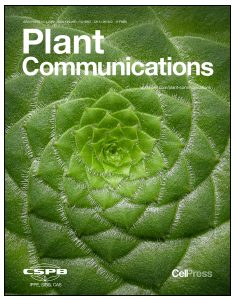
Journal Pre-proof



Genomic variation, environmental adaptation and feralization in ramie, an ancient fiber crop

Zeng-Yuan Wu, Mark A. Chapman, Jie Liu, Richard I. Milne, Ying Zhao, Ya-Huang Luo, Guang-Fu Zhu, Marc W. Cadotte, Ming-Bao Luan, Peng-Zhen Fan, Alex K. Monro, Zhi-Peng Li, Richard T. Corlett, De-Zhu Li

PII: S2590-3462(24)00212-8

DOI: https://doi.org/10.1016/j.xplc.2024.100942

Reference: XPLC 100942

To appear in: *PLANT COMMUNICATIONS*

Received Date: 19 November 2023

Revised Date: 20 December 2023

Accepted Date: 6 May 2024

Please cite this article as: Wu, Z.-Y., Chapman, M.A., Liu, J., Milne, R.I., Zhao, Y., Luo, Y.-H., Zhu, G.-F., Cadotte, M.W., Luan, M.-B., Fan, P.-Z., Monro, A.K., Li, Z.-P., Corlett, R.T., Li, D.-Z., Genomic variation, environmental adaptation and feralization in ramie, an ancient fiber crop, *PLANT COMMUNICATIONS* (2024), doi: https://doi.org/10.1016/j.xplc.2024.100942.

Genomic variation, environmental adaptation and feralization in ramie, an ancient fiber crop

Short title: Domestication and feralization of ramie

Zeng-Yuan Wu[1], Mark A. Chapman[2], Jie Liu[1, 3]*, Richard I. Milne[4], Ying Zhao[1], Ya-Huang Luo[3], Guang-Fu Zhu[3], Marc W. Cadotte[5], Ming-Bao Luan[6]*, Peng-Zhen Fan[1], Alex K. Monro[7], Zhi-Peng Li[3], Richard T. Corlett[7,8], De-Zhu Li[1, 3]*

[1] Germplasm Bank of Wild Species & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[2] School of Biological Sciences, University of Southampton, Southampton SO17 1BJ, UK

[3] CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[4] Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, UK

[5] Department of Biological Sciences, University of Toronto-Scarborough, Toronto, Ontario, Canada

[6] Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, Hunan 410205, China

[7] Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AE, UK

[8] Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China

*Correspondence:

dzl@mail.kib.ac.cn

liujie@mail.kib.ac.cn

luanmingbao@caas.cn

Abstract

Feralization is an important evolutionary process, but the mechanisms behind it remain poorly understood. Here, we use the ancient fiber crop, ramie (*Boehmeria nivea* (L.) Gaudich.) as a model to investigate genomic changes associated with both domestication and fertilization. We first produced a chromosome-scale *de novo* genome assembly of feral ramie and investigated structural variations between feral and domesticated ramie genomes. Next, 915 accessions from 20 countries were gathered, comprising cultivars, major landraces, feral populations and wild progenitor. Based on whole genome resequencing of these accessions, the most comprehensive ramie genomic variation map to date was constructed. Phylogenetic, demographic, and admixture signal detection analyses indicate that feral ramie is of exoferal or exo-endo origin, i.e., descended from hybridization between domesticated ramie and wild progenitor or ancient landraces. Feral ramie has greater genetic diversity than wild or domesticated ramie, and genomic regions affected by natural selection during feralization are different from those under selection during domestication. Ecological analyses showed that feral and domesticated ramie have similar ecological niches which are substantially different from the niche of the wild progenitor, and three environmental variables were associated with habitat-specific adaptation in feral ramie. Our findings advance our understanding of feralization, providing a scientific basis for the excavation of new crop germplasm resources and offering novel insights into the evolution of feralization in nature.

Teaser

We investigated the genomic, morphological and ecological factors underlying feralization of the fiber crop ramie. We elucidated where and when the crop was

domesticated and how it has been distributed and traded. We revealed the genetic and ecological differences between wild, crop, and feral ramie and identify candidate genes that could underlie their divergence. Importantly we determine that the feral ramie is derived from hybridization between domesticated and wild ramie and reveal that feralization involved different genes from domestication, therefore is not a simple reversal of that process.

Introduction

Feralization is the evolutionary process by which domesticated crops or livestock re-acquire some wild-like traits and escape from intensive management to form independent reproducing populations (Wu et al., 2021). Feralization has interested biologists since Darwin (1868), not only because of the implications for evolution but also because feral populations can become invasive and have severe ecological (Ellstrand et al., 2010; Qiu et al., 2017; Wu *et al.*, 2021) or agricultural impacts (Vigueira et al., 2013). On the other hand, feral populations might be significant reservoirs of genetic diversity for crop breeding (Farrant and Hilhorst, 2022; Gutaker et al., 2022; Mabry et al., 2023; Pisias et al., 2022). A better understanding of feral populations at the genetic level might therefore help to both mitigate their impacts as weeds (Qiu et al., 2020) and evaluate them as potential genetic reservoirs (Li et al., 2017). Three pathways to feralization have been recognized (Ellstrand *et al.*, 2010; Pisias *et al.*, 2022). Endoferalization involves spontaneous genetic mutations that influence key traits or selection favoring specific standing genetic variation in an ancestral crop population; exo-endoferalization occurs through natural hybridization between cultivated landraces or varieties with divergent genotypes, leading to novel genotypes that escape into the wild; finally, exoferalization occurs by hybridization or introgression between crops and wild relatives (Martin Cerezo et al., 2023; Wu *et al.*, 2021). The genetic signatures of these three modes can be difficult to distinguish (Zhang et al., 2020) which may contribute to the observation that, despite increasing attention,

the evolutionary mechanisms underlying feralization remain poorly understood (Gering et al., 2019; Mabry et al., 2021a; Wu *et al.*, 2021). Genomic studies have been conducted on grasses, such as weedy rice (Qiu *et al.*, 2017; Wedger et al., 2022), wheat (Guo et al., 2020) and barley (Zeng et al., 2018), but at least 14 feralization events in crops have been suggested (Wu *et al.*, 2021), and only one non-grass crop, *Brassica oleracea*, has so far been investigated at the genomic level (Mabry et al., 2021b).

Climate change is expected to have a strong impact on crop spread and adaptation (Gutaker and Purugganan, 2024; Zsögön et al., 2022), and the feral environment may differ from the ancestral wild range in many ways. Therefore, feralization should not be seen as simply a reversal of domestication, but rather as an adaptation to a new wild environment that applies novel selection pressures, including under a changing climate. Hence investigation of feralization offers opportunities to understand crop adaptation to a changing environment, and thus inform future crop improvements for climate resilience. However, the basis of adaptation and ecological niche range in plants escaping cultivation have yet to be investigated.

Ramie or China grass (*Boehmeria nivea* (L.) Gaudich.), is a subshrub grown for its fibers which are the longest, toughest, and most silky of all known plant fibers, and is an excellent model for studying the evolutionary mechanism of feralization. It was one of the first fiber crops to be domesticated; used since at least 6000 BC in China, where it has long been a symbol of status (Chen, 2007; Liao and Yang, 2016). Today, it is still widely cultivated for textiles and cordage products in tropical and subtropical regions around the world (Sen and Reddy, 2011). However, following the introduction of cotton to China around 1300 AD, many ramie landraces were gradually abandoned by farmers in favor of the new, more easily processed crop, removing the constraints of artificial selection, and permitting feralization. Moreover, the tiny, wind-dispersed seeds of ramie provide ample opportunity for regular escapes from cultivation, and feral populations are now widespread. Feral ramie populations have likely existed in China for centuries or even millennia, but almost nothing is known about their origins and adaptations, or how the plants changed during feralization.

Broad sampling of both wild and cultivated material is needed to understand evolution of feralization (Ellstrand *et al.*, 2010). *Boehmeria nivea* is separated into three morphologically distinct varieties: var. *nivea*, only known from cultivated or naturalized populations, var. *tenacissima* and var. *strigosa* (Zhao et al., 2024), which both occur in apparently natural populations. Previous attempts to understand ramie domestication used limited numbers of molecular markers (Liao et al., 2014; Liu et al., 2009) and narrow population sampling, giving an incomplete picture of the location and timing of domestication. To overcome these shortcomings, here we *de novo* assembled a chromosome-scale genome for a feral ramie accession and then analyzed resequencing data of 915 ramie accessions from 23 countries, covering the wild progenitor, feral populations, major landraces and cultivars. We then combined evidence from morphology, ecology, and genomics to determine the pathway leading to the origin of feral ramie and investigate how adaptation occurred in the feral populations.

Results

Chromosome-level genome of feral ramie and comparative analysis with domesticated ramie

Previous studies of feral plants have predominantly focused on the population genomics of SNPs, and the absence of a framework for studying genomic structural variants (SVs) has hampered progress towards a comprehensive understanding of the evolutionary mechanisms underlying feralization. A high-quality feral ramie genome was assembled (Fig. 1A) from a total of 19.74 Gb of PacBio long reads, with approximately 73-fold high-quality sequence coverage. The contig N50 length was 3.42 Mb, the final scaffold N50 was 21.64 Mb, and the final assembled genome size was 294 Mb (Figs. 1A & S1; Tables S1 & S2), considerably smaller than the estimated genome size of ~380 Mb determined by the k-mer method and flow cytometry (Fig. S2). Accurate genome size estimates are notoriously difficult to achieve for highly repetitive and heterozygous diploid genomes (Helmkampf et al., 2019; Pflug et al., 2020): for example, flow

cytometry may overestimate size due to effects from different plant compounds that affect binding of the stains (Mgwatyu et al., 2020), whereas higher levels of heterozygosity and repetitive sequences may cause inaccurate estimation when using the k-mer method (Pflug *et al.*, 2020). After genome annotation, we obtained 22,312 annotated protein-coding genes, plus 2164 noncoding RNA genes, and determined that more than half (54.85%) of the feral ramie genome was composed of repetitive elements (Table S1). Over 95% of the predicted genes showed homology to genes with known functional annotation in public databases (Table S3) and the BUSCO analysis revealed 1546 out of 1614 (95.8%) complete BUSCOs, 22 (1.4%) of which were duplicated (Table S4). These two results indicate that the newly assembled genome is of high quality and we are confident that our genome is well-assembled.

