centre for Agricultural Publishing and
595 Documentations.

596 da Silva, R. R., Wang, M., Nothias, L.-F., van der Hooft, J. J. J., Caraballo-Rodríguez, A. M.,
597 Fox, E., Balunas, M. J., Klassen, J. L., Lopes, N. P., & Dorrestein, P. C. (2018).
598 Propagating annotations of molecular networks using *in silico* fragmentation. *PLOS*
599 *Computational Biology*, *14*(4), e1006089. <https://doi.org/10.1371/journal.pcbi.1006089>

600 Dexter, K. G., Lavin, M., Torke, B. M., Twyford, A. D., Kursar, T. A., Coley, P. D., Drake, C.,
601 Hollands, R., & Pennington, R. T. (2017). Dispersal assembly of rain forest tree
602 communities across the Amazon basin. *Proceedings of the National Academy of*
603 *Sciences*, *114*(10), 2645–2650. <https://doi.org/10.1073/pnas.1613655114>

604 Endara, M.-J., Coley, P. D., Ghabash, G., Nicholls, J. A., Dexter, K. G., Donoso, D. A., Stone,
605 G. N., Pennington, R. T., & Kursar, T. A. (2017). Coevolutionary arms race versus host
606 defense chase in a tropical herbivore–plant system. *Proceedings of the National Academy*
607 *of Sciences*, *114*(36), E7499–E7505. <https://doi.org/10.1073/pnas.1707727114>

608 Endara, M.-J., Nicholls, J. A., Coley, P. D., Forrister, D. L., Younkin, G. C., Dexter, K. G.,
609 Kidner, C. A., Pennington, R. T., Stone, G. N., & Kursar, T. A. (2018). Tracking of host
610 defenses and phylogeny during the radiation of neotropical *Inga*-feeding sawflies
611 (Hymenoptera; Argidae). *Frontiers in Plant Science*, *9*, 1237.
612 <https://doi.org/10.3389/fpls.2018.01237>

613 Endara, M.-J., Weinhold, A., Cox, J. E., Wiggins, N. L., Coley, P. D., & Kursar, T. A. (2015).
614 Divergent evolution in antiherbivore defences within species complexes at a single
615 Amazonian site. *Journal of Ecology*, *103*(5), 1107–1118. [https://doi.org/10.1111/1365-](https://doi.org/10.1111/1365-2745.12431)
616 [2745.12431](https://doi.org/10.1111/1365-2745.12431)

617 Fine, P. V. A., Metz, M. R., Lokvam, J., Mesones, I., Zuñiga, J. M. A., Lamarre, G. P. A., Pilco,
618 M. V., & Baraloto, C. (2013). Insect herbivores, chemical innovation, and the evolution
619 of habitat specialization in Amazonian trees. *Ecology*, *94*(8), 1764–1775.
620 <https://doi.org/10.1890/12-1920.1>

621 Fine, P. V. A., Zapata, F., & Daly, D. C. (2014). Investigating processes of neotropical rainforest
622 tree diversification by examining the evolution and historical biogeography of the
623 Protieae (Burseraceae): Evolution and historical biogeography of Protieae. *Evolution*,
624 *68*(7), 1988–2004. <https://doi.org/10.1111/evo.12414>

625 Forrister, D. L., & Soule, A. J. (2020). *Evolution of Inga Chemistry* [Gitlab Repository].
626 https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry

627 Forrister, Dale L., Endara, M.-J., Younkin, G. C., Coley, P. D., & Kursar, T. A. (2019).
628 Herbivores as drivers of negative density dependence in tropical forest saplings. *Science*,
629 *363*(6432), 1213–1216. <https://doi.org/10.1126/science.aau9460>

630 Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, *81*(9),
631 2606–2621. [https://doi.org/10.1890/0012-9658\(2000\)081\[2606:NMAOSC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2)

632 Green, P. T., Harms, K. E., & Connell, J. H. (2014). Nonrandom, diversifying processes are
633 disproportionately strong in the smallest size classes of a tropical forest. *Proceedings of*

