



This is a repository copy of *Crop origins explain variation in global agricultural relevance*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/175410/>

Version: Accepted Version

Article:

Milla, R. and Osborne, C.P. orcid.org/0000-0002-7423-3718 (2021) Crop origins explain variation in global agricultural relevance. *Nature Plants*, 7 (5). pp. 598-607.

<https://doi.org/10.1038/s41477-021-00905-1>

This is a post-peer-review, pre-copyedit version of an article published in *Nature Plants*. The final authenticated version is available online at: <https://doi.org/10.1038/s41477-021-00905-1>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

1 **Crop origins explain variation in global agricultural relevance**

2

3 Rubén Milla^{1*} and Colin P. Osborne²

4

5 ¹Departamento de Biología, Geología, Física y Química Inorgánica. Universidad Rey
6 Juan Carlos. Tulipán s/n, 28933 Móstoles, Spain. Orcid ID: <https://orcid.org/0000-0001-8912-4373>.
7

8 ²Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN,
9 United Kingdom. Orcid ID: <https://orcid.org/0000-0002-7423-3718>.

10

11 *Author for correspondence:

12 Rubén Milla,

13 Tel: +34 914888288

14 E-mail: ruben.milla@urjc.es

15

16

17 **Abstract**

18 Human food production is dominated globally by a small number of crops. Why certain crops
19 have attained high agricultural relevance while others have remained minor might partially
20 stem from their different origins. Here, we analyse a dataset of 866 crops to show that seed
21 crops and species originating from seasonally dry environments tend to have the greatest
22 agricultural relevance, while phylogenetic affinities play a minor role. These patterns are
23 nuanced by root and leaf crops and herbaceous fruit crops having older origins in the
24 aseasonal tropics. Interestingly, after accounting for these effects, we find that older crops are
25 more likely to be globally important and are cultivated over larger geographical areas than
26 crops of recent origin. Historical processes have therefore left a pervasive global legacy on the
27 food we eat today.

28

29

30

31 **Introduction**

32 The diversity of crops supplying our global food system is remarkably narrow. Of the calories
33 we consume at the global scale, ca. 60% come from either rice, wheat, soya or maize, or from
34 livestock fed on those staple grains^{1,2}. Diets are less restricted at the national scale, yet 90% of
35 the food supply for every country on Earth can still be accounted for by just 103 species from a
36 pool of several thousand food plants³. Our restricted diets at the global scale contrast with the
37 nutrition of pre-agricultural, hunter-gatherer cultures, who collected a diverse range of locally
38 available wild plant and animal foods⁴. Historical developments in agriculture and the
39 connections between societies have thus moved our current interactions with food plants
40 towards specialization⁵.

41 A specialized nutritional niche is evolutionarily advantageous under ecological stability, but
42 entails high risks in a changing world⁶. For example, the capacity of major crops to meet
43 growing food demands is uncertain⁷, and is becoming less secure as the climate changes⁸.
44 Moreover, the widespread use of monocultures of elite varieties hinders resilience against
45 pathogen outbreaks or extreme climatic events, which are increasingly common^{9,10}. Other
46 consequences of specialization extend beyond agronomic resilience, and include health
47 problems like increased rates of adult diabetes or over-nourishment^{11,12}. Agricultural
48 homogenization is thus recognized as a major contemporary problem, with significant
49 consequences for food security and sustainability. Homogenization is typically blamed on a
50 number of recent changes, including imperialism and colonisation, intensified inter-continental
51 trade, increased global access to commodity markets, the expansion of western dietary habits,
52 and the growth of multinational food industries^{8,13}. However, the roots of convergence in food
53 production might lie deeper in time, arising from additional temporal, spatial and biological
54 drivers. In this work, we address this topic by investigating the historic, geographic and
55 phylogenetic causes of variation among crop species in their agricultural relevance.

56 Variation in the relevance of crops arises first from their profiles as food sources. Crops differ
57 widely in their nutritional values. Some sustain the caloric and protein needs of people or
58 livestock, while others supply essential nutrients required in smaller quantities, preserve other
59 foods against microbial spoilage, provide flavouring or have a mixture of roles ¹⁴. These
60 functions are required in differing amounts, which explains why the food system is dominated
61 by crops meeting bulk caloric needs, predominantly starchy grains and vegetables. However,
62 crops with similar food profiles differ widely in global production (Figure 1). For example, of
63 the several species of *Triticeae* cultivated for their seeds, bread wheat is one of the big four
64 contributors to global agriculture, while rye and oats play a minor role. Similarly, crop species
65 in the *Theobroma* or *Prunus* genera, or in the family Cucurbitaceae, differ widely in extent of
66 cultivation ¹. Therefore, factors other than food profiles clearly account for variation among
67 crops in agricultural relevance.

68 Crop origins, defined in terms of time, space and phylogenetic background, might underpin the
69 predominance of particular species. Ancient crops may have benefited from priority effects, by
70 filling agricultural niches, providing the foundations for agricultural cultures and thus
71 narrowing options for later crops. Older crops have also had longer to diversify under
72 cultivation, and thus to adapt to local environments as their geographic ranges widened. The
73 identities of the four globally dominant grain crops are consistent with this expectation, since
74 they are each crops domesticated during the earliest transitions from hunter-gatherer
75 subsistence to agricultural economies in three major geographical centres: western Asia
76 (wheat), China (rice and soya) and Mesoamerica (maize) ¹⁵. Archaeological evidence shows
77 that people expanded the ranges of these crops across contiguous continental regions during
78 pre-history ^{13,16}. Pre-emption of agricultural niches might also happen at the phylogenetic
79 scale. The emergence of a relevant domesticate from an evolutionary lineage might pre-empt
80 opportunities for other similar crops in that lineage to achieve relevance. If true, agricultural
81 relevance should be over-dispersed in the phylogenetic tree of flowering plants. Previous work

82 has only found modest phylogenetic dependence of the presence and frequency of crops in
83 the different families of flowering plants ¹⁷, but we lack similar analyses for agricultural
84 relevance. Similarly, environmental conditions at sites of origin might influence the proclivity
85 of crops to attain widespread use. The suitability of terrestrial lands for agriculture is diverse,
86 with savannas, grasslands and temperate biomes dominating the global area under cultivation
87 ¹⁸, and areas with moderate rainfall and temperature regimes harbouring the bulk of human
88 populations ¹⁹. Some of the most prominent starchy staples originated in savanna and
89 Mediterranean-type climate regions ²⁰. Therefore, crops from seasonal climates might be
90 better pre-adapted to a majority of global agricultural lands, and thus might become
91 widespread with higher likelihoods ²¹. Food usage also seems to have influenced the order in
92 which crops were domesticated, with starchy grain crops typically the earliest contributors to
93 agricultural development for food, followed by legumes, with leafy vegetables, spices, oil crops
94 and fruits generally arriving later on ^{22,23}. However, preservation biases might influence these
95 patterns, and more recent work shows that starchy tubers and crops from the humid tropics
96 could be more ancient than previously thought ^{24,25}. There is some early literature advocating
97 for this view ²³, but comprehensive quantitative analyses are missing. Crucially, the usage,
98 historical, geographical and phylogenetic origins of crops are highly inter-related ²⁶, which
99 makes the various effects challenging to disentangle.

100 Here, we combine a global quantitative database on the origins and usage of 866 food crops ²⁷
101 with global crop production data ¹, to statistically disentangle the functional, historical,
102 geographical and climatic drivers of variation in crop origins, and their consequences for global
103 agricultural relevance. In a first step, we analyse how crop antiquity differs among types of
104 crops, and among diverse phylogenetic and climatic origins. In a second step, we summarize
105 the global area of each crop species under cultivation during the period 2006-2016, as a
106 measure of each crop's current agricultural relevance, and investigate its dependence on crop
107 origins. We aim to address the following questions: 1) Do ancient and recent crops have

108 different phylogenetic and climate origins? and 2) Do historical, phylogenetic and climatic
109 origins help to explain variation in current agricultural relevance among crops? To our
110 knowledge, this is the first comprehensive, quantitative assessment of the role of crop origins
111 in explaining agricultural specialization.

112

113 Results

114 The associations between crop antiquity, type and climate may provide insights into crop
115 origins, but also contribute the information needed to control for confounding effects when
116 investigating the relationship of crop antiquity to modern agricultural relevance. As expected,
117 we found complex interactions between crop type, climate of origin and antiquity.

118 The rate at which plant species were adopted for cultivation increased steadily from ca. 12,000
119 years ago onwards, but began declining in historically recent times (Fig. 2A). The various types
120 of crops tended to be taken into cultivation at different paces (Fig. 2A, ANOVA main effect,
121 type of crop, $F_{4,281} = 5.63$, $P < 0.001$, Supplementary Table 1). Seed crops and a number of root
122 and herbaceous fruit crops tend to be of earlier origin, while woody fruit and leaf crops tend to
123 be over-represented among recent crops (Fig. 2A). Climate at the location of crop origins also
124 explained variation in crop antiquity, but with different directions and strengths for the various
125 types of crops (ANOVA interaction, type of crop x temperature climate regime, $F_{5,281} = 5.05$, $p <$
126 0.001 , Supplementary Table 1). In particular, root and leaf crops, and herbaceous fruit crops,
127 tended to originate earlier in hot, aseasonal temperature regimes (*i.e.* the tropics; Fig. 2B).