Aligning our new feral and existing cultivated reference genomes revealed high collinearity (Fig. 1B & S3), plus a considerable number of genomic variants between them (Fig. 1 C-D; Table S5). The distribution of variants was not uniform along the chromosomes. Among all classes of structural variants (SVs) examined, Highly Diverged Regions (HDRs) affected the greatest amount of the feral ramie genome (2780 events, 30.9 Mb), followed by inversions (INV), copy number variants (CNV), translocations, insertions (INS), SNPs, and deletions (DEL) (Fig. 1C, Table S5).

Genome-wide variation and population structure

We sequenced 915 ramie individuals (Figs. 2A & S4), with an average sequencing depth of 31.4× (Table S6). Reads were mapped to the ramie reference genome, with an average mapping rate of 92.2%. Through variant detection and filtering, we identified 8,035,826 high-quality SNPs and 796,139 InDels (Table S7). After filtering (see Methods), 1,260,336 SNPs were retained.

Maximum-likelihood (ML) and neighbor-joining (NJ) approaches produced similar topologies (Fig. 2B; Figs. S5 & S6). Considering the habitats of the individual accessions and the results of the admixture analysis (see below), we separated the ramie accessions into three groups, with Group I (all naturally wild accessions) forming a

monophyletic clade sister to all other accessions. This clade was comprised of three subclades: the first included all accessions of *B. nivea* var. *strigosa* from southern Yunnan, northern Vietnam, and Thailand; the second included all accessions of *B. nivea* var. *strigosa* from southwest Guangxi; and the third included only two accessions, one each from Guangxi and Jiangxi (Fig. S5). These two accessions were morphologically similar to *Archiboehmeria*, a monotypic genus dubiously distinct from *Boehmeria* (Chen, 1980), so we removed them from the subsequent analyses. Group II comprised accessions genetically more similar to domesticated than wild accessions, but with clear admixture in the genome. This group included the bulk of the feral accessions, including all feral accessions from China, plus nine domesticated accessions. Group III comprised all other domesticated accessions examined, plus eleven feral accessions from around the world. Group II was paraphyletic with respect to Group III (Figs. 2B & S5).

Two-dimensional principal component analysis (PCA) based on genomic data clearly separated group I from groups II/III along PC1, with groups II and III largely separated along PC2 (Fig. 2C). These results were concordant with the phylogenetic results and indicate a relatively deep divergence between wild ramie and the others, whereas feral and domesticated ramie grade into one another.

In admixture analysis, the cross-validation error decreased continuously as the number of subpopulations, K, increased, with no clear optimal K (up to $K = 10$; Fig. S7). We therefore discuss only the biologically meaningful groupings of the accessions. At $K = 2$, the wild and domesticated accessions formed groupings distinct from one another, and the feral accessions were mostly admixed with domesticated accessions. At $K = 3$, the wild material was clearly distinct, whereas the feral and domesticated accessions formed groups that graded into one another (Fig. 2B).

Nucleotide diversity ($\theta\pi$) differed between the three groups and was greatest for group II (predominantly feral), slightly lower for group III (predominantly domesticated), and lowest for group I (wild) (Fig. 2D). Genetic differentiation (F_{ST}) was greatest between the wild and domesticated groups, intermediate between the feral and wild groups, and least between the domesticated and feral groups (Fig. 2D).

198 **Demographic and divergence histories**

199 We used a supervised machine learning algorithm (DIYABC Random Forest) (see
200 Materials and Methods) to test different hypotheses concerning the origin of feral ramie.
201 Whether we consider feral ramie as a whole (Table S8) or treat the two largest
202 monophyletic subclades of feral ramie as discrete populations (Table S9), under the best
203 scenarios, feral ramie is shown to be product of hybridization between wild and
204 domesticated ramie (Figs. 3A-B, S8). We describe the results here entirely based on
205 three groups division (Fig. 3A, Table S10). Groups I and III are estimated to have
206 diverged 8,678 years before present (YBP) (95% quantile: 4181-10,800), indicating the
207 initial stages of ramie domestication. Group II is estimated to have originated as a
208 product of admixture between groups I and III (Fig. 3, Table S10), 5095 YBP (95%
209 quantile: 1677-8967), with a smaller portion of the admixture being from group I (the
210 wild group; 0.24; 95% quantile: 0.03-0.88) than group III (the domesticated group;
211 0.76).

212 To infer the demographic history of the three genetic groups and trace potential
213 historical fluctuations in population size, we used two analyses (MSMC2 and SMC++)
214 to examine this over a longer timescale. Both produced similar results (Figs. 3C & S9a),
215 so only those for MSMC2 are described here. The ancestors of the three ramie groups
216 experienced similar, continual increases of effective population size (*Ne*) until 48 ka
217 (thousand years before present) (Fig. 3C). For group I (wild ramie), *Ne* continued to
218 decline from 48 ka to 16 ka, but expanded to a peak at around 5.5 ka, which was
219 followed by a precipitous decline to ca. 4 ka. Group III (the domesticated ramie lineage)
220 experienced a continual reduction of *Ne* starting 48 ka until its lowest point ca. 4.2 ka
221 to 3 ka, which likely corresponds to an associated severe domestication bottleneck.
222 Group II (primarily feral accessions) resembles the wild lineage in having a bottleneck
223 ca. 13 ka to 9 ka, in this case *Ne* then increased considerably at 2.8 ka before a slight
224 reduction at ~1.2 ka (Fig. 3C).

225 We further used '*GONE*' to examine very recent demographic history and obtained
226 very different demographic trajectories for group III (domesticated) and II (feral)

lineages, but a relatively stable trend for group I (wild) populations (Fig. S9B). The population sizes of the domesticated and feral lineages started to decline ~150 generations ago, with the feral lineage exhibiting a very gradual decline, whereas for the domesticated lineage this was sharp and in two steps. This is probably related to the continuous reduction of ramie cultivation over this period, especially in China.

Admixture signal detection

Admixture signal detection analysis for each ramie group detected a strong signal of admixture in group II (feral) arising from both groups I and III (i.e., wild and domesticated). Signals of admixture were not recorded for any other combination of populations (Table S11).

To assess the ancestry of feral ramie compared to the ancestral populations (domesticated and wild ramie), we identified SNPs that were present in one or more accessions of domesticated ramie but not detected in wild ramie (crop-specific private SNPs) or vice versa (wild-specific private SNPs). Among all feral SNPs that matched one of these categories, 90.7% were shared with domesticated material, compared to 9.3% with the wild accessions. This pattern was evident across all 14 chromosomes (Fig. 3D; Table S12), apart from one genome region that had more wild than domesticated SNPs. Thus, both the DIYABC analysis and the admixture analysis support that the feral group was derived through admixture and is genetically more similar to the domesticated group.

Selection associated with domestication and feralization

Signatures of selection were detected in 728 and 605 putative regions within feral and domesticated ramie, respectively (Fig. 3E & 3F; Tables S13 & S14). We further performed GO and KEGG enrichment analysis for the genes in these putative regions. In feral ramie, GO enrichment analysis showed 72 enriched terms, including terms related to metabolic processes, cellular processes, and binding (Table S15), whereas KEGG enrichment analysis identified 17 terms (Table S16). Most of these items have

relationships with stress tolerance. For example, ABC transporters (ko02010) is related to resistance to heavy metal pollution (Wang et al., 2015; Xu et al., 2020). In domesticated ramie, GO enrichment analysis showed 97 enriched terms (Table S17), and 12 significant terms were found in KEGG enrichment analysis (Table S18), most of these terms also related to stress resistance, for example genes involved in Vitamin B6 metabolism (ko00750) may be associated with shade tolerance (Jiang et al., 2023), whereas Benzoxazinoid biosynthesis (ko00402) could be related to cold tolerance in wheat (Li et al., 2023). Regions affected by natural selection during feralization are different from those under selection during domestication (Fig. 3E &3F), and hence that feralization is not a simple reversal of domestication.

Niche differentiation among wild, feral and domesticated ramies

We used several ecological analyses to reveal differences in the niche of each group and to identify candidate ecological factors associated with habitat-specific adaptation during feralization. Empirically observed values for Hellinger's *I* and Schoener's *D* were significantly lower than those expected from pseudoreplicated datasets in paired analyses between Groups I and Group II (wild and feral), and between Groups I and III (wild and domesticated) (Fig. 4 A-B), indicating niche differentiation between these pairs. However, observed values for *I* and *D* were close to 1 between Groups II and III (feral and domesticated), indicating only slight differentiation (Fig. 4C). Niche overlap between the Groups II and III was greatest (D = 0.63), with niches shared between groups accounting for 86.8%, while overlap between Groups I and III was the lowest (D = 0.38), with shared niches accounting for only 37% (Fig. 4 D-F; Table S19). In the PCA analysis, the first two axes explained 43.16% (PCA1: 24.44%; PCA2: 18.72%) of the variation in environmental variables. PCA1 was positively correlated with soil properties (including total nitrogen and organic carbon stocks) and topographic variables (including slope), while PCA2 was correlated with precipitation variables including the precipitation of driest month (bio14) and warmest quarter (bio18), and precipitation seasonality (bio15) (Fig. 4G). ANOVA showed that Group II differed

significantly from Groups I and III in PCA1 and the mean value of PCA2 in Group I was significantly larger than for Group II or III (Table S20). All 12 environmental variables investigated had a statistically significant phylogenetic signal (Table S21), with *K* values less than 1, indicating that closely related populations are more likely to share niches than populations drawn at random.

To identify loci associated with local ecological adaptation in feral ramie, we carried out genome-environment association (GEA) analysis (Grummer et al., 2019; Manel et al., 2018). The result identified 8 regions (at $-\log_{10}(p)>7.83$) significantly associated with 3 of the 12 environmental variables in feral ramie, i.e., mean temperature of wettest quarter (bio16), precipitation of driest month (bio14), and total nitrogen (tn) (Figs. 5A & S10; Table S22). In total, 13 genes were recognized, and the largest number were related to temperature (bio8) (Table S22), e.g., Bnt01G001074 on chromosome 1 was involved in blue light signaling pathway (GO:0009785) and circadian rhythm of plant (KEGG: ko04712), which is proposed to be associated with temperature adaptation (Ben Michael et al., 2020). Other examples include Bnt12G017285 on chromosome 12 with transmembrane transporter activity (GO: GO:0022857), which is thought to be related to drought stress tolerance in maize (Jiao et al., 2022) and Bnt04G005975 with ATP binding activity (GO:0005524), which is involved in low nitrogen (Borah et al., 2018).