634 *the National Academy of Sciences*, 111(52), 18649–18654.
635 <https://doi.org/10.1073/pnas.1321892112>

636 Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H.-Z., & Xu, X. (2013). Use of natural products as
637 chemical library for drug discovery and network pharmacology. *PLoS ONE*, 8(4),
638 e62839. <https://doi.org/10.1371/journal.pone.0062839>

639 Hadfield, J. (2019). *MCMCglmm* (2.29) [R].

640 Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative*
641 *Biology*, 22(0), 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039>

642 Janzen, D. H. (1970). Herbivores and the number of trees in tropical forests. *The American*
643 *Naturalist*, 104(940), 501–529.

644 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017).
645 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*,
646 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>

647 Kembel, S. W., Ackerly, D. D., Blomberg, S. P., Cornwell, W. K., Cowan, P. D., Helmus, M. R.,
648 Morlon, H., & Webb, C. O. (2020). *picante: Integrating phylogenies and ecology* (1.8.1)
649 [R]. <https://cran.r-project.org/web/packages/picante/picante.pdf>

650 Koenen, E. J. M., Kidner, C., Souza, É. R., Simon, M. F., Iganci, J. R., Nicholls, J. A., Brown, G.
651 K., Queiroz, L. P., Luckow, M., Lewis, G. P., Pennington, R. T., & Hughes, C. E. (2020).
652 Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid
653 legumes and reveals the polytomous origins of a large pantropical radiation. *American*
654 *Journal of Botany*, 107(12), 1710–1735. <https://doi.org/10.1002/ajb2.1568>

655 Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015).
656 Community assembly, coexistence and the environmental filtering metaphor. *Functional*
657 *Ecology*, 29(5), 592–599. <https://doi.org/10.1111/1365-2435.12345>

658 Kraft, N. J. B., Godoy, O., & Levine, J. M. (2015). Plant functional traits and the
659 multidimensional nature of species coexistence. *Proceedings of the National Academy of*
660 *Sciences*, 112(3), 797–802. <https://doi.org/10.1073/pnas.1413650112>

661 Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: An
662 integrated strategy for compound spectra extraction and annotation of liquid
663 chromatography/mass spectrometry data sets. *Analytical Chemistry*, 84(1), 283–289.
664 <https://doi.org/10.1021/ac202450g>

665 Kursar, T. A., & Coley, P. D. (2003). Convergence in defense syndromes of young leaves in
666 tropical rainforests. *Biochemical Systematics and Ecology*, 31(8), 929–949.
667 [https://doi.org/10.1016/S0305-1978\(03\)00087-5](https://doi.org/10.1016/S0305-1978(03)00087-5)

668 Kursar, T. A., Dexter, K. G., Lokvam, J., Pennington, R. T., Richardson, J. E., Weber, M. G.,
669 Murakami, E. T., Drake, C., McGregor, R., & Coley, P. D. (2009). The evolution of
670 antiherbivore defenses and their contribution to species coexistence in the tropical tree