128 Crop antiquity showed a modest, though statistically significant, degree of phylogenetic signal
129 (Fig. 3), and only 19 out of 301 crops showed over-dispersion in their LIPA scores for antiquity
130 (Supplementary Data 1). Certain clades harboured mostly ancient crops, such as Triticeae (*i.e.*
131 cereals) and Fabaeae (*i.e.* pulses) tribes (Fig. 3, Supplementary Data 1). Others, like Potentilleae,
132 Grossulariaceae or Ericaceae only included crops of recent origin (Fig. 3, Supplementary Data
133 1). However, since the phylogenetic structure of crop antiquity was modest, a number of
134 phylogenetically sister crops entered into cultivation at widely different historical times, such
135 as yam species in the *Dioscorea* genus, or *Oxalis*, *Opuntia* and *Stenocereus* crops (Fig. 3,
136 Supplementary Data 1).

137 The results outlined above showed that crop type, climatic origins and phylogeny are all
138 related to crop antiquity. It was therefore vital to account for these factors when investigating
139 the role of antiquity in modern agricultural relevance. Below we describe how, after these
140 factors had been statistically accounted for, crop antiquity still explained a remarkable part of
141 the variation among crops in their agricultural relevance.

142 There was statistical evidence that both the type of crop and the climate regime influenced the
143 likelihood of becoming a major crop (Supplementary Table 2, Fig 4). Woody fruit species from
144 cool, seasonal temperature regimes were more likely to become major crops than fruit trees
145 originating in non-seasonal and hot climates (*i.e.* the tropics) (Fig 4B). Root and herbaceous
146 fruit crops from warm regions without pronounced thermal but high rainfall seasonality (*i.e.*
147 the seasonally dry tropics and sub-tropics) were also more likely to become major (Fig 4 B,C).
148 Finally, after accounting for crop type and climate, there was a positive relationship between
149 crop antiquity and the likelihood of becoming a major crop (Analysis of Deviance main effect,
150 crop antiquity, $Dev_{1,293} = 37.74$, $P < 0.001$, Supplementary Table 2, Fig 4A), such that crops first
151 cultivated in antiquity were more likely to become major crops.

152 The variation in global production among major crops (Fig. 1) partially depended on the crop
153 type, with seed crops more important than leaf, root and fruit crops (ANOVA main effect, type
154 of crop, $F_{4,97} = 4.95$, $p = 0.001$, Supplementary Table 3, Fig. 5D). However, variation within crop
155 types was high (Figs. 1, 5D). Thus, global production also depended on the climate origins of
156 crops, such that those originating in seasonally dry climates tended to be cultivated more
157 widely than species of wet, aseasonal climates (Fig. 5C). Crucially, after these effects of crop
158 type and climate were accounted for, the antiquity of crop origins still had a highly significant
159 effect on global production (ANOVA main effect, antiquity, $F_{1,97} = 6.56$, $p = 0.012$,
160 Supplementary Table 3; Fig. 5A). This finding was robust to whether recent or older data on
161 global production were used as response (Supplementary Fig. 1, Supplementary Table 4), and

162 to whether crops that tend to preserve well in the archaeological record, and thus yield more
163 reliable estimates of crop antiquity, or crops with poorer preservation were used
164 (Supplementary Figs. 2-3, Supplementary Tables 5-6).

165 Global agricultural production showed a very modest phylogenetic signal, regardless of
166 whether modern or older production data were analysed (Fig. 6, Supplementary Fig. 4).
167 Indeed, many of the widely cultivated crops, such as soybean, sunflower, peanuts, coffee or
168 cassava showed significant phylogenetic over-dispersion (Supplementary Data 1), and thus
169 their global production was distinct when compared to their phylogenetic neighbourhoods
170 (Fig. 6). However, most clades included a large number of minor crops, which contributed to
171 some degree of phylogenetic clustering (Fig. 6, Supplementary Data 1). A few clades, like the
172 *Prunus* genus and the Benincaseae tribe, clustered together species with modest to high global
173 production (Fig. 6, Supplementary Data 1).

174 **Discussion**

175 Our results show that 36% of the variation in global agricultural production of major crops is
176 explained by their antiquity, climatic and phylogenetic diversities. Climatic influences
177 depended on crop type, such that major root and herbaceous fruit crops were more likely to
178 originate in the seasonally dry (sub)tropics, while major fruit trees were more likely to be
179 temperate in origin. Seed crops and crops from seasonally dry climates tended to be cultivated
180 over a larger area at the global scale. Interestingly, after accounting for these complex
181 interactions between crop type and climate, antiquity remained a significant predictor of
182 global production. Thus, crop antiquity stood out as an important correlate of global relevance,
183 despite the undoubted influence of cultural differences, dissemination histories, and of many
184 other peculiarities of the history of each individual crop. Explaining when and why certain
185 plants were brought into cultivation, and what makes human food provision specialized, are
186 long-standing questions in the crop sciences ^{20,28}. However, this type of research only rarely
187 investigates large samples or diverse groups of crops ²⁶. Our findings suggest that, when
188 examined across the majority of crop species, priority effects of early crops and climatic origins
189 have made important contributions to the specialization of our modern food system.

190

191 Only around a hundred of the ca. thousand cultivated food plants are grown to a significant
192 extent, with no more than 12 species accounting for most food provision ²⁹. However,
193 determining what makes a crop a good candidate for widespread adoption is elusive and often
194 attributed to contingencies or common knowledge (*e.g.*, ³⁰). We have found that crops from
195 seasonally-dry climates and ancient crops tend to dominate the global area under production.
196 The fact that ancient crops were more likely to become agriculturally relevant suggests priority
197 effects on the adoption and expansion of crop species. Crops adopted early on would have
198 enjoyed a head-start advantage, taking over agricultural niches, such that new opportunities

199 became progressively scarcer. Furthermore, since older crops have been under selection for a
200 longer period, they have had a longer period to adapt to local environments during agricultural
201 range expansion. Thus, for a new crop to be adopted, it probably had to displace species
202 already established and with longer histories of adaptation. These mechanisms might
203 synergize with the fact that plants sharing phylogenetic affinities tend to be functionally alike,
204 such that they occupy similar ecological niches ³¹. Early success of a crop might thus limit the
205 success of other crops in its phylogenetic neighbourhood, supported here by a very modest
206 phylogenetic signal in agricultural relevance, and the vast majority of major crops being
207 surrounded by minor crops in their phylogenetic vicinity (Fig. 6 and Supplementary Data 1).
208 Therefore, priority effects of ancient crops hindered the progressive emergence of
209 agriculturally relevant crops, but did not preclude the advent of new minor crops in their
210 phylogenetic lineage. Further, this indicates that agricultural needs (*i.e.*, ecological niches for
211 crops) have changed little through history ³². Overall, we suggest that low levels of
212 phylogenetic redundancy in our crop system, together with priority effects in the pace, timing
213 and geography of domestication events, have played important roles in agricultural
214 specialization.

215

216 Specialization is not unique to human relations with staple crops, but is widespread in nature
217 ³³. Ecological specialization boosts performance in a small subset of the ancestral niche space,
218 often at the cost of decreased abilities to face environmental change ^{6,34}. Niche narrowing
219 tends to result in reduced geographic ranges and higher susceptibilities to extinction ³⁵.
220 Moreover, natural selection tends to favour the evolution of specialism in stable environments
221 ^{35,36}. Some of those commonalities of specialization apply to the relationship of humans with
222 staple crops. For example, the Holocene has been the most climatically stable period of human
223 evolution ³⁷, which might have permitted the contraction of our dietary niche. Agricultural

224 specialization did not result in a reduction of our overall geographic range, as expected from
225 specialism. However, hunter-gatherers inhabited most terrestrial ecosystems at sparse
226 population densities ³⁷, while agricultural societies have specialized, achieving much higher
227 densities, within regions encompassing a small climate envelope, which the bulk of humans
228 inhabit ¹⁹. Specialization through time is a common theme in other agricultural systems
229 evolved by natural selection, for example in attine ant fungi-farmers. Basal lineages of attines
230 exploit a wide range of fungi, which are only partly domesticated, have wide nutritional niches,
231 and engage in gene flow with their free-living relatives ^{38,39}. But the evolutionary recent attines
232 specialize on a very small set of clones of leucocoprinaceous fungi, which have a narrow
233 nutritional niche compared to free-living fungi ⁴⁰. Evolution of ant farming, spanning ca. 50 My,
234 thus parallels trends in agricultural homogenization between humans and crops, which took
235 place in a remarkably shorter time span.