Potential geographic distribution and ecological drivers of feral ramie

To predict changes in the areas potentially suitable for feral ramie under past and future climate change, we carried out ecological niche modeling (ENM). Results showed both wild and feral ramies had an area under the receiver operating characteristic curve (AUC) value of ≥0.9 (Table S23), indicating a better than random prediction. The suitable area was greatly influenced by climate change (Fig. 5 B-E). The potential suitable area for wild ramie was greater in the Last Interglacial (LIG) than the Last Glacial Maximum (LGM) and the present, and is predicted to increase in the future (2090). The area suitable for feral ramie is predicted to remain stable to 2090 (Fig. 5 F-

G).

Discussion

All samples of *B. nivea* var. *strigosa* formed a well-supported, monophyletic group, clearly distinct from both the feral and domesticated accessions. This strongly suggests that var. *strigosa* is either the direct progenitor of domesticated ramie, or at least a close relative of the wild progenitor if that is now extinct. *B. nivea* var. *strigosa* is distributed in southern Yunnan, southwest Guangxi, and the Indo-Chinese Peninsula. These are all places where ramie is cultivated, so it seems likely that it was domesticated within this native range, although we were unable to sample all reported wild populations, and some may have become extinct during the agricultural expansion over the last few millennia (He et al., 2023; Xie et al., 2021). This might explain why wild ramie has lower genetic diversity that the other two groups. Our data shows that wild ramie is genetically distinct from feral and domestic ramies, and therefore is likely to possess novel genetic diversity that could be useful in future breeding.

Feral and domesticated ramie together form a monophyletic group (Groups II+III) (Figs. 2B & S5). Most feral accessions fall into Group II and comprise a phylogenetic grade, with most cultivated accessions forming a single derived lineage. Accessions identified morphologically as *B. nivea* var. *nivea* exist among both cultivated and naturalized feral material (Fig. S5). *B. nivea* var. *tenacissima* has previously been suggested as the original wild form of ramie (Chen et al., 2003), but our results indicate that individuals with this morphology are feral and derived from, and not the ancestor of, domesticated ramie (Fig. S5).

If we assume that feral populations generally occur close to where they originated, then this allows us to infer the origin and subsequent routes of spread for cultivated ramie across the world. Following this, it appears that basal populations in Group II (Fig. S5), which are mainly from Jiangxi, Guangdong, and Guangxi provinces, and form a subgroup at K=5 (Figs. 2B & S7), may represent the earliest ramie feralization

events. This suggests that these are the first places where ramie was cultivated, and the likely region of its domestication, largely consistent with a previous study, based on nuclear SSR marker analysis, suggesting that ramie domestication began in the Yangtze River Valley of China (Liao *et al.*, 2014). Southern China is an important hotspot of domestication for several crop species, including rice, apricot, and peach (Groppi et al., 2021; Larson et al., 2014; Li et al., 2019), and our results further highlight the importance of this region for crop domestication. Starting from Jiangxi, the putative cradle of domestication, ramie cultivars followed a predominantly westward pattern of dispersal within China to Hunan (which contains the basal individuals within the mainly cultivated Group III), and from there across the rest of China, especially along the Yangtze River Valley (e.g., Chongqing, Zhejiang) and Fujian in southeastern China (Fig. S5).

All feral accessions from Japan and Korea were grouped with feral populations from Zhejiang, Anhui and northern Jiangxi (Fig. S5), suggesting that these locations were the source of the Japanese and Korean accessions, likely driven by human migration and maritime trade. This disjunctive grouping contains no domesticated material, suggesting that these individuals might be all that remains of lineages no longer in cultivation. In Japan and the Philippines, concentrated efforts were made to produce ramie during the Second World War so there was probably a large amount of recent trade between these regions around that time (Roy and Lutfar, 2012). Taiwanese indigenous people have used ramie fiber for thousands of years until the period of Japanese colonial rule (1895−1945), when the availability of other types of clothing caused ramie cultivation there to gradually peter out (Taru and Watan, 2020). Considering the sister grouping of an accession from the Philippines (W531) with ones from Taiwan (Fig. S5), ramie material now in Taiwan most probably originated from the Philippines. African ramie accessions were closely related to Chinese cultivated material, and two accessions from the USA (B344 and W160) were nested among the cultivated individuals of Guangdong and Jiangxi. Despite the fiber's use for a wide variety of products, it was little known in North American markets or widely traded

until the 1980s (Hester and Yuen, 1989), but our data indicate at least two introductions of Chinese material into the USA.

Feral organisms usually revert to the wild-like morphology of their ancestors, and such restoration of ancestral phenotypes can involve novel genetic mechanisms (Dwivedi et al., 2023; Thurber et al., 2010). Feral ramie contains accessions referable to both var. *tenacissima* and var. *nivea*, but the var. *tenacissima* accessions are closer to the base of the phylogeny (Fig. S5). *B. nivea* var. *tenacissima* shares with var. *strigosa* a branched stem, a partly connate stipule, and mostly green abaxial surfaces of the leaf blades (Fig. 6), so these characteristics in var. *tenacissima* might be atavistic and/or due to crossing with var. *strigosa*. Other features, for example an assurgent or appressed strigose stem, differentiate var. *tenacissima* from var. *strigosa*. Most accessions in Group II identified as var. *nivea* are more similar to the domesticated accessions and appear to contain a smaller proportion of the wild genome.

Nucleotide diversity ($\theta\pi$) was greater in feral than domesticated ramie, whereas, in contrast, some feral populations of both corn and rice were found to have lower genetic diversity than crop populations (Qiu *et al.*, 2017; Vigueira *et al.*, 2013). Our observation is best explained by feral ramie populations expanding their gene pools via hybridization from wild material and/or landraces. Admixture signal detection showed a strong signal for admixture of wild and domesticated populations, and DIYABC Random Forest analysis showed that hybridization between wild and domesticated ramie gave rise to feral ramie (Fig. 3).

However, there is little overlap between the geographical ranges of var. *strigosa* and feral material (Fig. S4), indicating that gene flow from the former into the latter is unlikely. One possible explanation for the observed admixture is that var. *strigosa* was previously more widespread. Given the geographical and climatic differences between the ranges of the varieties (Fig. 4), this seems unlikely. Alternatively, gene flow may have come from now extinct (or undetected) landraces, derived independently from var. *strigosa* and closer to it genetically than existing var. *nivea*.

Together, our findings support the idea that feral ramie resulted from hybridization

between domesticated ramie and wild progenitor or landrace material, probably growing in close proximity on the edges of farms, so feral ramie is most probably of exoferal or exo-endoferal origin (Wu *et al.*, 2021). Crucially, feral ramie may contain genetic diversity which may be of use in ramie breeding going forward.

Demographic analysis reveals that the ancestors of wild, feral, and domesticated ramie lineages all exhibited a parallel reduction in *Ne* from 48–16 ka, with the recent end roughly coinciding with the LGM (19-26.5 ka) (Clark et al., 2009). The prolonged decrease in *Ne* of domesticated material may have resulted from a protracted period of low-intensity cultivation and/or management before full domestication 4.2 ka, similar to the situation in grapes (Li *et al.*, 2017) and African rice (Meyer et al., 2016), and some archaeological evidence suggests that humans had already used fibers from ramie at least 30,000 years ago (Kvavadze et al., 2009). More recent changes in the population dynamics of the ancestors of domesticated ramies might, in turn, have been linked to human expansion as the Holocene (11.7 ka) began. The timing of a recent bottleneck in wild ramie, from 4 ka onwards, is consistent with anthropogenic destruction of its habitat (Xiao et al., 2018; Xie *et al.*, 2021). The dramatic reduction in *Ne* for domesticated ramie ~4.2 ka to 3 ka likely represents a domestication bottleneck.

Crop domestication was realized through niche construction (Purugganan, 2022), but little is known about niche change and ecological adaptation of feral plants after they return to the natural environment (Gering *et al.*, 2019). All niche differentiation analyses (Fig. 4) indicated that the niche of feral ramie is substantially different from that of wild ramie, but similar to that of domesticated ramie. Temperature and precipitation-related variables and total nitrogen in the soil were identified as candidate ecological factors associated with habitat-specific adaptation in feral ramie.

Investigations identifying loci involved in domestication and their significance for feralization have been carried out in many animal taxa, but this is limited in plants. Genome scans have become routine and offer potential to investigate adaptive variation (Grummer *et al.*, 2019). We found the feral and domesticated genomes to be largely collinear. Small SVs were mostly located in intergenic regions or introns. Further work

could identify whether any of these SVs demonstrated fixed differences between wild, feral and/or domesticated populations. Selective sweeps analysis revealed that the genomic regions targeted by the domestication and the feralization processes were largely non-overlapping, suggesting that feralization is determined by novel genetic mechanisms, distinct from those involved in domestication.

In short, in this study, the largest genomic resource for ramie to date has been generated and explored, unveiling the domestication and feralization history, and the genetic basis of environmental adaptation for feral ramie. Our results not only support that feral ramie can be a source material for improving current domesticates and even *de novo* domestication (Yu and Li, 2022), but also provide many important scientific insights into the feralization process. However, feralization is a complex biological process, so more work is needed that examines the molecular genetic basis of fitness-related phenotypes in feral settings, and the universality of the evolutionary mechanisms during feralization needs to be examined in more plants.

Materials and Methods

Sample collection

A total of 915 ramie accessions were sampled from 23 ramie-producing countries across Asia, Europe, Africa, and the Americas. China, where the cultivation history is most ancient, was extensively sampled from all 19 provinces or autonomous regions where ramie is currently cultivated. Our sampling covered all major ramie production areas and the full spectrum of wild, feral, and domesticated (including landrace and cultivar) material, so all three varieties of *B. nivea* (vars. *nivea*, *tenacissima*, and *strigosa*) were comprehensively sampled (Figs. 2A & S4). Among sampled material, the term ‘wild’ is used exclusively to refer to wild progenitor that appears to have no history of domestication, and the term ‘feral’ to refer to plants that have escaped cultivation and evolved independently, typically adapting to their local environments (Ellstrand *et al.*, 2010; Pisias *et al.*, 2022). Because feralization can occur at both landrace and cultivar

453 stages (Wu *et al.*, 2021), all ramie referable to var. *tenacissima* or var. *nivea* growing in
454 the wild without human control are considered as feral in our study. Moreover, the term
455 "landrace" encompasses a range of different concepts that have varied over time
456 (Casañas et al., 2017); our study followed the landrace definition of Villa et al. (2005),
457 Zeven (1998), and Dwivedi et al. (2016), i.e., as a dynamic population of a cultivated
458 species that has a historic origin and distinct identity and lacks formal crop
459 improvement, as well as often being genetically diverse, locally adapted and associated
460 with traditional farming systems or a low input agriculture system. Because most
461 farmers in China have given up growing ramie (see Introduction), there is hardly any
462 domesticated ramie in the farms, samples of all cultivars and most landraces were
463 acquired from National Infrastructure for Bast Fiber Crop Germplasm Resources of
464 China (Table S6).