- 671 genus *Inga*. *Proceedings of the National Academy of Sciences*, 106(43), 18073–18078.
672 <https://doi.org/10.1073/pnas.0904786106>
- 673 Levi, T., Barfield, M., Barrantes, S., Sullivan, C., Holt, R. D., & Terborgh, J. (2019). Tropical
674 forests can maintain hyperdiversity because of enemies. *Proceedings of the National*
675 *Academy of Sciences*, 116(2), 581–586. <https://doi.org/10.1073/pnas.1813211116>
- 676 Lokvam, J., Brenes-Arguedas, T., Lee, J. S., Coley, P. D., & Kursar, T. A. (2006). Allelochemic
677 function for a primary metabolite: The case of L-tyrosine hyper-production in *Inga*
678 *umbellifera* (Fabaceae). *American Journal of Botany*, 93(8), 1109–1115.
679 <https://doi.org/10.3732/ajb.93.8.1109>
- 680 Lokvam, J., & Kursar, T. A. (2005). Divergence in structure and activity of phenolic defenses in
681 young leaves of two co-occurring *Inga* species. *Journal of Chemical Ecology*, 31(11),
682 2563–2580. <https://doi.org/10.1007/s10886-005-7614-x>
- 683 Mace, E. S., Buhariwalla, K. K., Buhariwalla, H. K., & Crouch, J. H. (2003). A high-throughput
684 DNA extraction protocol for tropical molecular breeding programs. *Plant Molecular*
685 *Biology Reporter*, 21(4), 459–460. <https://doi.org/10.1007/BF02772596>
- 686 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler,
687 A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for
688 phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5),
689 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- 690 Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with
691 many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12), i44–i52.
692 <https://doi.org/10.1093/bioinformatics/btv234>
- 693 Pearse, W. D., Purvis, A., Cavender-Bares, J., & Helmus, M. R. (2014). Metrics and Models of
694 Community Phylogenetics. In L. Z. Garamszegi (Ed.), *Modern Phylogenetic*
695 *Comparative Methods and Their Application in Evolutionary Biology* (pp. 451–464).
696 Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_19
- 697 Pennington, R. T., Richardson, J. E., & Lavin, M. (2006). Insights into the historical construction
698 of species-rich biomes from dated plant phylogenies, neutral ecological theory and
699 phylogenetic community structure. *New Phytologist*, 172(4), 605–616.
700 <https://doi.org/10.1111/j.1469-8137.2006.01902.x>
- 701 Pennington, T. D., Revelo, N., Linklater, R., & Wise, R. (1997). *El género Inga en el Ecuador:*
702 *Morfología, distribución y usos*. Royal Botanic Gardens, Kew.
- 703 Richardson, J. E. (2001). Rapid diversification of a species-rich genus of neotropical rain forest
704 trees. *Science*, 293(5538), 2242–2245. <https://doi.org/10.1126/science.1061421>
- 705 Ricklefs, R. E. (2007). History and Diversity: Explorations at the intersection of ecology and
706 evolution. *The American Naturalist*, 170(S2), S56–S70. <https://doi.org/10.1086/519402>

707 Salazar, D., Jaramillo, A., & Marquis, R. J. (2016a). The impact of plant chemical diversity on
708 plant–herbivore interactions at the community level. *Oecologia*, *181*(4), 1199–1208.
709 <https://doi.org/10.1007/s00442-016-3629-y>

710 Salazar, D., Jaramillo, M. A., & Marquis, R. J. (2016b). Chemical similarity and local
711 community assembly in the species rich tropical genus *Piper*. *Ecology*, *97*(11), 3176–
712 3183. <https://doi.org/10.1002/ecy.1536>

713 Salazar, D., Lokvam, J., Mesones, I., Vásquez Pilco, M., Ayarza Zuñiga, J. M., de Valpine, P., &
714 Fine, P. V. A. (2018). Origin and maintenance of chemical diversity in a species-rich
715 tropical tree lineage. *Nature Ecology & Evolution*, *2*(6), 983–990.
716 <https://doi.org/10.1038/s41559-018-0552-0>

717 Schmitt, U. (2016). *Raw2mzML* (0.0.3) [Python]. <https://pypi.org/project/raw2mzml/#description>

718 Sedio, B. E. (2013). *Trait evolution and species coexistence in the hyperdiverse tropical forest*
719 *genus Psychotria* [PhD]. University of Michigan.

720 Sedio, B. E., Boya P., C. A., & Rojas Echeverri, J. C. (2018). A protocol for high-throughput,
721 untargeted forest community metabolomics using mass spectrometry molecular networks.
722 *Applications in Plant Sciences*, *6*(3), e1033. <https://doi.org/10.1002/aps3.1033>

723 Sedio, B. E., & Ostling, A. M. (2013). How specialised must natural enemies be to facilitate
724 coexistence among plants? *Ecology Letters*, *16*(8), 995–1003.
725 <https://doi.org/10.1111/ele.12130>