236

237 Global reliance on a small set of ancient crops from specific geographic regions has brought
238 benefits, ranging from the harnessing of shared cultural and technological expertise ⁴¹, to the
239 sustained availability of grain stocks as back-up for calamities or periods of scarcity ⁴².
240 However, risks associated with specialization are multiple and pressing. Food specialization
241 and trade of economic plants drive the homogenization of the wild biotas of the world and
242 thus contribute to the current biodiversity crisis ⁴³. Further, in times of intensified global
243 change, a species-poor food system is more unstable and less resilient ⁸. The results of this
244 work suggest that a focus on promoting historically recent or new ⁴⁴ crops, crops from diverse
245 climate origins, and crops functionally and phylogenetically redundant with current staples ⁴⁵,
246 might help to avert some of the risks posed by agricultural homogenization. Since
247 specialization and dietary narrowing are labile ecological traits ^{34,46}, implementing shifts in the

248 relative abundances of crops species should be compatible with the fast reaction times needed
249 to tackle global changes.

250 **Materials and Methods**

251 **Data gathering**

252 Data on antiquity of cultivation, type of crop (organ harvested for primary use and growth
253 form), and the identity, distribution and climatic niche of each crop's wild progenitor were
254 retrieved from the live version of the Crop Origins database ^{27,47}, accessed the 1st December
255 2020. Crop Origins is a comprehensive database containing the identity and descriptors of the
256 plant species cultivated for food. The antiquity of each crop (Crop Antiquity, y ago) was taken
257 as its earliest record of cultivation. Crop Origins provided crop antiquity data for 301 crop
258 species. To estimate the climate associated with each crop's origins, the procedures used by
259 Crop Origins can be summarized as follows. First, the identities of the most likely wild
260 progenitors of all crops were searched and retrieved from the literature, and checked for
261 taxonomic accuracy and validity. Then, the geographical occurrences of each wild progenitor
262 were searched in GBIF ⁴⁸ and cleaned using a number of procedures described in ²⁷.
263 Afterwards, current climate (yearly averages for the years 1970-2000, at 2.5 minutes spatial
264 resolution) was taken for the 19 Bioclim variables of WorldClim v 2 ⁴⁹ at each geographic
265 occurrence. To summarize climate data for each crop, all occurrences of the wild progenitors
266 for a given crop were used as replicates, and medians of climate data per crop were computed.
267 Climate data were available for 802 out of the 866 crop species in Crop Origins.

268 Crop species were subset into five types: (1) Woody Fruit: woody crops cultivated for their fruit
269 (mostly) or seed (rarely); (2) Herb Seed: herbaceous crops cultivated for their seed, including
270 grain crops and oil seeds; (3) Root: woody or herbaceous crops cultivated for their
271 belowground organs; (4) Leaf: crops cultivated for their leaf or shoot organs, including
272 sugarcane (the term "vegetables" is used to refer to root and/or leaf crops and was thus
273 discarded); and (5) Herb Fruit: herbaceous crops cultivated for their fruit. If a given crop could
274 be allocated to more than one type based on its use or growth form, it was assigned to the

275 type that contributed most to its agricultural relevance as food. We avoided splitting crops into
276 more categories in order to keep a number of crops per category sufficiently balanced for
277 statistical analyses. A number of supplementary criteria for classifying crops into usage types
278 were considered, including diversity of uses, calorific vs non calorific, and others. However,
279 implementing a combination of those criteria to the whole set of 866 crops proved impossible,
280 thus we retained a simple typology. Finally, a dated phylogenetic tree of the pool of species
281 used in this study was obtained from the live version of Phylo Food ⁴⁷, accessed the 1st
282 December 2020. Phylo Food contained all binomials of the crop species used in this paper.
283 Depending on the statistical analysis Phylo Food was either used as provided, or pruned to
284 sub-groups of species using the drop.tip function of the ape package in R 3.6.2. ⁵⁰.

285 We further compiled information on the current agricultural relevance of each crop. To do this,
286 on the 3rd of October 2018 we queried FAOSTAT production data (area harvested, ha) from
287 2006 to 2016 for all FAO commodity groups ¹. Production data at the regional level were
288 summed up to world totals per year, and averaged for the 2006-2016 period (Global
289 Production – ha –). To match crop species binomials to FAO commodity groups we used ref. ⁵¹.
290 That source used FAO commodity group codes, in conjunction with the FAO Commodity List
291 tool (www.fao.org/economic/ess/ess-standards/commodity), to identify the crop binomials
292 that FAO associates with each group. Prior to matching, the FAO Commodity List was checked
293 for taxonomic accuracy and validity against The Plant List ⁵² using function TPL of the
294 Taxonstand package in R 3.6.2. ⁵³. Commodity groups composed of a single crop species were
295 automatically matched, but the vast majority of groups included several taxa. However, all
296 groups but wheat were composed of a major contributor to agricultural production, together
297 with congeneric minor crops. Therefore, production data were assigned to the major crop of
298 each multi-specific group. For wheat we assigned 93% of production to bread wheat and 7% to
299 durum ⁵⁴. The 112 crop species with a tracked record in FAOSTAT were operationally coded as

300 major crops, and the remaining 754 crop species that were included in the Crop Origins
301 database, but which lacked a FAOSTAT record, were considered as minor crops.

302

303 **Statistical analyses**

304 To address our two questions, we measured phylogenetic signals, and indicators of
305 phylogenetic clustering and overdispersion for the antiquity and global production values of
306 each crop species. Further, we used phylogenetic comparative methods⁵⁵ to test hypotheses
307 about the causes of variation in crop antiquity and of global production, while accounting for
308 phylogenetic relationships.

309 The phylogenetic signal of crop antiquity was calculated using Pagel's λ ⁵⁶. Pagel's λ compares
310 the similarity of the covariances among species trait scores with the covariances expected
311 under Brownian motion evolution (hereafter BM). $\lambda = 0$ indicates an absence of phylogenetic
312 signal, whereas $\lambda = 1$ indicates that close relatives are as similar to each other as would be
313 expected under BM. Lambdas between $0 < \lambda < 1$ indicates phylogenetic signals ranging from
314 very low to near BM expectations. Lambda was computed using the function `phylosig` of the
315 `phytools` package in R 3.6.2.⁵⁷ Global production showed a zero-inflated log-normal
316 distribution (zero for all crops lacking FAOSTAT data), which precludes the use of Pagel's λ .
317 Thus, phylogenetic signal was computed as in¹⁷. In brief, the phylogenetic heritability
318 parameter (λ) of an intercept-only phylogenetic mixed model, which allows family distributions
319 deviating from Gaussian, was taken as the metric of phylogenetic signal⁵⁸. That procedure is
320 analogous to a custom calculation of Lambda, but accounting for zero-inflated binomial
321 distribution of data. Finally, we computed Local Indicators of Phylogenetic Association (LIPA),
322 based on Local Moran's I ⁵⁹ to detect individual crops surrounded by phylogenetic
323 neighbourhoods with similar or distinctive (*i.e.* positive or negative autocorrelation,
324 respectively) global production and crop antiquity values. For each LIPA score, statistical

325 significance was assigned by performing a non-parametric two-sided randomizations test,
326 performed by reshuffling the scores of the tips 999 times. Local Moran's I were computed
327 using the `lipaMoran` function of the `phylosignal` package in R 3.6.2.⁵⁹

328 Before testing hypotheses about the causes of variation in the antiquity of cultivation and
329 agricultural relevance, we conducted two analyses to test and account for collinearity among
330 predictors. First, since WorldClim's descriptors are correlated⁶⁰, a Principal Components
331 Analysis (PCA) of the 19 bioclimatic variables was performed. The PCA was run using crops'
332 median scores for each variate and the function `prcomp` of the `stats` package in R 3.6.2.⁶¹ The
333 19 bioclim variates were scaled and centred prior to analysis. The first two axes of the PCA
334 accounted for 72% of variation in the bioclimatic variables (PCA axis 1 51%, PCA axis 2 21%,
335 Figure S5). The loadings of the 12 bioclimatic variables on PCA axis 1 (Supplementary Data 2)
336 indicated that this axis represents a gradient of increasing seasonality in temperature and
337 decreasing temperature. To facilitate interpretation of results and comparability among both
338 axes, axis 1 was transformed to its inverse ($-$ axis 1) prior to statistical analyses, such that low
339 values were associated with cool, seasonal climates, while high values were associated with
340 hot, aseasonal climates. This axis was therefore named "High T seas \leftrightarrow High T". The loadings
341 of the 12 bioclimatic variables on PCA axis 2 represented a gradient of increasing precipitation
342 and decreasing seasonality in precipitation, such that low values were associated with
343 seasonally dry climates, while high values were associated with wet, aseasonal climates. This
344 axis was thus named "High Ppt seas \leftrightarrow High Ppt". The "High T seas \leftrightarrow High T" and "High Ppt
345 seas \leftrightarrow High Ppt" axes were thus orthogonal and used as descriptors of climatic niche at
346 origin. Second, the set of predictors in the several models was checked for collinearity in the
347 full dataset, and separately for each of the subsets (*e.g.* only the 301 crops with data on crop
348 antiquity). Collinear terms were removed from models until all remaining predictors showed
349 Variance Inflation Factors (VIFs) well below 10, and thus problems associated to collinearity

350 could be discarded. Diagnoses of collinearity were performed using the function `vif` of the `car`
351 package in R 3.6.2.⁶²

352 To address question 1, we used Phylogenetic Generalized Least Squares models (PGLSs) to
353 investigate causes of variation in Crop Antiquity among the 301 crops with such data available
354 ⁶³. Crop antiquity was the response variable, while the fixed-effect predictors were type of
355 crop, and its interactions with “High T seas <-> High T”, and with “High Ppt seas <-> High Ppt”.
356 The main effects of precipitation and temperature climate regimes, and their bi-variate
357 interaction, were not tested because of their high collinearity with other predictors in the
358 model. PGLS models were fitted by restricted log-likelihood using the function `gls` of the `nlme`
359 package in R 3.6.2.⁶⁴. `CorPagel` was specified in the correlation term of `gls`, which took the
360 phylogenetic variance-covariance matrix in the model, weighed by the phylogenetic signal in
361 the residuals of the model. Statistical significance of the predictors was qualified using the
362 `anova.gls` function of the `nlme` package in R 3.6.2.⁶⁴. Observed vs predicted, and residual vs
363 predicted plots were generated to check the validity of models.