465 No feral ramie genome has yet been reported to date, although three whole
466 genomes of cultivated ramie have been reported (Chen et al., 2023; Wang et al., 2021).
467 In this study, we collected for this purpose fresh material from a feral adult (lab No. is
468 HZS10, Table S6) in Shennong Valley National Forest Park, Hunan Province, China (N
469 26.503°, E 114.001°). Living collections and seeds of this individual are preserved in
470 the Germplasm Bank of Wild Species, Kunming Institute of Botany, CAS.

471 **Genomic DNA extraction and sequencing**

472 Genomic DNA was extracted from the leaves of feral ramie HZS10 using a modified
473 CTAB method. The quality of the extracted DNA was examined using a NanoDrop
474 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), and its
475 quantity determined by electrophoresis on a 0.8% agarose gel. Illumina sequencing
476 libraries were generated using the VAHTS Universal DNA Library Prep Kit for MGI
477 (Vazyme, Nanjing, China) following the manufacturer's recommendations, and index
478 codes were added to attribute sequences to each sample. The library was quantified
479 using a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and
480 Bioanalyzer 2100 (Agilent Technologies, CA, USA). Finally, the MGI-SEQ 2000

platform was used to generate paired-end sequencing data, which generated a total of 12.7 Gb. To construct sequencing libraries for PacBio sequencing, genomic DNA was fragmented into ~15 kb fragments by g-TUBE, then end-repaired, with adapters ligated and digested with exonuclease as recommended by Pacific Biosciences. The SMRTbell library was constructed using the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences). Library size and quantity were assessed using the FEMTO Pulse and the Qubit dsDNA HS reagents Assay kit, and DNA libraries were sequenced on the PacBio Sequel II platform (Pacific Biosciences), generating a total of 19.74 Gb of PacBio long read data. A Hi-C library was constructed and sequenced on an MGI-SEQ 2000 platform for chromosome-level scaffolding, generating a total of 156.73 million paired-end reads and 46.33 Gb of sequencing data.

To aid genome annotation, we generated RNA-seq data for four different tissues, i.e., root, stem, leaf, and flowers from the same individual. All fresh tissues were frozen in liquid nitrogen and stored at –80 ℃ before processing. Paired-end RNA libraries were constructed using the VAHTS Universal V6 RNA-seq Library Kit for MGI (Vazyme, Nanjing, China) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The quantification and size of libraries were measured using Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Sequencing was performed on an MGI-SEQ 2000 platform.

Genome *de novo* assembly and annotation

To estimate the genome size of individual HZS10, the Illumina short raw reads were pre-processed to remove the adaptors and low-quality bases using SOAPnuke (Chen et al., 2018b) with default settings, and the clean data were recruited to determine the k-mer distributions using the GCE software (Liu et al., 2013). Genome size was also estimated by flow cytometry using tomato as an internal standard. The PacBio long-read data were *de novo* assembled into contigs using Hifiasm (Cheng et al., 2021). The 12.7 Gb (~47× coverage) of Illumina pair-end short reads were used to further correct

systematic errors in the PacBio contigs using Pilon (Walker et al., 2014). Subsequently, to anchor the corrected contigs into chromosomes, we aligned the Hi-C sequencing data into these contigs using Juicer (Durand et al., 2016) and the contigs were finally linked into 14 chromosomes by 3D-DNA (Dudchenko et al., 2017). The completeness and accuracy of genome assembly were quantitatively assessed using BUSCO (Simão et al., 2015) and the eudicotyledons_odb10 gene set.

For annotation of repetitive sequences, two methods were employed to identify repeats in the feral ramie genome. First, we used homology-based analysis, in which known TEs were identified using RepeatMasker (version 4.0.9) (Chen, 2004), and the results were compiled into the Repbase TE library (Jurka et al., 2005). RepeatProteinMask searches were also conducted using the TE protein database as a query library. Second, we used *de novo* prediction, i.e., a de novo repeat library of the feral ramie genome was constructed using RepeatModeler, which can automatically execute two core de novo repeat-finding programs, namely RECON (version 1.08) (Bao and Eddy, 2002) and RepeatScout (version 1.0.5) (Price et al., 2005). Furthermore, we performed a de novo search for long terminal repeat (LTR) retrotransposons using LTR_FINDER (version 1.0.7) (Xu and Wang, 2007) and identified tandem repeats using the Tandem Repeat Finder (TRF) package (Benson, 1999). Finally, we merged the library files of the two methods and used Repeatmaker (Chen, 2004) to identify all repeats.

Protein-coding genes were predicted by three methods, which were *ab initio*, homology-based and RNA-Seq-aided gene prediction. For *ab initio* prediction, we used the gene predictor softwares Augustus (version 3.3.1) (Stanke et al., 2006) and Genescan (Burge and Karlin, 1997). Models used for each gene predictor were trained from a set of high-quality proteins generated from the RNA-Seq dataset. Homology-based gene prediction was conducted using Exonerate (version 2.2.0) with default parameters (Slater and Birney, 2005). For RNA-Seq-aided gene prediction, we first removed low quality reads and bases using SOAPnuke (Chen *et al.*, 2018b), and then assembled clean RNA-Seq reads into transcripts using Trinity (Grabherr et al., 2011),

following which gene structure was defined using PASA (Haas et al., 2003). Finally, Maker (version 3.0) (Cantarel et al., 2008) was used to integrate the results of all three methods. The output included a set of consistent and non-overlapping sequence assemblies, which were used to describe the gene structures.

For the annotation of non-coding RNAs (rRNA, small nuclear RNA, and microRNAs), we used RNAmmer (version 1.2) (Lagesen et al., 2007) and Infernal (version 1.1.2) (Nawrocki and Eddy, 2013) by searching the Rfam database (version 14.1) (Kalvari et al., 2018) with default parameters. We used tRNAscan-SE (version 1.3.1) (Lowe and Eddy, 1997) with default parameters to identify the genes associated with tRNA.

For functional annotation of protein-coding genes, BLASTP was used to align the feral ramie protein sequences with those on public databases including NCBI, NR, TrEMBL, InterPro, Swiss-Prot, and KEGG database, with an E-value threshold of 1E-5. Motifs, and domains were annotated using PfamScan (Mistry et al., 2007) and InterProScan (Jones et al., 2014). Motifs and domains within gene models were identified by PFAM databases. GO IDs for each gene were obtained from Blast2GO (Conesa and Götz, 2008).

Synteny analysis and comparative genomics

To determine the pairwise similarity of protein sequences between feral and domestic ramie genomes (Wang *et al.*, 2021), gene synteny analysis was performed using the JCVI package (Tang et al., 2015).

To identify structural variants (SVs) between the feral and domesticated assemblies, comparative genomics analysis was performed. The contigs of the feral *de novo* assembly were ordered along a chromosome-level reference genome of cultivated ramie (Zhongsizhu 1) (Wang *et al.*, 2021) using Minimap2 (Li, 2018) with parameter setting "-ax asm20 -eqx". SyRI (Goel et al., 2019) (-k -F S) was used to identify structural rearrangements and local variants between two genomes. All these variants were annotated using the SnpEff program (Cingolani et al., 2012) with parameter -ud

2000, and a dot plot was drawn using the software plotsr (Goel and Schneeberger, 2022) with parameters -m 20000 -x -q 500000 -s -t.

Variant calling and filtering

Genome resequencing was carried out for 915 ramie accessions (Table S6) and an outgroup using the same methods as above, but using the Illumina NovoSeq platform. Raw data were subjected to a quality check and then filtered by fastp (version 0.20.0) (Chen et al., 2018a). Clean paired-end reads of each accession were then mapped to the latest reference genome of domesticated ramie (Qingyezhuma) (Wang *et al.*, 2021) using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010) with default parameters. After alignment, Picard (version 2.18.17, http://broadinstitute.github.io/picard/) was employed to mark duplicate reads, and SAMtools (Li et al., 2009) was employed to convert alignment format.

To analyze population genetics, we focused on SNPs and small indels (1–10 bp). GATK (version 3.8.1) (McKenna et al., 2010) was used for calling and filtering whole-genome variants (SNPs and InDels). SNPs were filtered with the following parameters: QD<2.0, MQ<40.0, FS>60.0, SOR>3.0, MQRank- Sum<− 12.5, ReadPosRankSum<− 8.0, and indels filtered with the parameters QD<2.0, FS>200.0, MQ<40.0, SOR>10.0, ReadPosRankSum<− 20.0. From this we defined a core SNP set by removing SNPs with more than two alleles and >20% missing calls. Heterozygous sites were also filtered to retain SNPs with minor allele frequency (MAF) greater than 1%. All variants were annotated using Annovar (Wang et al., 2010).

Population structure and phylogenetic analyses

Before inferring the population structure, PLINK (Purcell et al., 2007) was used to filter out SNPs that were in linkage disequilibrium with the parameters *indep-pairwise 50 5 0.5*. In total we retained 1,260,336 SNPs, and then ADMIXTURE (Alexander et al., 2009) was employed to infer the optimum number of clusters (*K*) among all ramie accessions. *K* values from two to ten were examined, and the cross-validation error was

593 calculated to identify the most likely number of clusters. A principal component
594 analysis (PCA) was performed using EIGENSOFT (Price et al., 2006). To infer
595 relationships among accessions, two kinds of rooted phylogenetic trees were
596 reconstructed. First, using the same 1,260,336 SNPs, a NJ phylogenetic tree was
597 obtained by calculating the pairwise genetic distances using PLINK (Purcell *et al.*,
598 2007), and the tree was constructed using PHYLIP (Retief, 2000). Second, an ML tree
599 was constructed based on fourfold-degenerate sites in the 915 ramie accessions. SNPs
600 were extracted and compared to the 7,460,735 fourfold degenerate sites identified in
601 the ramie genome using iTools (20180520) (Dinov et al., 2008). SNPs from each
602 individual were merged into one file using mafft (version 7.407) (Katoh and Standley,
603 2013) followed by trimming low quality regions with trimAl (version 1.4.rev22)
604 (Capella-Gutiérrez et al., 2009). The 120,201 SNPs were then used to construct
605 a rooted maximum likelihood tree using IQ-TREE (version 1.6.12) (Nguyen et al.,
606 2015) with the parameters -alrt 1000-bb 1000 (ultrafast bootstrap). *Girardinia*
607 *diversifolia* (sample ID is W1000) was used as outgroup.