726 Sedio, B. E., Rojas Echeverri, J. C., Boya P., C. A., & Wright, S. J. (2017). Sources of variation
727 in foliar secondary chemistry in a tropical forest tree community. *Ecology*, *98*(3), 616–
728 623. <https://doi.org/10.1002/ecy.1689>

729 Singer, M. S., & Stireman, J. O. (2005). The tri-trophic niche concept and adaptive radiation of
730 phytophagous insects. *Ecology Letters*, *8*(12), 1247–1255. [https://doi.org/10.1111/j.1461-](https://doi.org/10.1111/j.1461-0248.2005.00835.x)
731 [0248.2005.00835.x](https://doi.org/10.1111/j.1461-0248.2005.00835.x)

732 Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing
733 mass spectrometry data for metabolite profiling using nonlinear peak alignment,
734 matching, and identification. *Analytical Chemistry*, *78*(3), 779–787.
735 <https://doi.org/10.1021/ac051437y>

736 Smith, S. A., & O’Meara, B. C. (2012). treePL: Divergence time estimation using penalized
737 likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690.
738 <https://doi.org/10.1093/bioinformatics/bts492>

739 Song, C. M., Bernardo, P. H., Chai, C. L. L., & Tong, J. C. (2009). CLEVER: Pipeline for
740 designing *in silico* chemical libraries. *Journal of Molecular Graphics and Modelling*,
741 *27*(5), 578–583. <https://doi.org/10.1016/j.jmgm.2008.09.009>

742 ter Steege, H., Prado, P. I., Lima, R. A. F. de, Pos, E., de Souza Coelho, L., de Andrade Lima
743 Filho, D., Salomão, R. P., Amaral, I. L., de Almeida Matos, F. D., Castilho, C. V.,

744 Phillips, O. L., Guevara, J. E., de Jesus Veiga Carim, M., Cárdenas López, D.,
745 Magnusson, W. E., Wittmann, F., Martins, M. P., Sabatier, D., Irumé, M. V., ...
746 Pickavance, G. (2020). Biased-corrected richness estimates for the Amazonian tree flora.
747 *Scientific Reports*, 10(1), 10130. <https://doi.org/10.1038/s41598-020-66686-3>

748 Treutler, H., Tsugawa, H., Porzel, A., Gorzolka, K., Tissier, A., Neumann, S., & Balcke, G. U.
749 (2016). Discovering regulated metabolite families in untargeted metabolomics studies.
750 *Analytical Chemistry*, 88(16), 8082–8090. <https://doi.org/10.1021/acs.analchem.6b01569>

751 Tuomisto, H., Van doninck, J., Ruokolainen, K., Moulatlet, G. M., Figueiredo, F. O. G., Sirén,
752 A., Cárdenas, G., Lehtonen, S., & Zuquim, G. (2019). Discovering floristic and
753 geoeological gradients across Amazonia. *Journal of Biogeography*, 46(8), 1734–1748.
754 <https://doi.org/10.1111/jbi.13627>

755 Valencia, R., Foster, R. B., Villa, G., Condit, R., Svenning, J.-C., Hernandez, C., Romoleroux,
756 K., Losos, E., Magard, E., & Balslev, H. (2004). Tree species distributions and local
757 habitat variation in the Amazon: Large forest plot in eastern Ecuador. *Journal of Ecology*,
758 92(2), 214–229. <https://doi.org/10.1111/j.0022-0477.2004.00876.x>

759 Vleminckx, J., Salazar, D., Fortunel, C., Mesones, I., Dávila, N., Lokvam, J., Beckley, K.,
760 Baraloto, C., & Fine, P. V. A. (2018). Divergent secondary metabolites and habitat
761 filtering both contribute to tree species coexistence in the Peruvian Amazon. *Frontiers in*
762 *Plant Science*, 9, 836. <https://doi.org/10.3389/fpls.2018.00836>

763 Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D.,
764 Watrous, J., Kaponó, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A.
765 V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E.,
766 Sandoval-Calderón, M., ... Bandeira, N. (2016). Sharing and community curation of
767 mass spectrometry data with Global Natural Products Social Molecular Networking.
768 *Nature Biotechnology*, 34(8), 828–837. <https://doi.org/10.1038/nbt.3597>