364 To address question 2 we tested whether crop origins and descriptors influence 1) the
365 probability that a species is a major or a minor crop, and 2) the variation in global production
366 among major crops. To address probabilities of becoming a major crop, we first used
367 Phylogenetic Logistic Regressions ⁶⁵ which yielded negligible phylogenetic signals in the
368 residuals. Thus, for the sake of simplicity we shifted to binomial general models. Binomial
369 models were conducted separately in the full 866 species dataset, and in the 301 species
370 dataset with data on Crop Antiquity. Major vs minor status was the response variable and type
371 of crop, crop antiquity (only in the 301 species model), “High T seas <-> High T”, and “High Ppt
372 seas <-> High Ppt” were the fixed-effect predictors. Main effects, two- and three-way
373 interactions between predictors were included in the models unless significant collinearity was
374 detected ($VIF > 10$), in which case, the specific term was removed. Binomial models were run

375 specifying the “binomial” family in the glm function of the stats package in R 3.6.2.⁶¹.
376 Statistical significances of each predictor were assessed through log-likelihood ratio tests
377 against a null model using the anova function of the stats package in R 3.6.2.⁶¹. Finally, we
378 tested whether and how the variation in global production among major crops depended on
379 crop origins and typology. A model structure analogous to the binomial model was specified,
380 but model fitting, testing of statistical significance, and model validation were run as described
381 above for the PGLS test used for antiquity of cultivation. Global production data were log₁₀-
382 transformed prior to analyses.

383

384 **Data availability statement**

385 All data used in this paper are publicly available at:

386 https://github.com/rubenmilla/Crop_Origins_Phylo and <http://www.fao.org/faostat/en>.

387

388 **Code availability statement**

389 The analyses carried out in this paper did not require the development of custom code.

390 Functions were run as provided by the R packages mentioned in the Materials and Methods

391 section.

392

393 **References**

- 394 1. FAOSTAT. <http://www.fao.org/faostat/en/#data/QC> (2019).
- 395 2. Mottet, A. *et al.* Livestock: On our plates or eating at our table? A new analysis of the
396 feed/food debate. *Glob. Food Sec.* **14**, 1–8 (2017).
- 397 3. Prescott-Allen, R. & Prescott-Allen, C. How Many Plants Feed the World? *Conserv. Biol.*
398 **4**, 365–374 (1990).
- 399 4. Crittenden, A. N. & Schnorr, S. L. Current views on hunter-gatherer nutrition and the
400 evolution of the human diet. *Am. J. Phys. Anthropol.* **162**, 84–109 (2017).
- 401 5. Khoury, C. K. *et al.* Origins of food crops connect countries worldwide. *Proc. R. Soc. B*
402 *Biol. Sci.* **283**, 20160792 (2016).
- 403 6. Poisot, T., Canard, E., Mouquet, N. & Hochberg, M. E. A comparative study of ecological
404 specialization estimators. *Methods Ecol. Evol.* **3**, 537–544 (2012).
- 405 7. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield Trends Are Insufficient to
406 Double Global Crop Production by 2050. *PLoS One* **8**, (2013).
- 407 8. Khoury, C. K. *et al.* Increasing homogeneity in global food supplies and the implications
408 for food security. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4001–4006 (2014).
- 409 9. Renard, D. & Tilman, D. National food production stabilized by crop diversity. *Nature*
410 **571**, 257–260 (2019).
- 411 10. Newton, A. C., Johnson, S. N. & Gregory, P. J. Implications of climate change for
412 diseases, crop yields and food security. *Euphytica* **179**, 3–18 (2011).
- 413 11. Hawkesworth, S. *et al.* Feeding the world healthily: The challenge of measuring the
414 effects of agriculture on health. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3083–3097 (2010).

- 415 12. Popkin, B. M. Technology, transport, globalization and the nutrition transition food
416 policy. *Food Policy* **31**, 554–569 (2006).
- 417 13. Spengler III, R. N. *Fruit from the Sands: The Silk Road origins of the foods we eat*. (Univ
418 of California Press, 2019).
- 419 14. Vaughan, J. & Geissler, C. *The new Oxford book of food plants*. (OUP Oxford, 2009).
- 420 15. Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication.
421 *Nature* **457**, 843–8 (2009).
- 422 16. Wang, L. *et al.* The interplay of demography and selection during maize domestication
423 and expansion. *Genome Biol.* **18**, 1–13 (2017).
- 424 17. Milla, R., Bastida, J. M., Turcotte, M. M. & Al, E. Phylogenetic patterns and phenotypic
425 profiles of the species of plants and mammals farmed for food. *Nat. Ecol. Evol.* **2**, 1808–
426 1817 (2018).
- 427 18. Ellis, E. C., Klein Goldewijk, K., Siebert, S., Lightman, D. & Ramankutty, N. Anthropogenic
428 transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* **19**, 589–606 (2010).
- 429 19. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J.-C. & Scheffer, M. Future of the human
430 climate niche. *Proc. Natl. Acad. Sci.* **117**, 11350–11355 (2020).
- 431 20. Harlan, J. R. *Crops and man*. (ASA, 1992).
- 432 21. Blumler, M. A. & others. Ecology, evolutionary theory and agricultural origins. in *The*
433 *origins and spread of agriculture and pastoralism in Eurasia* 25–50 (Washington, DC:
434 Smithsonian Institution Press, 1996).
- 435 22. Hancock, J. F. *Plant evolution and the origin of crop species*. (CABI, 2012).
- 436 23. Harlan, J. R. *The living fields: our agricultural heritage*. (Cambridge University Press,
437 1998).

- 438 24. Lombardo, U. *et al.* Early Holocene crop cultivation and landscape modification in
439 Amazonia. *Nature* **581**, 190–193 (2020).
- 440 25. Denham, T. *et al.* The domestication syndrome in vegetatively propagated field crops.
441 *Ann. Bot.* **125**, 581–597 (2020).
- 442 26. Meyer, R. S., DuVal, A. E. & Jensen, H. R. Patterns and processes in crop domestication:
443 an historical review and quantitative analysis of 203 global food crops. *New Phytol.* **196**,
444 29–48 (2012).
- 445 27. Milla, R. Crop Origins and Phylo Food: A database and a phylogenetic tree to stimulate
446 comparative analyses on the origins of food crops. *Glob. Ecol. Biogeogr.* **29**, 606–614
447 (2020).
- 448 28. Larson, G. *et al.* Current perspectives and the future of domestication studies. *Proc.*
449 *Natl. Acad. Sci. U. S. A.* **111**, 6139–46 (2014).
- 450 29. Esquinas-Alcázar, J. Protecting crop genetic diversity for food security: political, ethical
451 and technical challenges. *Nat. Rev. Genet.* **6**, 946–953 (2005).
- 452 30. Clement, C. R. 1492 and the loss of amazonian crop genetic resources. I. The relation
453 between domestication and human population decline. *Econ. Bot.* **53**, 188 (1999).
- 454 31. Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and
455 Community Ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
- 456 32. Tauger, M. B. *Agriculture in world history*. (Routledge, 2013).
- 457 33. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol.*
458 *Syst.* **19**, 207–233 (1988).
- 459 34. Forister, M. L., Dyer, L. A., Singer, M. S., Stireman III, J. O. & Lill, J. T. Revisiting the
460 evolution of ecological specialization, with emphasis on insect--plant interactions.