608 Based on population structure and each individual's habitat (see results), we
609 defined three groups of individuals. Group I included only wild individuals and was
610 distinct from all feral and domesticated material. Group II comprised all but 11 feral
611 individuals plus nine domesticated accessions; this group was genetically similar to
612 domesticated material, but with apparently admixed genomic composition. Group III
613 comprised the vast majority of cultivated landraces and modern cultivar accessions
614 from the National Infrastructure for Bast Fiber Crop Germplasm Resources of China,
615 plus eleven feral individuals from around the world. Overall, our dataset comprised 552
616 group III accessions (primarily domesticated), 286 group II accessions (primarily feral)
617 and 77 group I accessions (all wild).
618

Diversity statistics estimation, population demography, and inference of selective sweeps

To more accurately estimate diversity and divergence statistics and demography, we assigned an individual to a cluster if it had an estimated posterior probability > 0.80 to that cluster at K = 3. This resulted in a 'non-admixed' dataset which included 522 accessions (51, 144 and 327 individuals, respectively, from Group I, Group II and Group III; Table S24).

Nucleotide diversity ($\theta\pi$) and a measure of genetic differentiation (F_{ST}) were calculated for each of the three groups using VCFtools (version 0.1.17) (Danecek et al., 2011). In demographical analyses, we first used MSMC2 (Schiffels and Wang, 2020), which has advantages in estimating recent histories (Liu and Fu, 2020), with default parameters. We selected four individuals from each of the three groups that had the highest mean depth (all >20 ×) and ancestral component (based on admixture results) to ensure the quality of consensus sequences, and then used SHAPEIT4 (Delaneau et al., 2019) to phase each chromosome. MSMC-tools (https://github.com/stschiff/msmc-tools) were used to generate the input files for MSMC2 for each chromosome. Average generation time was set to one year and the mutation rate was assumed as $\mu = 1.5 \times 10^{-8}$ mutations × bp^{-1} × $generation^{-1}$ (Koch et al., 2000). Next, demographic history was also inferred with SMC++ (version v1.15) (Terhorst et al., 2017), which analyzes multiple genotypes without phasing. Finally, we estimated *Ne* in the recent past using *GONE* (Santiago et al., 2020), which is found to be accurate up for at least recent 200 generations.

To test alternative evolutionary scenarios for the origin of feral ramie, and their relationship to wild and domesticated ramies, we employed Approximate Bayesian Computation and supervised machine learning methods implemented in DIYABC-RF v1.0 (Collin et al., 2021). For Group II, one analysis treated it as a whole, and in another we defined as separate groups the two largest monophyletic groups of individuals (subclades 2A & 2B in Fig. S5). To generate the input file, using the unlinked SNP dataset, we filtered out sites that were missing from more than half of the individuals,

and sites that were monomorphic across populations, leaving 1,268,798 and 1,172,407 SNPs for six models (Fig. S8a) and eight models (Fig. S8b), respectively. For all scenarios, training sets were generated using 4,000 simulations per model, and 50 default summary statistics were calculated for observed and simulated data to train the model. Prior values were drawn from uniform distributions (Table S10). Following the recommendations in the manual, and the RF algorithm for model choice based on linear discriminant analysis, we used five noise variables and generated 2,000 Random Forest trees per model to select the most likely scenario of each set.

To identify potential selective sweeps associated with domestication and feralization, based on non-admixed individuals, selective sweeps across the ramie genome in the feral group (Group II) and in the domesticated group (Group III) were identified using SweeD (version 4.0.0) (Pavlidis et al., 2013). Genome-wide SNPs were trimmed with parameter setting "-maf 0.05, -missing 0.1", and the empirical estimate of the effective population size derived from the MSMC2 analysis described above was incorporated. Composite likelihood ratios (CLR) were calculated in windows with average size 10-kb across the genome by setting grid numbers according to chromosome lengths (number of grid = chromosome length/10000). Those with the top 5% highest CLR values were identified as potential selective sweeps, and sweeps with physical distance no larger than 100 bp were merged. Candidate genes within these genomic regions and their biological functions were retrieved according to annotations from functional databases KEGG and GO, and statistical enrichment of terms was determined.

Admixture detection and genomic composition of feral ramie

To recover the admixture history in the formation of the *B. nivea* complex, we employed the qp3Pop program in ADMIXTOOLS (Patterson et al., 2012) with default parameters.

To assess the genomic composition of feral ramie in comparison with domesticated

and wild ramies, we identified a subset of SNPs that were present in one or more accessions of domesticated material but not detected in wild material, which we termed 'crop-specific private SNPs'. Likewise, those detected in wild but not domestic material formed the subset termed 'wild-specific private SNPs' (Li *et al.*, 2017). We estimated the numbers of wild-specific and domestic-specific private SNPs in each 100-kb window across feral genomes and visualized this by plotting the log value of the ratio between crop- and wild-specific private SNPs using the ggplot2 R package (Wickham, 2016). A negative value indicates that there are more wild-specific than domestic-specific private SNPs within the genomic window.

Ecological analyses

We used several ecological analyses to reveal differences in the niche of each group, and to identify candidate ecological factors associated with habitat-specific adaptation during feralization. Most of the samples collected in China were obtained through our own fieldwork and have accurate GPS information, but samples from outside China are mainly collected from herbarium specimens, so here we only used samples collected in China. Using R package spThin (Aiello-Lammens et al., 2015), we only kept records of the same groups that were separated from each other by ≥5 km, and so the final dataset consisted of 367 unique sample locations. We obtained 26 environmental variables from the WorldClim (Fick and Hijmans, 2017), WoSIS (Batjes et al., 2020), GCAM-Demeter (Chen et al., 2020), and Human-Footprint (Venter et al., 2016), which together included bioclimatic, topographical, pedologic and anthropogenic variables (Table S25). To reduce collinearity among environmental variables, using the R package usdm (Naimi et al., 2014), we kept only those variables with VIF<5, which resulted in 12 environmental variables being retained (Table S25).

Using these data, three kinds of analysis were employed to study niche differentiation among the three groups: 1) using R package ENMTools version 1.0.4 (Warren et al., 2021), we carried out niche identity tests among the three groups, niche equivalency was quantified by Schoener's *D* and Hellinger's *I*, where a value of 0

suggest no overlap and 1 means complete overlap; 2) to quantify degree of niche overlap among the groups, we used the R package ecospat (Di Cola et al., 2017); and 3) we performed a PCA analysis, and then tested for significant differences between these three groups using ANOVA.

In addition, genome-environment association analyses were performed with PCA controlled as fixed effects using EMMAX (Zhou and Stephens, 2012), taking environmental data as phenotypes (Table S26), and employing a linear mixed model. Manhattan plots were visualized using the ggplot2 R package (Wickham, 2016), and the *p* value threshold for significance was estimated as 0.05/n (where n corresponds to the number of SNPs).

Furthermore, using the 12 variables, we predicted the potential geographic distributions for wild and feral ramie under past and future climate change. For wild ramie, we studied the potential distribution during the LIG, LGM, the present and the future. For feral ramie, ENM was only carried out for the present and future. For the future, we took the year 2090 under the pessimistic RCP8.5 scenario (IPCC, 2013). For each sample location, ENM was conducted using the 'biomod2' R package (Thuiller et al., 2009), in which we used an ensemble of six models (GBM, CTA, FDA, MARS, RF and MAXNET), with 10 bootstrap replicates, employing 75% of the localities to train the model, and applying the 'equal training sensitivity and specificity threshold' rule (Liu et al., 2005) to define the minimum threshold of suitable habitat. We assessed the quality of the predictions using the area under the receiver operator curve (AUC).

Finally, to estimate the phylogenetic conservatism of each climate variable, we quantified the phylogenetic signal using Blomberg's *K* for the 12 environmental variables (Blomberg et al., 2003). The significance was estimated through 999 randomizations with the niche distribution randomly shuffled across phylogenetic tips. We conducted Blomberg's *K* using the *multiphylosignal* functions in the R package *picante* (Kembel et al., 2010).

Funding

This study was supported by the CAS Strategic Priority Research Program (XDB31000000), the National Natural Science Foundations of China (31970356, 42171071, 32170398), the Yunnan Young & Elite Talents Projects (YNWR-QNBJ-2020-293, YNWR-QNBJ -2018-146), the Key Research Program of Frontier Sciences, CAS (ZDBS-LY-7001), the CAS 'Light of West China' Program (to Zeng-Yuan Wu and Jie Liu), the Applied and Fundamental Research Foundation of Yunnan Province (202401AT070190), CAS' Youth Innovation Promotion Association (2019385), and the Central Public-interest Scientific Institution Basal Research Fund (Y2023PT11). Richard Milne and Mark Chapman also thank the CAS President's International Fellowship Initiative for its financial support (2022VBA0004 and 2020VBB0016, respectively).

Authors' contributions

Z-Y Wu, D-Z Li, J Liu and M-B Luan conceived the study. Z-Y Wu, J Liu, Y Zhao, and M-B Luan did field work, AK Monro helped collect most samples outside of China. Z-Y Wu and Y Zhao carried out lab work. Z-Y Wu, J Liu, MA Chapman, Y-H Luo, G-F Zhu, P-Z Fan and Z-P Li performed data analyses. Z-Y Wu organized the data and wrote the first draft. MA Chapman, RT Corlett, RI Milne and MK Cadotte helped improve the focus and discussion. All authors revised and approved the final manuscript.

Data availability

The genome sequence data of feral ramie reported in this paper has been deposited in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation, under accession number GWHERBU00000000 (BioProject PRJCA015489), and is publicly accessible at https://ngdc.cncb.ac.cn/gwh. The raw resequencing data of 915 individuals reported in this paper have been deposited in the Genome Sequence Archive

in National Genomics Data, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA011837 and CRA010145) under project accession number PRJCA015489 and is publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Acknowledgements

We thank Prof. Hong Wang, Dr. Wei Xu, and Mr. Jin-Xuan Shi for their insightful discussions, and Mr. Zhi-Ming Sun for kind help during field work. We acknowledge valuable contribution of Dr. Ting Zhang, Dr. Chun-Yuan Zhang, Dr. Dong An, Dr. Song-Bo Wang and Mr. Ren-Gang Zhang for their kind assistance with software. The herbaria of the Royal Botanic Gardens, Kew (K) and Institute of Botany, the Chinese Academy of Sciences (PE) are thanked for providing some DNA materials. This work was facilitated by the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences. Wuhan Frasergen Bioinformatics Co. Ltd. is thanked for valuable technical support in whole genome sequencing.

Declaration of interest

No conflict of interest is declared.

References

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., and Anderson, R.P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography **38**:541-545.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research **19**:1655-1664.

Bao, Z., and Eddy, S.R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Research **12**:1269-1276.