769 Webb, C. O., Ackerly, D. D., McPeck, M. A., & Donoghue, M. J. (2002). Phylogenies and
770 community ecology. *Annual Review of Ecology and Systematics*, 33(1), 475–505.
771 <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>

772 Wiggins, N. L., Forrister, D. L., Endara, M.-J., Coley, P. D., & Kursar, T. A. (2016).
773 Quantitative and qualitative shifts in defensive metabolites define chemical defense
774 investment during leaf development in *Inga*, a genus of tropical trees. *Ecology and*
775 *Evolution*, 6(2), 478–492. <https://doi.org/10.1002/ece3.1896>

776 Wills, C., Harms, K. E., Wiegand, T., Punchi-Manage, R., Gilbert, G. S., Erickson, D., Kress, W.
777 J., Hubbell, S. P., Gunatilleke, C. V. S., & Gunatilleke, I. A. U. N. (2016). Persistence of
778 neighborhood demographic influences over long phylogenetic distances may help drive
779 post-speciation adaptation in tropical forests. *PLOS ONE*, 11(6), e0156913.
780 <https://doi.org/10.1371/journal.pone.0156913>

781 Yguel, B., Bailey, R., Tosh, N. D., Vialatte, A., Vasseur, C., Vitrac, X., Jean, F., & Prinzing, A.
782 (2011). Phytophagy on phylogenetically isolated trees: Why hosts should escape their
783 relatives. *Ecology Letters*, *14*(11), 1117–1124. [https://doi.org/10.1111/j.1461-](https://doi.org/10.1111/j.1461-0248.2011.01680.x)
784 [0248.2011.01680.x](https://doi.org/10.1111/j.1461-0248.2011.01680.x)

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807 **Tables**

808 Table 1. Results for the phylogenetic community structure analyses for each *Inga* community
809 sampled. Marginally significant values are bolded. N= number of taxa in each community.
810 MPD= mean pairwise distance, MNTD= mean nearest taxon distance. Observed Z=
811 standardized effect size of mean pairwise distance versus null model.

812

Site	N	Metric	Observed	Observed Z	p-value
Peru	41	MPD	0.0334	-0.1683	0.471
		MNTD	0.0152	0.5350	0.710
French Guiana	43	MPD	0.0347	0.5464	0.713
		MNTD	0.0161	1.5045	0.929
Ecuador	41	MPD	0.0348	0.6032	0.735
		MNTD	0.0163	1.3836	0.913
Brazil	29	MPD	0.0336	-0.0603	0.535
		MNTD	0.0163	0.0686	0.534

813

814

815

816

817

818

819

820

821

822

823 Table 2. Within-site chemical dissimilarity analyses separated by compound class. Significant
 824 values are bolded. O-E indicates the average difference in observed chemical similarity values
 825 compared to a randomized null model.

Site	Compound Class	O- E	95% CI (lower)	95% CI (upper)	p-value
Peru	All Chemistry	0.061	0.05300958	0.070	2.20E-16
	Phenolics	0.063	0.05517029	0.072	2.20E-16
	Saponins	-0.008	0.0180129614	0.000	0.05
Brazil	All Chemistry	0.003	-0.009744713	0.016	0.6088
	Phenolics	-0.39	-0.3999554	-0.380	2.20E-16
	Saponins	0.04	0.03395981	0.055	2.10E-15
French Guiana	All Chemistry	0.021	0.01112708	0.031	4.29E-05
	Phenolics	0.150	0.1409909	0.160	2.20E-16
	Saponins	-0.119	-0.1288449	-0.110	2.20E-16
Ecuador	All Chemistry	0.131	0.1218379	0.140	2.20E-16
	Phenolics	0.248	0.2385003	0.258	2.20E-16
	Saponins	0.070	0.05877019	0.083	2.20E-16

826

827 Table S1. Proportion of variation in *Inga* occurrence data attributed to phylogeny and chemistry
 828 effects. Columns contain the posterior modes (with 95% confidence intervals in
 829 parentheses) for the estimates. See Materials and Methods for a description of each term.