- 461 *Ecology* **93**, 981–991 (2012).
- 462 35. Colles, A., Liow, L. H. & Prinzing, A. Are specialists at risk under environmental change?
463 Neocological, paleoecological and phylogenetic approaches. *Ecol. Lett.* **12**, 849–863
464 (2009).
- 465 36. McKinney, M. L. & Lockwood, J. L. Biotic homogenization: a few winners replacing many
466 losers in the next mass extinction. *Trends Ecol. Evol.* **14**, 450–453 (1999).
- 467 37. Richerson, P. J., Boyd, R. & Bettinger, R. L. Was agriculture impossible during the
468 Pleistocene but mandatory during the Holocene? A climate change hypothesis. *Am.*
469 *Antiq.* **66**, 387–411 (2001).
- 470 38. Mueller, U. G. & Rabeling, C. A breakthrough innovation in animal evolution. *Proc. Natl.*
471 *Acad. Sci.* **105**, 5287–5288 (2008).
- 472 39. Schultz, T. R. & Brady, S. G. Major evolutionary transitions in ant agriculture. *Proc. Natl.*
473 *Acad. Sci.* **105**, 5435–5440 (2008).
- 474 40. Mueller, U. G., Scott, J. J., Ishak, H. D., Cooper, M. & Rodrigues, A. Monoculture of
475 leafcutter ant gardens. *PLoS One* **5**, e12668 (2010).
- 476 41. Kingsbury, N. *Hybrid, the history and science of plant breeding*. (University of Chicago
477 Press, 2009).
- 478 42. FAO. *Food Outlook – Biannual Report on Global Food Markets: June 2020. Food*
479 *Outlook, 1*. (FAO - Food and Agriculture Organisation of the United Nations., 2020).
- 480 43. van Kleunen, M. *et al.* Economic use of plants is key to their naturalization success. *Nat.*
481 *Commun.* **11**, 3201 (2020).
- 482 44. Li, T. *et al.* Domestication of wild tomato is accelerated by genome editing. *Nat.*
483 *Biotechnol.* **36**, 1160–1163 (2018).

- 484 45. Siddique, K. H. M., Li, X. & Gruber, K. Rediscovering Asia's forgotten crops to fight
485 chronic and hidden hunger. *Nat. Plants* **7**, 116–122 (2021).
- 486 46. Lancaster, L. T. Host use diversification during range shifts shapes global variation in
487 Lepidopteran dietary breadth. *Nat. Ecol. Evol.* 1–7 (2020).
- 488 47. Milla, R. Crop Origins Phylo. *GitHub* https://github.com/rubenmilla/Crop_Origins_Phylo
489 (2020).
- 490 48. The Global Biodiversity Information Facility, G. What is GBIF?
491 <https://www.gbif.org/what-is-gbif> (2018).
- 492 49. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for
493 global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
- 494 50. Paradis, E., Claude, J. & Strimmer, K. {APE}: Analyses of Phylogenetics and Evolution in R
495 language. *Bioinformatics* **20**, 289–290 (2004).
- 496 51. Martin AR; Cadotte, M. W. I. M. M. R. V. C. V. D. Regional and global shifts in crop
497 diversity through the Anthropocene. *PLoS One* **14**, e0209788 (2019).
- 498 52. The Plant List. Version 2. Published on the Internet; <http://www.theplantlist.org/>.
499 <http://www.theplantlist.org/> (2013).
- 500 53. Cayuela, L., la Cerda, Í. G., Albuquerque, F. S. & Golicher, D. J. taxonstand: An r package
501 for species names standardisation in vegetation databases. *Methods Ecol. Evol.* **3**,
502 1078–1083 (2012).
- 503 54. <https://en.wikipedia.org/wiki/Durum>. <https://en.wikipedia.org/wiki/Durum>.
- 504 55. Paradis, E. An Introduction to the Phylogenetic Comparative Method BT - Modern
505 Phylogenetic Comparative Methods and Their Application in Evolutionary Biology:
506 Concepts and Practice. in (ed. Garamszegi, L. Z.) 3–18 (Springer Berlin Heidelberg,

- 507 2014). doi:10.1007/978-3-662-43550-2_1.
- 508 56. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884
509 (1999).
- 510 57. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other
511 things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
- 512 58. de Villemereuil, P. & Nakagawa, S. General Quantitative Genetic Methods for
513 Comparative Biology. in *Modern Phylogenetic Comparative Methods and Their*
514 *Application in Evolutionary Biology* (ed. Garamszegi, L.) 287–304 (Springer-Verlag,
515 2014).
- 516 59. Keck, F., Rimet, F., Bouchez, A. & Franc, A. phylsignal: an R package to measure, test,
517 and explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016).
- 518 60. Bush, S. E. *et al.* Unlocking the black box of feather louse diversity: A molecular
519 phylogeny of the hyper-diverse genus *Brueelia*. *Mol. Phylogenet. Evol.* **94**, 737–751
520 (2016).
- 521 61. R Core Team. R: A Language and Environment for Statistical Computing. (2018).
- 522 62. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, 2019).
- 523 63. Grafen, A. & Hamilton, W. D. The phylogenetic regression. *Philos. Trans. R. Soc. London.*
524 *B, Biol. Sci.* **326**, 119–157 (1989).
- 525 64. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & Team Core, R. nlme: Linear and
526 Nonlinear Mixed Effects Models. (2020).
- 527 65. Ives, A. R. & Garland Jr., T. Phylogenetic Logistic Regression for Binary Dependent
528 Variables. *Syst. Biol.* **59**, 9–26 (2009).
- 529

530

531 **Acknowledgements**

532 This study has been supported by grants CGL2014-56567-R and CGL2017-83855-R (Ministerio
533 de Economía y Competitividad, MINECO, Spain), and REMEDINAL TE (Comunidad de Madrid).
534 Jose María Iriondo, Pablo García-Palacios, Manuel Delgado-Baquerizo, Itay Mayrose and three
535 anonymous reviewers provided valuable comments to earlier versions of this work. Crop icons
536 in Figures 3 and 6 were kindly donated by C. Khoury and C. Navarro-Racines. We thank the
537 GBIF (<https://www.gbif.org/>) and WordClim (<https://worldclim.org/data/bioclim.html>)
538 initiatives for providing geographic and climate data.

539

540 **Authors contributions**

541 R.M. and C.P.O. conceived the study. R.M. analysed data and wrote a first draft of the
542 manuscript. R.M. and C.P.O. contributed to subsequent rounds of writing and gave the
543 approval for submission of the final version.

544

545 **Competing interests**

546 The authors declare no competing financial interests.

547

548 **Figure titles and legends**

549

550 **Figure 1: Global production of food crops included in FAOSTAT**
551 **(<http://www.fao.org/faostat/>)**. Size of each square is proportional to the global area annually
552 harvested per crop, averaged for the period 2006-2016. See Materials and Methods section for
553 procedures used to assign crop species to FAO commodity groups, and for criteria used to
554 cluster species into five types of crops. N = 112 crop species. Throughout this paper the 112
555 crop species with a track record in FAOSTAT were operationally defined as major crops, and
556 the remaining 754 crop species included in Crop Origins database (see Materials and Methods
557 section) but lacking a FAOSTAT record were considered as minor.

558

559 **Figure 2: Predictors of the antiquity of cultivation. A.** Density distribution of crop antiquity
560 (earliest records of cultivation) for each type of crop with raw data on antiquity shown as X-
561 axis ticks. **B and C.** Relationships between temperature and precipitation climate regimes and
562 crop antiquity, shown separately for each type of crop. “High T seas <-> High T” indicates a
563 gradient of decreasing seasonality in temperature and increasing temperature, such that low
564 values were associated with cool, seasonal climates, while high values were associated with
565 hot, aseasonal climates. “High Ppt seas <-> High Ppt” represents a gradient of increasing
566 precipitation and decreasing seasonality in precipitation, such that low values were associated
567 with seasonally dry climates, while high values were associated with wet, aseasonal climates
568 (see Materials and Methods section). Insets in panels of panels 2B and 2C are simple Pearson
569 correlation coefficients plus p-values. See Supplementary Table 1 for effect sizes, statistical
570 significance of terms, and variance explained, for the full statistical model (phylogenetic
571 generalised least squares model). N = 301 crop species. N of individual sub-groups shown in

572 panel 2A. Coloured tick marks in the x-axes of panel 2A are antiquity sample data of each
573 individual crop in each of the five types.

574

575 **Figure 3: Phylogenetic structure of crop antiquity.** Distribution of crop antiquities (earliest
576 records of cultivation or domestication) across the phylogeny of flowering plants. The
577 phylogenetic signal, computed as the magnitude of Pagel's Lambda and its 95% CI, is shown in
578 the centre of the tree. Trait mapping on the internal branches of the tree is shown only to
579 facilitate visualization of phylogenetic clusters with similar or diverse antiquities. N = 301 crop
580 species. Crop icons kindly donated by C. Khoury and C. Navarro-Racines. Several of the icons
581 were published in Khoury, et al. Proc. Biol. Sci., 283 (2016), 20160792 under a CC-BY 4.0
582 licence. The tree can be magnified online to read species identities and to identify their
583 antiquity score.

584

585 **Figure 4: Probability that a crop is major or minor as a function of crop antiquity and climate.**

586 Density distributions of the abundances of major and minor crops (operationally defined here
587 as crops with or without a track record at FAOSTAT, respectively), shown separately for each
588 type of crop and as functions of crop antiquities (**A**, N = 301 crop species), temperature (**B**, N =
589 802 crop species) and precipitation (**C**, N = 802 crop species) regimes. See legend of Figure 2
590 for meaning of "High T seas <-> High T" and "High Ppt seas <-> High Ppt". Statistical analysis
591 fitted two alternative logistic general linear models. See Supplementary Table 2 for details,
592 effect sizes and statistical significance of terms.