Batjes, N.H., Ribeiro, E., and Van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth System

Science Data **12**:299-320.

Ben Michael, T.E., Faigenboim, A., Shemesh-Mayer, E., Forer, I., Gershberg, C., Shafran, H., Rabinowitch, H.D., and Kamenetsky-Goldstein, R. (2020). Crosstalk in the darkness: Bulb vernalization activates meristem transition via circadian rhythm and photoperiodic pathway. BMC Plant Biology **20**:77.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Research **27**:573-580.

Blomberg, S.P., Garland Jr, T., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution **57**:717-745.

Borah, P., Das, A., Milner, M.J., Ali, A., Bentley, A.R., and Pandey, R. (2018). Long non-coding RNAs as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. Genes **9**:459.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology **268**:78-94.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., and Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research **18**:188-196.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972-1973.

Casañas, F., Simó, J., Casals, J., and Prohens, J. (2017). Toward an evolved concept of landrace. Frontiers in Plant Science **8**:145.

Chen, C.J. (1980). *Archiboehmeria* C. J. Chen–a new genus of Urticaceae. Acta Phytotaxonomica Sinica **18**:476-481.

Chen, C.J., Lin, Q., Friis, I., C.M., W.-D., and Monro, A.K. (2003). Urticaceae. In Flora of China, Z.Y.Wu and P.H. Raven, ed. (Science Press, Bejing & Missouri Botanical Garden Press: Bejing), pp. 76-189.

Chen, K., Ming, Y., Luan, M., Chen, P., Chen, J., Xiong, H., Chen, J., Wu, B., Bai, M., and Gao, G. (2023). The chromosome-level assembly of ramie (*Boehmeria nivea* L.) genome provides insights into molecular regulation of fiber fineness. Journal of Natural Fibers **20**:2168819.

Chen, M., Vernon, C.R., Graham, N.T., Hejazi, M., Huang, M., Cheng, Y., and Calvin, K. (2020). Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. Scientific Data **7**:320.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics **5**:4.10.11-14.10.14.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018a). fastp: An ultra-fast all-in-one fastq preprocessor. Bioinformatics **34**:i884-i890.

Chen, X. (2007). The history, status and future of ramie textile industry in China. Plant Fiber Sciences in China **29**:77-85.

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018b). SOAPnuke: A mapreduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data.

GigaScience **7**:1-6.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods **18**:170-175.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.Y., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. Fly **6**:80-92.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The last glacial maximum. Science **325**:710-714.

Collin, F.-D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J.-M., and Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC random forest. Molecular Ecology Resources **21**:2598-2613.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics **2008**:619832.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156-2158.

Darwin, C. (1868). The variation of animals and plants under domestication. Volume ii (John Murray).

Delaneau, O., Zagury, J. F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nature Communications **10**:5436.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., et al. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. Ecography **40**:774-787.

Dinov, I.D., Rubin, D., Lorensen, W., Dugan, J., Ma, J., Murphy, S., Kirschner, B., Bug, W., Sherman, M., Floratos, A., et al. (2008). iTools: A framework for classification, categorization and integration of computational biology resources. PLoS ONE **3**:e2265.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., and Aiden, A.P. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356**:92-95.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Systems **3**:95-98.

Dwivedi, S.L., Chapman, M.A., Abberton, M.T., Akpojotor, U.L., and Ortiz, R. (2023). Exploiting genetic and genomic resources to enhance productivity and abiotic stress adaptation of underutilized pulses. Frontiers in Genetics **14**:1193780.

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K., and Ortiz, R. (2016). Landrace germplasm for improving yield and abiotic stress adaptation. Trends in Plant Science **21**:31-42.

Ellstrand, N.C., Heredia, S.M., Leak-Garcia, J.A., Heraty, J.M., Burger, J.C., Yao, L., Nohzadeh-Malakshah, S., and Ridley, C.E. (2010). Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. Evolutionary Applications **3**:494-504.

Farrant, J.M., and Hilhorst, H. (2022). Crops for dry environments. Current Opinion in Biotechnology **74**:84-91.

Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Gering, E., Incorvaia, D., Henriksen, R., Conner, J., Getty, T., and Wright, D. (2019). Getting back to nature: Feralization in animals and plants. Trends in Ecology & Evolution **34**:1137-1151.

Goel, M., and Schneeberger, K. (2022). Plotsr: Visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics **38**:2922-2926.

Goel, M., Sun, H., Jiao, W.-B., and Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biology **20**:277.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology **29**:644-652.

Groppi, A., Liu, S., Cornille, A., Decroocq, S., Bui, Q.T., Tricon, D., Cruaud, C., Arribat, S., Belser, C., Marande, W., et al. (2021). Population genomics of apricots unravels domestication history and adaptive events. Nature Communications **12**:3956.

Grummer, J.A., Beheregaray, L.B., Bernatchez, L., Hand, B.K., Luikart, G., Narum, S.R., and Taylor, E.B. (2019). Aquatic landscape genomics and environmental effects on genetic variation. Trends in Ecology & Evolution **34**:641-654.

Guo, W.L., Xin, M.M., Wang, Z., Yao, Y., Hu, Z., Song, W., Yu, K., Chen, Y., Wang, X., and Guan, P. (2020). Origin and adaptation to high altitude of Tibetan semi-wild wheat. Nature Communications **11**:5085.

Gutaker, R.M., and Purugganan, M.D. (2024). Adaptation and the geographic spread of crop species. Annual Review of Plant Biology **75**:2.1-2.28.

Gutaker, R.M., Chater, C.C.C., Brinton, J., Castillo-Lorenzo, E., Breman, E., and Pironon, S. (2022). Scaling up neodomestication for climate-ready crops. Current Opinion in Plant Biology **66**:102169.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research **31**:5654-5666.

He, X., Ziegler, A.D., Elsen, P.R., Feng, Y., Baker, J.C.A., Liang, S., Holden, J., Spracklen, D.V., and Zeng, Z. (2023). Accelerating global mountain forest loss threatens biodiversity hotspots. One Earth **6**:303-315.

Helmkampf, M., Bellinger, M.R., Geib, S.M., Sim, S.B., and Takabayashi, M. (2019). Draft genome of the rice coral *Montipora capitata* obtained from linked-read sequencing. Genome Biology and Evolution **11**:2045-2054.

Hester, S.B., and Yuen, M.L. (1989). Ramie: Patterns of world production and trade. Journal of the Textile Institute **80**:493-505.

Jiang, A., Liu, J., Gao, W., Ma, R., Zhang, J., Zhang, X., Du, C., Yi, Z., Fang, X., and Zhang, J. (2023). Transcriptomic and metabolomic analyses reveal the key genes related to shade tolerance in soybean. International Journal of Molecular Sciences **24**:14230.

Jiao, P., Ma, R., Wang, C., Chen, N., Liu, S., Qu, J., Guan, S., and Ma, Y. (2022). Integration of mRNA and microRNA analysis reveals the molecular mechanisms underlying drought stress tolerance in maize (*Zea mays* L.). Frontiers in Plant Science **13**:932667.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics **30**:1236-1240.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research **110**:462-467.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Research **46**:D335-D342.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution **30**:772-780.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**:1463-1464.

Koch, M.A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis,* and related genera (Brassicaceae). Molecular Biology and Evolution **17**:1483-1498.

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., and Meshveliani, T. (2009). 30,000-year-old wild flax fibers. Science **325**:1359-1359.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research **35**:3100-3108.

Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., and Dobney,

K. (2014). Current perspectives and the future of domestication studies. Proceedings of the National Academy of Sciences of the United States of America **111**:6139-6146.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094-3100.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows–wheeler transform. Bioinformatics **26**:589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009). The sequence alignment/map format and samtools. Bioinformatics **25**:2078-2079.

Li, L., Han, C., Yang, J., Tian, Z., Jiang, R., Yang, F., Jiao, K., Qi, M., Liu, L., and Zhang, B. (2023). Comprehensive transcriptome analysis of responses during cold stress in wheat (*Triticum aestivum* L.). Genes **14**:844.

Li, L.F., Li, Y.L., Jia, Y., Caicedo, A.L., and Olsen, K.M. (2017). Signatures of adaptation in the weedy rice genome. Nature Genetics **49**:811-814.

Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., Zhao, P., Guo, J., Ding, T., Guan, L., et al. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biology **20**:36.

Liao, J., and Yang, X. (2016). Study on the evolution of grass cloth. Asian Social Science **12**:109.

Liao, L., Li, T., Zhang, J., Xu, L.L., Deng, H.S., and Han, X.J. (2014). The domestication and dispersal of the cultivated ramie (*Boehmeria nivea* (L.) Gaud. In Freyc.) determined by nuclear SSR marker analysis. Genetic Resources and Crop Evolution **61**:55-67.

Liu, B.H., Shi, Y.J., Yuan, J.Y., Hu, X.S., Zhang, H., Li, N., Li, Z.Y., Chen, Y.X., Mu, D.S., and Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv preprint. arXiv:1308.2012.

Liu, C., Berry, P.M., Dawson, T.P., and Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28**:385-393.

Liu, L.J., Meng, Z.Q., Wang, B., Wang, X.X., Yang, J.Y., and Peng, D.X. (2009). Genetic diversity among wild resources of the genus *Boehmeria* Jacq. from west China determined using inter-simple sequence repeat and rapid amplification of polymorphic DNA markers. Plant Production Science **12**:88-96.

Liu, X., and Fu, Y.X. (2020). Stairway Plot 2: Demographic history inference with folded SNP frequency spectra. Genome Biology **21**:280.

Lowe, T.M., and Eddy, S.R. (1997).tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research **25**:955-964.

Mabry, M.E., Rowan, T.N., Pires, J.C., and Decker, J.E. (2021a). Feralization: Confronting the complexity of domestication and evolution. Trends in Genetics **37**:302-305.

Mabry, M.E., Turner-Hissong, S.D., Gallagher, E.Y., McAlvay, A.C., An, H., Edger, P.P., Moore, J.D., Pink, D.A.C., Teakle, G.R., Stevens, C.J., et al. (2021b). The

evolutionary history of wild, domesticated, and feral *Brassica oleracea* (Brassicaceae). Molecular Biology and Evolution **38**:4419-4434.

Mabry, M.E., Bagavathiannan, M.V., Bullock, J.M., Wang, H., Caicedo, A.L., Dabney, C.J., Drummond, E.B.M., Frawley, E., Gressel, J., Husband, B.C., et al. (2023). Building a feral future: Open questions in crop ferality. Plants, People, Planet **5**:635-649.