830

831

832

833

834 **Figure Captions**

835 Figure 1. Map of study sites at (1) Nouragues, French Guiana, (2) Tiputini, Ecuador, (3) Los
836 Amigos, Peru, and (4) Manaus, Brazil. Size is weighted by the sample size (n) of *Inga* trees
837 present at each site.

838 Figure 2. Proportion of variation in *Inga* community assembly explained by phylogeny and
839 chemistry at the regional (between sites) and local (within sites) scales. Bars represent the
840 mean +/- the standard error.

841 Figure 3. Principal Coordinates Analysis (PCoA) of chemical distance between species accessions
842 (estimated as 1 minus our chemical similarity score). Accessions are colored by site, and
843 ellipsoids for each site represent the 95% confidence interval around their mean position
844 in chemical space.

845 Figure 4. Relationship between the number of *Inga* species sampled and the mean pairwise
846 phylogenetic distance (MPD, left), and the mean nearest taxon distance (MNTD, right) in
847 the Amazon. Solid line represents the null expectation for MPD and MNTD respectively.
848 Dotted line represents the 95% confidence interval of the null expectation. Sites are
849 differentiated by shape, and significance is indicated by shading (no fill = non-significant,
850 grey = $p < 0.10$).

851 Figure 5. Within-site chemical similarity analysis. Boxplots represent 1000 bootstrap iterations of
852 the difference between observed (real data) and expected (null model) chemical similarity
853 values at each site, separated by compound class. Significance is indicated by asterisks (ns
854 = non-significant ; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$). All country names
855 excepting French Guiana (F. G.) are spelled out.

856 Figure S1. Compound accumulation curve calculated using *specaccum* (method = “random”,
857 permutations = 100) from the “vegan” package in R on a focal taxon with 22 distinct plant
858 samples (N65). 122 out of 165, or 73.9% of total compounds are captured with just 5 plants.

859 Figure S2. Proportion of variation in *Inga* community assembly explained by phylogeny and
860 chemistry at the regional (between sites) and local (within sites) scales including Panama
861 and Amazonia. Bars represent the mean +/- the standard error.

862 Figure S3. Within-site chemical similarity analysis including Panama and Amazonia. Boxplots
863 represent 1000 bootstrap iterations of the difference between observed (real data) and
864 expected (null model) chemical similarity values at each site, separated by compound class.
865 Significance is indicated by asterisks (ns = non-significant ; * = $p < 0.05$; ** = $p < 0.01$;
866 *** = $p < 0.001$). All country names excepting French Guiana (F. G.) are spelled out.