593

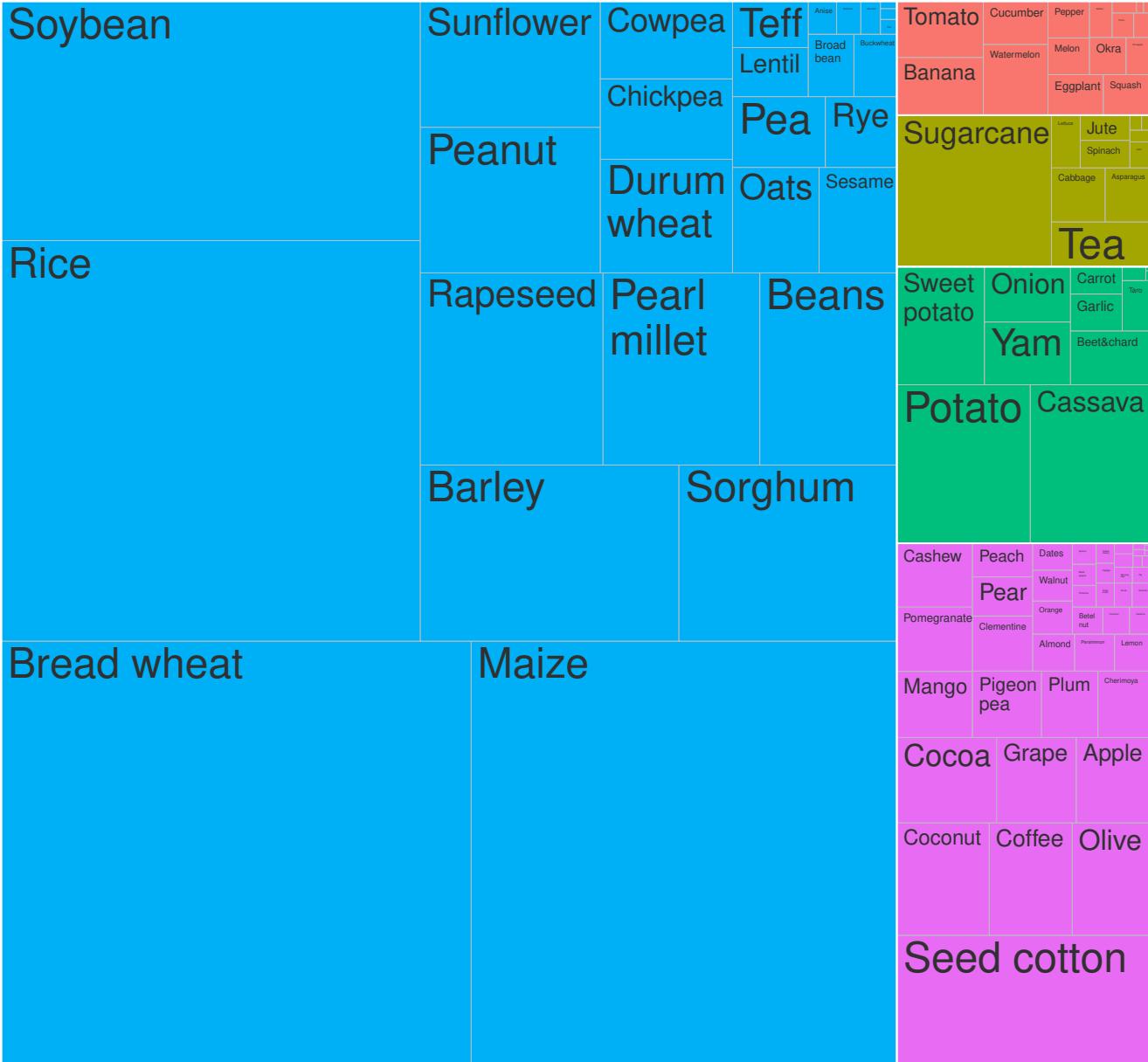
594 **Figure 5: Global production as a function of crop origins and crop type. A.** Global area
595 harvested as a function of crop antiquity. **B** and **C.** global area harvested as functions of

596 temperature and precipitation regimes, respectively, at crops' origins. **D.** Global area
597 harvested as a function of crop type. Global area harvested (ha) is a mean for 2006-2016.
598 Insets and fit lines in **A-C** are raw OLS bivariate regressions between the variates in the plots. In
599 **D**, the central line is the median, box limits are 25-75th percentiles, and whiskers are the 1.5 x
600 interquartile range. See legend of Figure 2 for meaning of "High T seas <-> High T" and "High
601 Ppt seas <-> High Ppt". We also fitted a phylogenetic generalized least squares model to the
602 data, and the details, effect sizes, statistical significance of terms, and variance explained by
603 this model are provided in Supplementary Table 3. N = 105 crops.

604

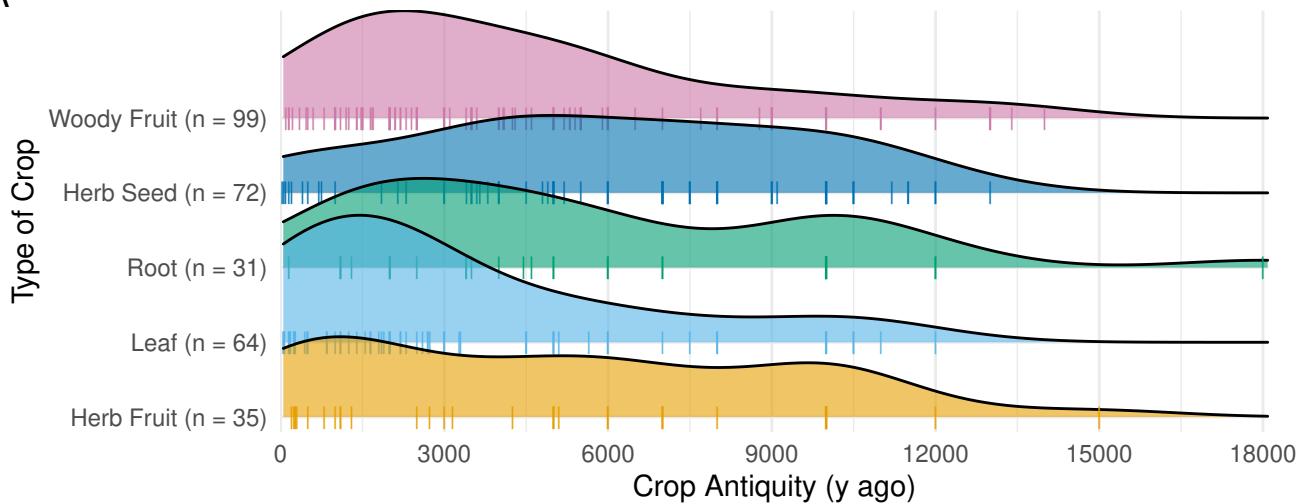
605 **Figure 6: Phylogenetic structure of global production.** Distribution of agricultural relevance of
606 all species included in this paper, measured as the 2006-2016 arithmetic mean of the area
607 harvested in the world for each crop (ha), as reported in FAOSTAT
608 (<http://www.fao.org/faostat/>). The phylogenetic signal, estimated as the magnitude of
609 Lambda and its 95% CI, is shown in the centre of the tree. Trait mapping on the internal
610 branches of the tree is shown only to facilitate visualization of phylogenetic clusters with
611 similar area harvested. The areas harvested are log(+1)-scaled to facilitate visualization. N =
612 866 crop species. Crop icons kindly donated by C. Khoury and C. Navarro-Racines. Several of
613 the icons were published in Khoury, et al. Proc. Biol. Sci., 283 (2016), 20160792 under a CC-BY
614 4.0 licence. The tree can be magnified online to read species identities and to identify their
615 production score.