Manel, S., Andrello, M., Henry, K., Verdelet, D., Darracq, A., Guerin, P.-E., Desprez, B., and Devaux, P. (2018). Predicting genotype environmental range from genome–environment associations. Molecular Ecology **27**:2823-2833.

Martin Cerezo, M.L., López, S., van Dorp, L., Hellenthal, G., Johnsson, M., Gering, E., Henriksen, R., and Wright, D. (2023). Population structure and hybridisation in a population of Hawaiian feral chickens. Heredity **130**:154-162.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**:1297-1303.

Meyer, R.S., Choi, J.Y., Sanches, M., Plessis, A., Flowers, J.M., Amas, J., Dorph, K., Barretto, A., Gross, B., Fuller, D.Q., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. Nature Genetics **48**:1083-1088.

Mgwatyu, Y., Stander, A.A., Ferreira, S., Williams, W., and Hesse, U. (2020). Rooibos (*Aspalathus linearis*) genome size estimation using flow cytometry and k-mer analyses. Plants **9**:270.

Mistry, J., Bateman, A., and Finn, R.D. (2007). Predicting active site residue annotations in the Pfam database. BMC Bioinformatics **8**:298.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., and Toxopeus, A.G. (2014). Where is positional uncertainty a problem for species distribution modelling? Ecography **37**:191-203.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933-2935.

Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q.J.M.b., and evolution (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution **32**:268-274.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics **192**:1065-1093.

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Molecular Biology and Evolution **30**:2224-2234.

Pflug, J.M., Holmes, V.R., Burrus, C., Johnston, J.S., and Maddison, D.R. (2020). Measuring genome sizes using read-depth, k-mers, and flow cytometry: Methodological comparisons in beetles (Coleoptera). G3 Genes|Genomes|Genetics **10**:3047-3060.

Pisias, M.T., Bakala, H.S., McAlvay, A.C., Mabry, M.E., Birchler, J.A., Yang, B.,

and Pires, J.C. (2022). Prospects of feral crop de novo redomestication. Plant and Cell Physiology **63**:1641-1653.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics **21**:i351-i358.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics **38**:904-909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics **81**:559-575.

Purugganan, M.D. (2022). What is domestication? Trends in Ecology & Evolution **37**:663-671.

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., Yu, Y., Fu, F., Wang, Y., Qian, F., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. Nature Communications **8**:15323.

Qiu, J., Jia, L., Wu, D., Weng, X., Chen, L., Sun, J., Chen, M., Mao, L., Jiang, B., Ye, C., et al. (2020). Diverse genetic mechanisms underlie worldwide convergent rice feralization. Genome Biology **21**:70.

Retief, J.D. (2000). Phylogenetic analysis using phylip. In Bioinformatics methods and protocols, S. Misener and S.A. Krawetz, eds. (Humana Press: Totowa, New Jersey, USA), pp. 243-258.

Roy, S., and Lutfar, L.B. (2012). Bast fibres: Ramie. In handbook of natural fibres: Types, properties and factors affecting breeding and cultivation, R.M. Kozłowski, ed. (Woodhead Publishing: Cambridge, England, UK), pp. 47-55.

Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., and Caballero, A. (2020). Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. Molecular Biology and Evolution **37**:3642-3653.

Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The multiple sequentially markovian coalescent. In Statistical population genomics (Human Press: New York, USA), pp. 147-166.

Sen, T., and Reddy, H.J. (2011). Various industrial applications of hemp, kinaf, flax and ramie natural fibres. International Journal of Innovation, Management and Technology **2**:192.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210-3212.

Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**:31.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Research **34**:W435-W439.

Tang, H., Krishnakumar, V., and Li, J. (2015). JCVI: JCVI utility libraries (v0.5.7). Zenodo. https://zenodo.org/record/31631/export/xd.

Taru, Y., and Watan, B. (2020). The route to ramie cultural ecology. Senri Ethnological Studies **103**:11-20.

Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics **49**:303-309.

Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M.B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. Ecography **32**:369-373.

Thurber, C.S., Reagon, M., Gross, B.L., Olsen, K.M., Jia, Y., and Caicedo, A.L. (2010). Molecular evolution of shattering loci in US weedy rice. Molecular Ecology **19**:3271-3284.

Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K.R., Possingham, H.P., Laurance, W.F., Wood, P., Fekete, B.M., et al. (2016). Global terrestrial human footprint maps for 1993 and 2009. Scientific Data **3**:160067.

Vigueira, C., Olsen, K., and Caicedo, A. (2013). The red queen in the corn: Agricultural weeds as models of rapid adaptive evolution. Heredity **110**:303-311.

Villa, T.C.C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. Plant Genetic Resources **3**:373-384.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS ONE **9**:e112963.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research **38**:e164-e164.

Wang, L., Yang, H., Liu, R., Fan, G.J.E.S., and Research, P. (2015). Detoxification strategies and regulation of oxygen production and flowering of *Platanus acerifolia* under lead (Pb) stress by transcriptome analysis. Environmental Science and Pollution Research **22**:12747-12758.

Wang, Y., Li, F., He, Q., Bao, Z., Zeng, Z., An, D., Zhang, T., Yan, L., Wang, H., and Zhu, S. (2021). Genomic analyses provide comprehensive insights into the domestication of bast fiber crop ramie (*Boehmeria nivea*). The Plant Journal **107**:787-800.

Warren, D.L., Matzke, N.J., Cardillo, M., Baumgartner, J.B., Beaumont, L.J., Turelli, M., Glor, R.E., Huron, N.A., Simões, M., Iglesias, T.L., et al. (2021). Enmtools 1.0: An R package for comparative ecological biogeography. Ecography **44**:504-511.

Wedger, M.J., Roma-Burgos, N., and Olsen, K.M. (2022). Genomic revolution of US weedy rice in response to 21st century agricultural technologies. Communications Biology **5**:885.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (Springer-Verlag).

Wu, D., Lao, S., and Fan, L. (2021). De-domestication: An extension of crop evolution. Trends in Plant Science **26**:560-574.

Xiao, X., Haberle, S.G., Li, Y., Liu, E., Shen, J., Zhang, E., Yin, J., and Wang, S.

(2018). Evidence of Holocene climatic change and human impact in northwestern Yunnan province: High-resolution pollen and charcoal records from Chenghai Lake, southwestern China. The Holocene **28**:127-139.

Xie, Y., Wang, Y., Liu, X., Shen, J., and Wang, Y. (2021). Increasing human activities during the past 2,100 years in southwest China inferred from a fossil pollen record. Vegetation History and Archaeobotany **30**:477-488.

Xu, X., Zhang, S., Cheng, Z., Li, T., Jia, Y., Wang, G., Yang, Z., Xian, J., Yang, Y., Zhou, W.J.E.S., et al. (2020). Transcriptome analysis revealed cadmium accumulation mechanisms in hyperaccumulator *Siegesbeckia orientalis* L. **27**:18853-18865.

Xu, Z., and Wang, H. (2007). Ltr_finder: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research **35**:W265-W268.

Yu, H., and Li, J. (2022). Breeding future crops to feed the world through de novo domestication. Nature Communications **13**:1171.

Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., Mao, L., Liu, Q., Xia, Z., Zhou, J., et al. (2018). Origin and evolution of qingke barley in Tibet. Nature Communications **9**:5433.

Zeven, A.C. (1998). Landraces: A review of definitions and classifications. Euphytica **104**:127-139.

Zhang, S.J., Wang, G.D., Ma, P., Zhang, L.L., Yin, T.T., Liu, Y.H., Otecko, N.O., Wang, M., Ma, Y.P., Wang, L., et al. (2020). Genomic regions under selection in the feralization of the dingoes. Nature Communications **11**: 671.

Zhao, Y., Yi, T.-S., Milne, R., Li, Z.-P., Yin-Lei, L., Kipkoech, A., Li, K., Fu, X.-G., Li, D.-Z., and Wu, Z.-Y. (2024). *Boehmeria nivea* var. *strigosa* (Urticaceae), a new variety from southwest China. Guihaia **In press**.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics **44**:821-824.

Zsögön, A., Peres, L.E.P., Xiao, Y., Yan, J., and Fernie, A.R. (2022). Enhancing crop diversity for food security in the face of climate uncertainty. The Plant Journal **109**:402-414.

Figure legends

Fig. 1. Genomic landscape of feral ramie and comparative genomic analyses between the feral and domesticated genomes. A, Genomic features of feral ramie. 1) pseudochromosomes, 2) Gene density per 100 kb window, 3) the distribution of repetitive sequences, 4) GC content, and 5) the inner lines show syntenic blocks within the feral genome. **B,** Genome collinearity between the feral and domesticated ramie assemblies. **C,** Doughnut chart showing the distribution of variants between feral and

1179 domestic ramies. Numbers in parentheses (x/y) indicate x, the total length of each type
1180 of variation, and y, the number of events. **D,** Syntenic analyses between the assemblies
1181 of domesticated (reference) and feral (query) ramies; syntenic regions and SVs are
1182 highlighted with different colors.
1183
1184 **Fig. 2**. **Population structure and genetic diversity of ramie. A,** Geographic
1185 distribution of *Boehmeria nivea* based on occurrence points from GBIF (black circles).
1186 Sampling sites for the current study are shown as blue (wild), yellow (feral), and purple
1187 (domesticated) circles. **B,** Admixture analyses with different numbers of groups (K = 2
1188 to 5). Each vertical bar represents one ramie accession, and the x axis shows the three
1189 genetic groups. Each color represents one putative ancestral background, and the y axis
1190 quantifies ancestry membership. **C,** Two-dimensional PCA plot showing the clustering
1191 of accessions color-coded in the same scheme as panel b. **D,** Nucleotide diversity ($\theta\pi$)
1192 within and genetic differentiation (F_{ST}) between the groups.
1193
1194 **Fig. 3. Demographic history and candidate genome regions with evidence for**
1195 **selective sweeps between groups. A,** Best scenario when feral ramie is considered as
1196 a whole**. B,** Best scenario when the two largest monophyletic subclades of feral ramie
1197 are treated as discrete populations. **C,** Demographic history of wild (group I), feral (II)
1198 and domesticated (III) ramies using MSMC2. The y axis represents inferred effective
1199 population size over time and the x axis represents time. **D,** Distribution of wild- and
1200 cultivar-specific SNPs for each chromosome in feral material based on log10 of the
1201 ratio of crop specific to wild specific SNPs in 100 kb regions. The box shows the 95%
1202 confidence interval and the black bar within each box is the mean. The horizontal dotted
1203 line represents zero, and positive and negative values represent excesses of domestic-
1204 like and wild-like SNPs, respectively. **E,** Distribution of the regions under selection in
1205 feral ramie. **F,** Distribution of the regions under selection in feral ramie, with horizontal
1206 dotted lines representing the cutoff fulfilling the requirement for the selected regions.
1207

Fig. 4. Ecological analyses results. A-C, Niche identity tests among Groups I (wild), II (feral), and III (domesticated). The arrows indicate the observed niche equivalency, and the histograms represent the simulated (expected) equivalency. All differences between the observed index and the expected index rejected the hypothesis that environmental niches between regions were identical (P<0.01). **D-F,** Niche overlap analysis based on pairwise comparisons among the three groups. The solid and dashed contour lines delimit the 100th and 75th quantiles, respectively, of the density at the available environment. Blue, yellow, and purple represent Group I, Group II, and Group III, respectively. Pink in each figure means stability between two groups. **G,** Principal Coordinate Analysis for 12 environmental variables, arrow lengths indicate the relative contributions of each environmental factor to the principal components. The details of the variables refer to Table S25.