867

868

869

870

871

872

873

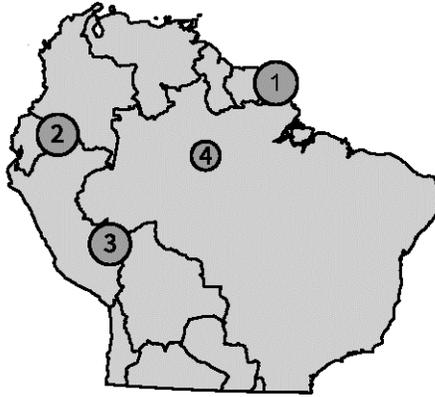
874

875

876

877

878



879

880 **Figure 1.**

881

882

883

884

885

886

887

888

889

890

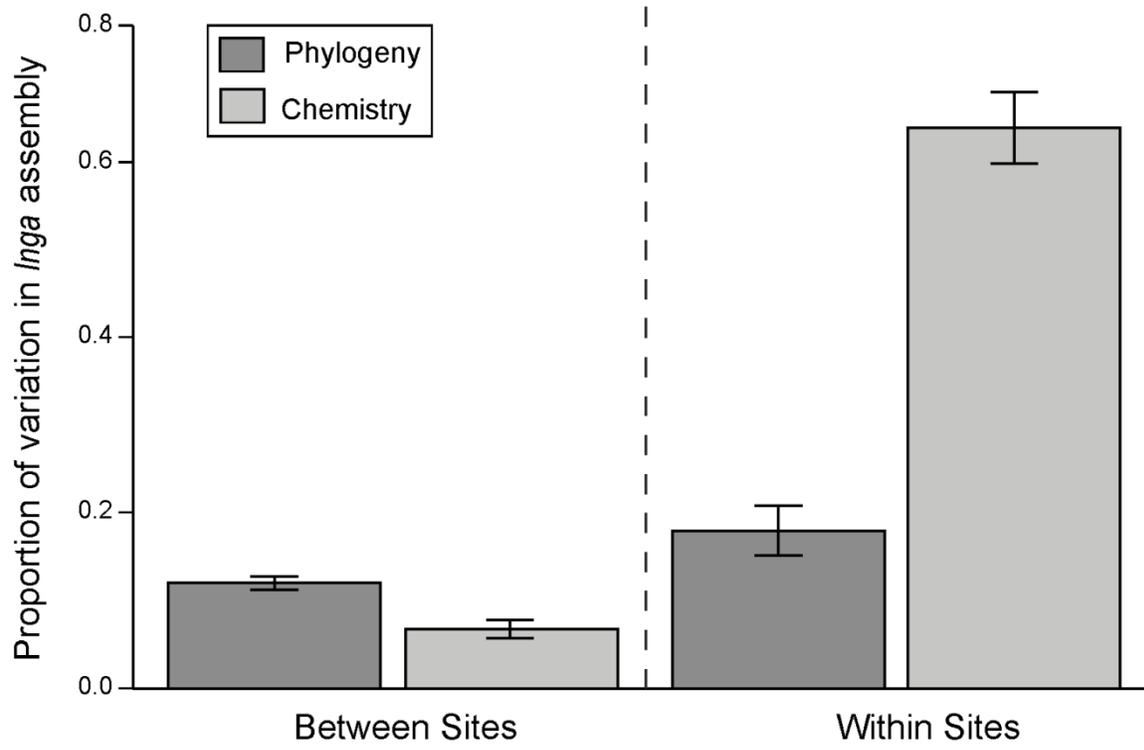
891

892

893

894

895



896

897 **Figure 2**

898

899

900

901

902

903

904

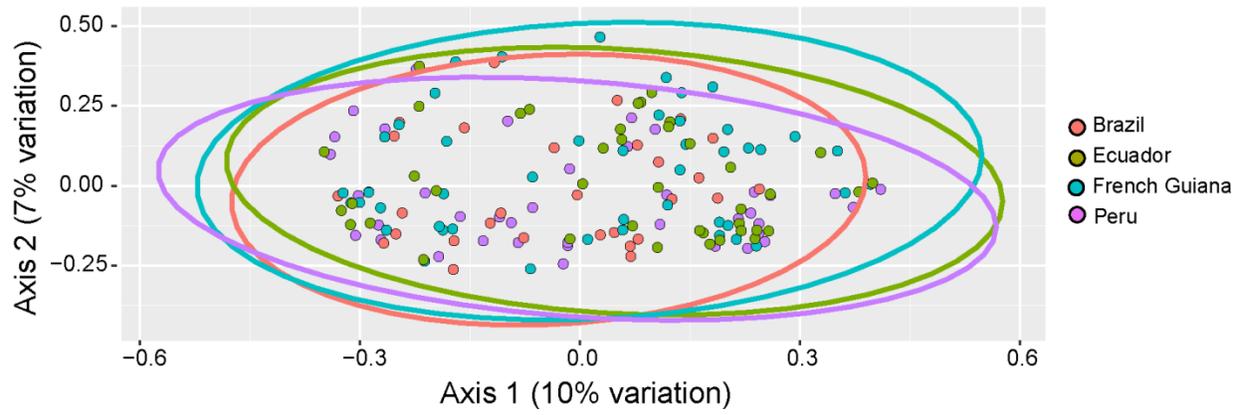
905

906

907

908