616

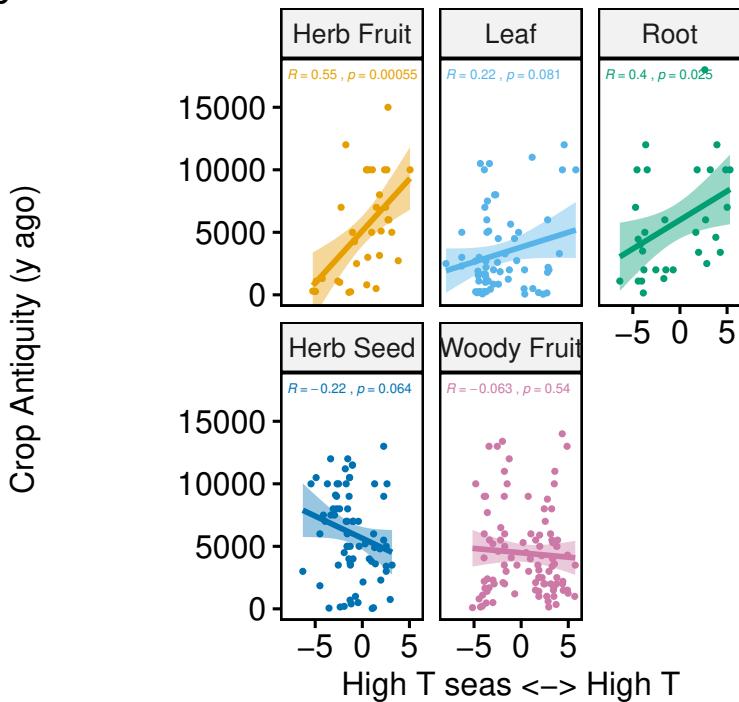


Herb Fruit
 Leaf
 Root
 Herb Seed
 Woody Fruit

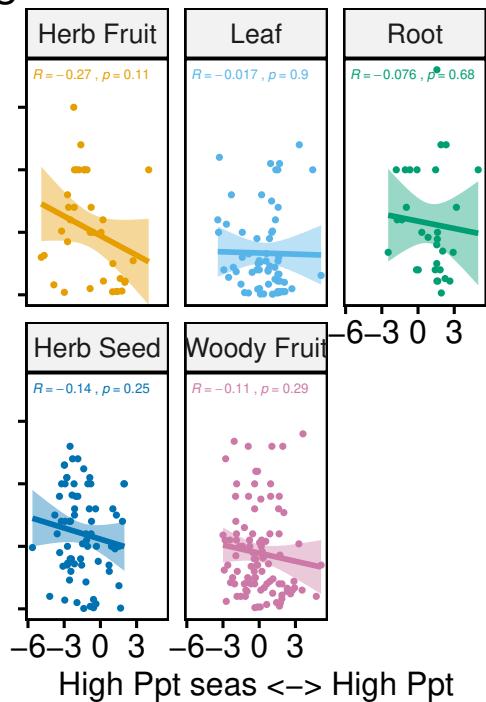
A

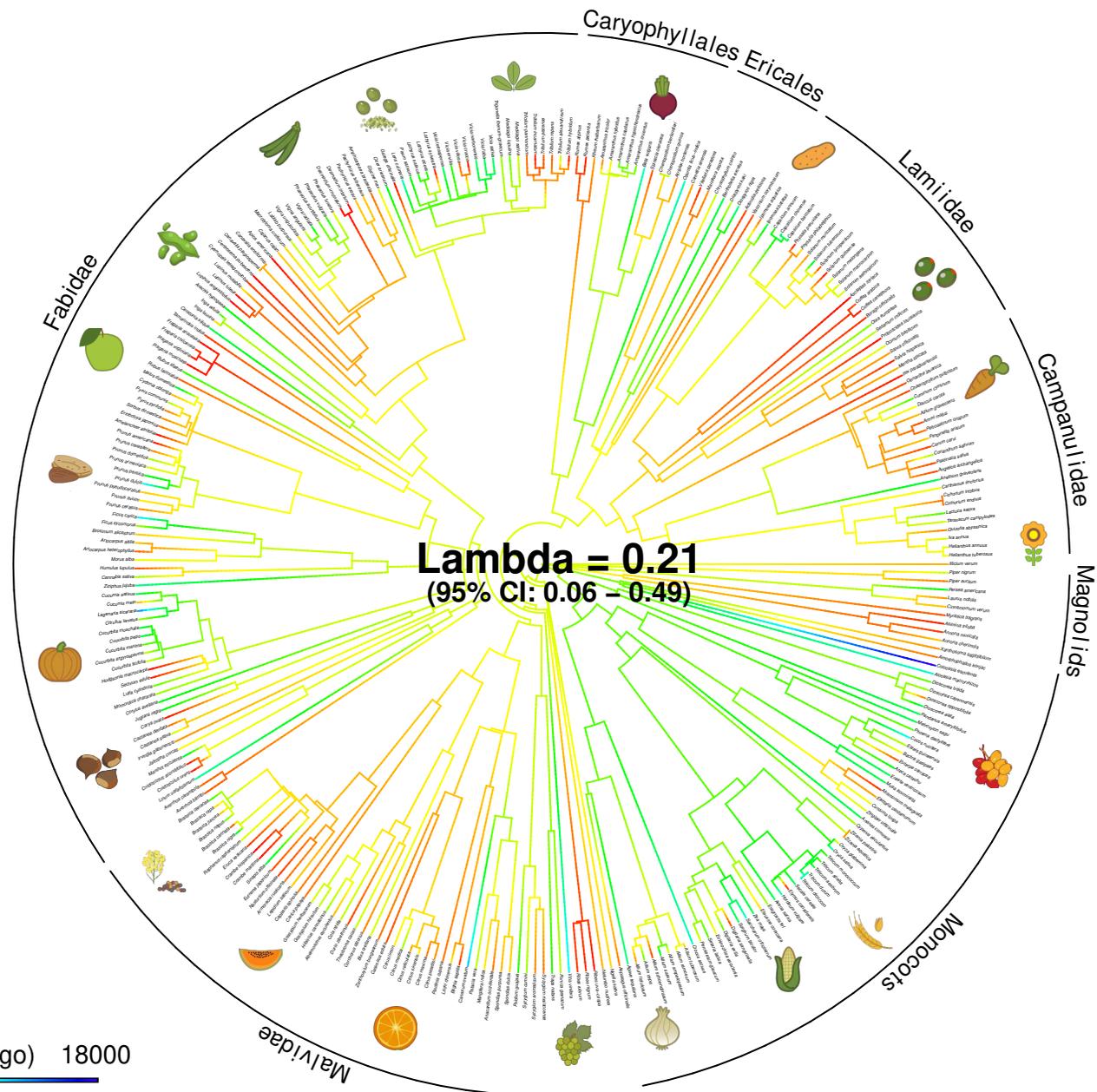


B

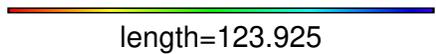


C

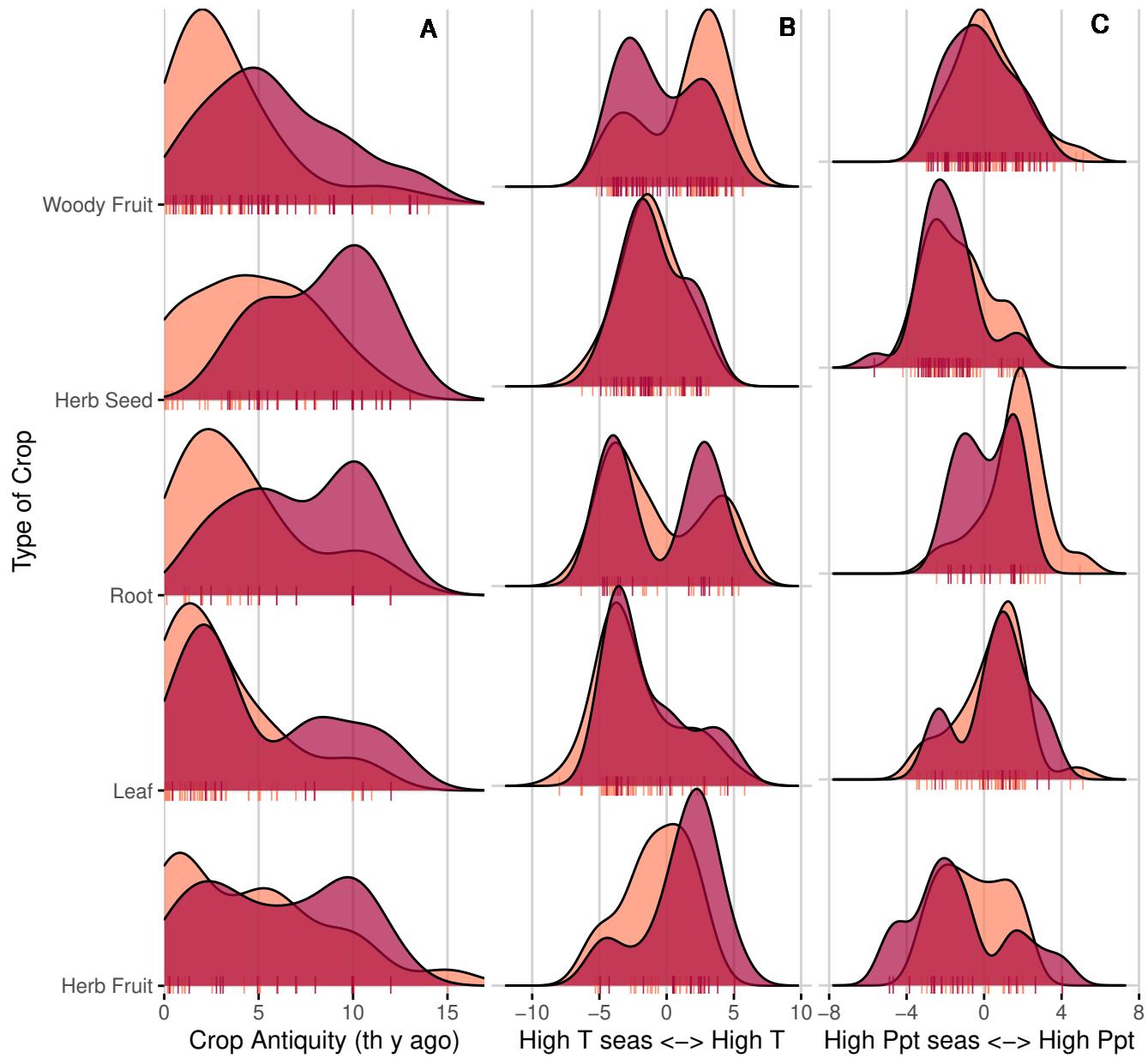