909



910

911 **Figure 3**

912

913

914

915

916

917

918

919

920

921

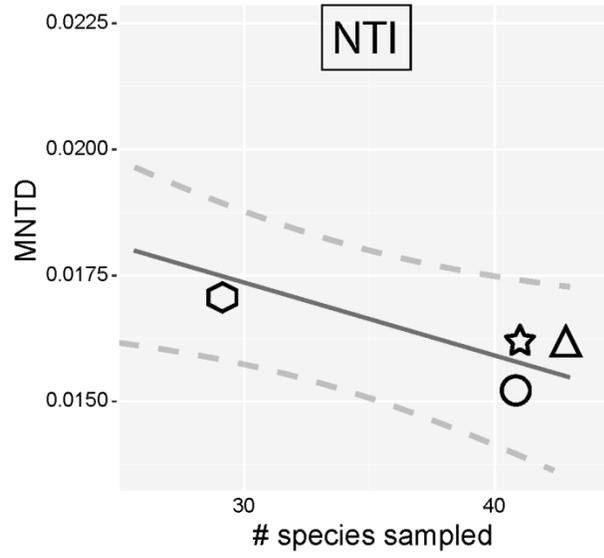
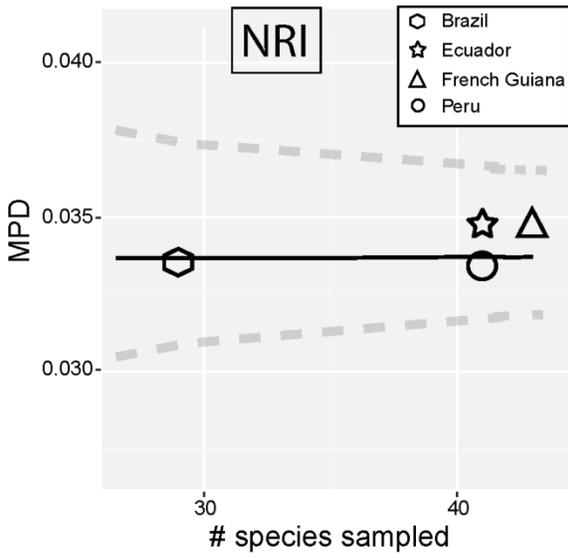
922

923

924

925

926



927

928 **Figure 4**

929

930

931

932

933

934

935

936

937

938

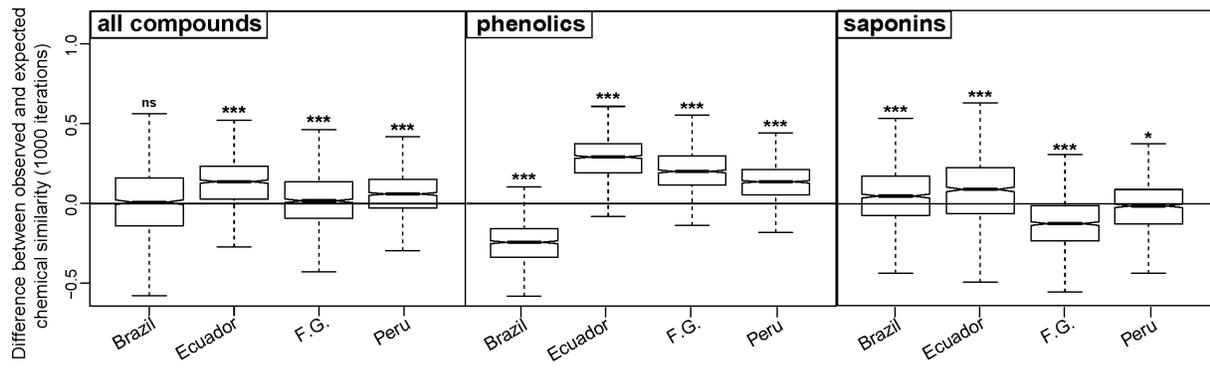
939

940

941

942

943



944

945

946

947 **Figure 5**

948

949