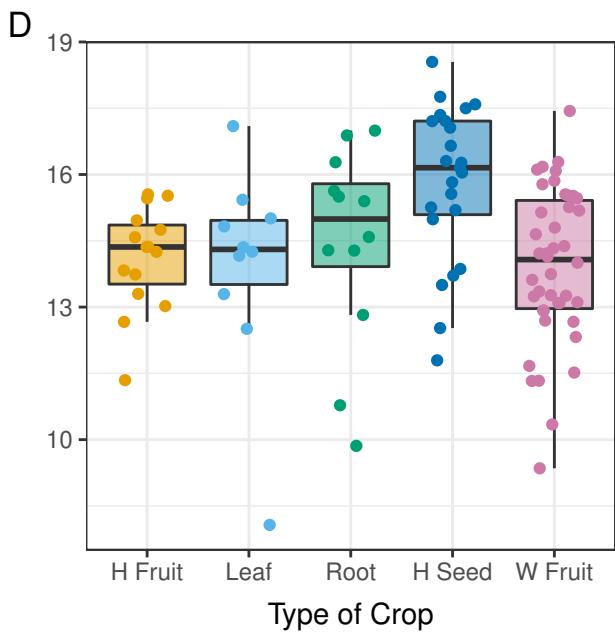
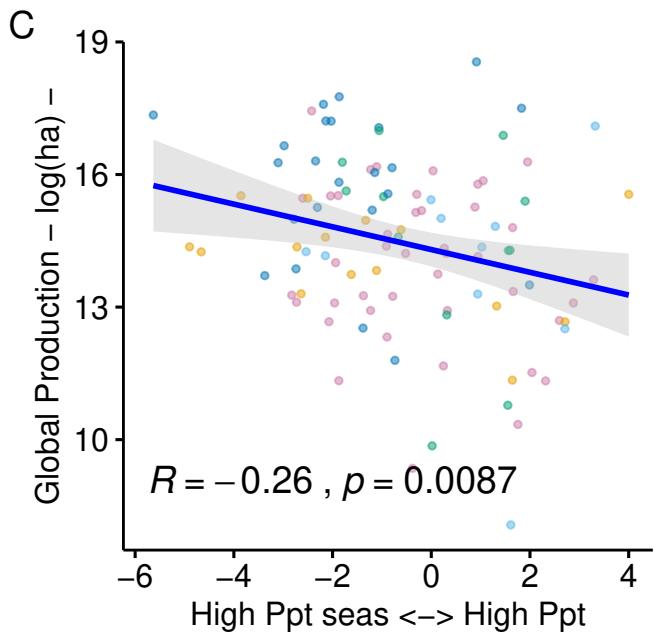
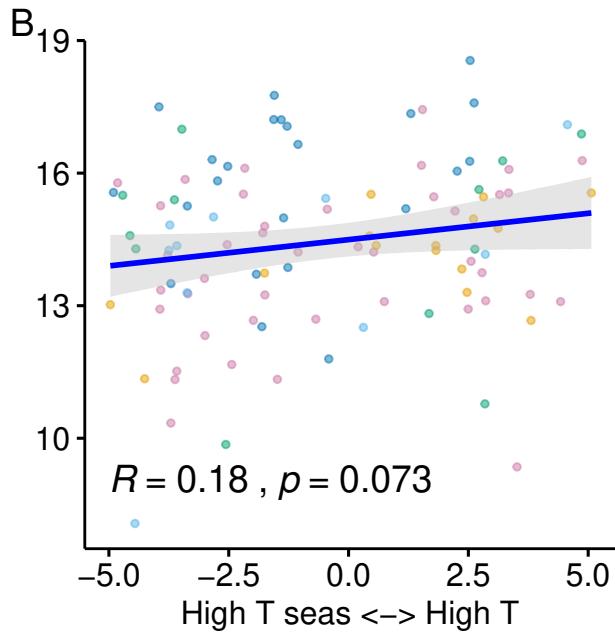
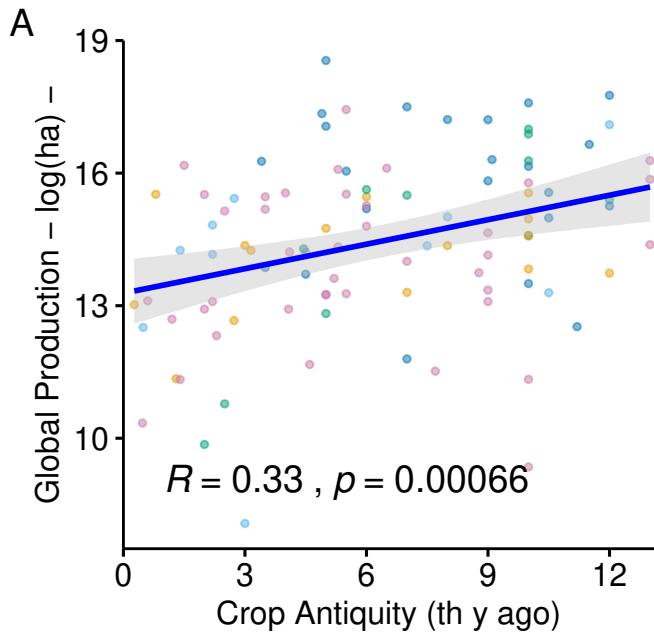


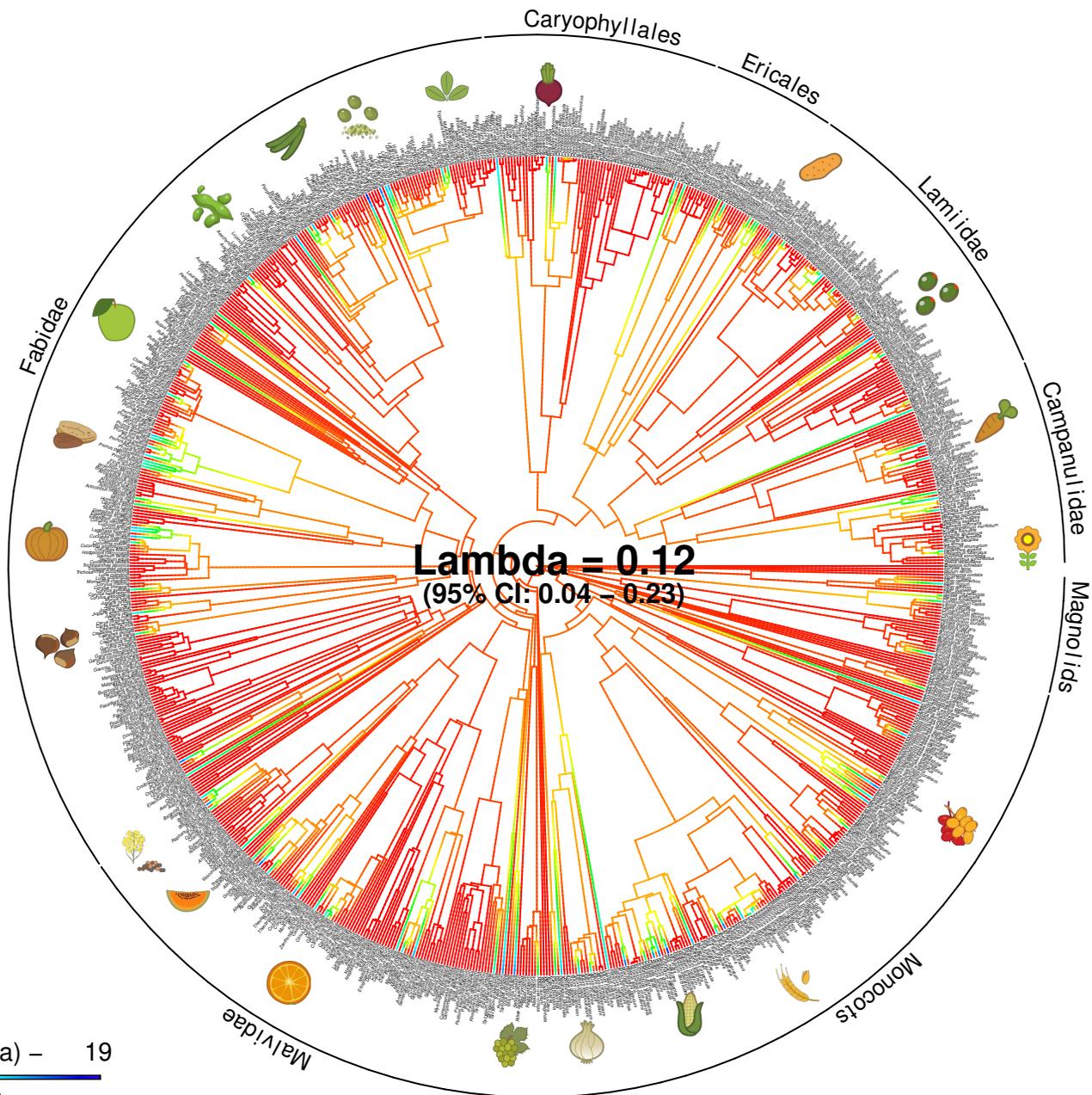
30 Crop Antiquity (y ago) 18000



Minor Major







0 Production - log (ha) - 19
length=125.069