Fig. 5. Potential range shift and genome-environment associations. A, Results of GEA analysis. Genomic locations of SNPs associated with environmental factors, genes mentioned in the text are indicated with red arrows. **B-G,** Potential distribution range of wild ramie (**A-D**) and feral ramie (**E-F**) by ENM using 12 environmental variables and species occurrence points.

Fig. 6. Morphological comparison among three varieties. **A-C,** *Boehmeria nivea* var. *nivea*; **A,** habit with unbranched stem; **B,** white abaxial leaf blade; **C,** free and patent hirsute stem. **D-F,** *B. nivea* var. *strigosa*; **D,** habit with branched stem; **E,** green abaxial leaf blade; **F,** patent strigose stem and partly connate stipule. **G-I,** *B. nivea* var. *tenacissima*; **G,** habit with branched stem; **H,** mixed color of abaxial leaf blade; **I,** appressed hirsute and partly connate stipule.

Fig. S1. Hi-C chromatin interaction map of the feral ramie genome assembly.

Fig. S2. Genome size estimate for feral ramie based on (**A**) the *K*-mer method and (**B**)

flow cytometry. *Solanum lycopersicum* L. was used as an internal standard.

Fig. S3. Syntenic relationship dot plot between feral (y axis) and domesticated (x axis) ramie genomes. Dots closest to the diagonal line represent collinearity between the two genomes with fragments <20 Kb filtered out.

Fig. S4. Distribution of ramie (*Boehmeria nivea*) and sampling sites for the current study. Black circles represent herbarium records, blue, yellow, and purple circles represent wild, feral and domesticated ramies, respectively. **A**, Distribution and sampling all over the world; **B**, Distribution and sampling in Asia.

Fig. S5. Maximum-Likelihood (ML) phylogenetic tree of ramie resequencing samples using 120,201 high-confidence SNPs. The numbers on the nodes indicate bootstrap values. Blue, yellow and purple lines represent wild, feral, and domesticated ramie, respectively. Each node consists of lab code_variety name_Country_Province, Bs, Bn, and Bt represent *Boehmeria nivea* var. *strigosa*, *B. nivea* var. *nivea*, and *B.* var. *tenacissima*, respectively.

Fig. S6. A rooted NJ tree of 915 ramie accessions based on single-nucleotide polymorphisms (SNPs), using *Girardinia diversifolia* as out group; The colored lines represent the sample source (see Fig. S5)

Fig. S7. Population structure analysis in ramie. **A**, Cross validation error with increasing values of *K*. **B**, ADMIXTURE plots for all accessions. *K* (the number of clusters) from 2 to 10 are shown.

Fig. S8. Results from the Approximate Bayesian Computation analysis implemented in the program DIYABC-RF to infer the most likely demographic scenario. In each panel the best scenario is surrounded by a box. **A,** Feral ramie is

considered as a whole, and six models analyzed. **B,** The two largest monophyletic subclades of feral ramie are used as discrete populations, and eight models analyzed. Groups 2A and 2B are showed in Fig. S5. Each colored segment depicts a distinct effective population size. tx represents coalescence time (in generations), Nx represents estimated population size, and rx represents the proportion of admixture between groups.

Fig. S9. Demographic history of wild (group I), feral (II) and domesticated (III) ramies using SMC++ (A) and *GONE* (B).

Fig. S10. Manhattan plots and Quantile-quantile plots comparing the observed $-\log_{10}(p)$ with expected $-\log_{10}(p)$ for 12 environmental variables in feral ramie. The genome-wide significant value threshold ($-\log_{10}(p) = 7.83$) is indicated by a horizontal dash–dot line.

Table S1. Summary of assembly and annotation of the feral ramie genome.

Table S2. The number and distribution per chromosome of protein-coding genes and non-coding RNAs in the feral ramie genome.

Table S3. Functional annotation of the feral ramie genome.

Table S4. BUSCO (Benchmarking Universal Single-Copy Orthologs) evaluation of genome completeness of feral ramie.

Table S5. Structural variants between feral and domesticated genomes.

Table S6. Origin of ramie accessions used in this study and their sequencing and mapping statistics. Group division is based on the ML tree.

Table S7. Summary of single nucleotide polymorphism (SNP) and insertions and deletions (indels) among 915 ramie accessions.

Table S8. Scenario choice in DIY ABC-RF when feral ramie is considered as a whole; six models analyzed.

Table S9. Scenario choice in DIY ABC-RF when the two largest monophyletic subclades of feral ramie are treated as discrete populations; eight models analyzed.

Table S10. Parameter estimates for selected best scenarios and associated 95% Confidence Intervals defined by the 0.05 and 0.95 quantiles (Q) of the posterior distribution. Units are number of individuals for effective population size parameters (N) and years before present (yrs BP) for divergence time parameters (t).

Table S11. Result of detecting gene flow using ADMIXTOOLS. A significantly negative f3 value indicates that Target is an admixed population of ancestries, gene flow occurred from Source 1 and Source 2.

Table S12. Number and proportion of wild- and domesticated-specific private SNPs in feral ramie.

Table S13. Regions putatively under selection in feral ramie.

Table S14. Regions putatively under selection in domesticated ramie.

Table S15. GO analysis of feralization-related genes identified by SweeD analysis.

Table S16. KEGG analysis of feralization-related genes identified by SweeD analysis.

Table S17. GO analysis of domestication-related genes identified by SweeD analysis.

Table S18. KEGG analysis of domestication-related genes identified by SweeD analysis.

Table S19. Results of niche overlap analysis and proportion of niche change among three groups.

Table S20. One-way analysis of variance (ANOVA) for PCA1 and PCA2 followed by LSD multiple comparison test among three groups. The mean difference is significant at the 0.05 level.

Table S21. Phylogenetic signal of each climatic variable. The value represents Blomberg's *K* and significance was estimated through 999 randomizations with the trait distribution randomly shuffled across phylogenetic tips.

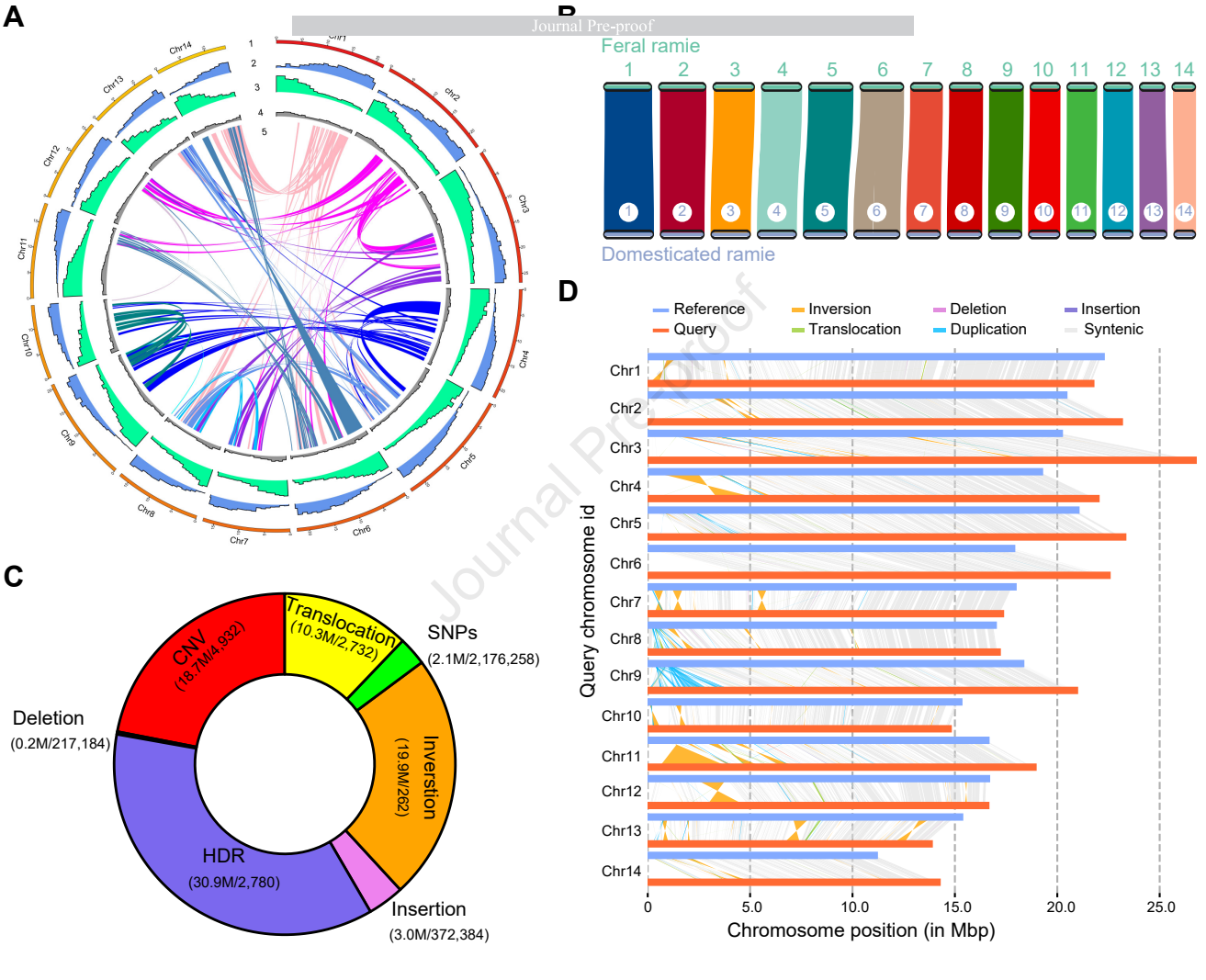
Table S22. Results of genome-environment association in feral ramie.

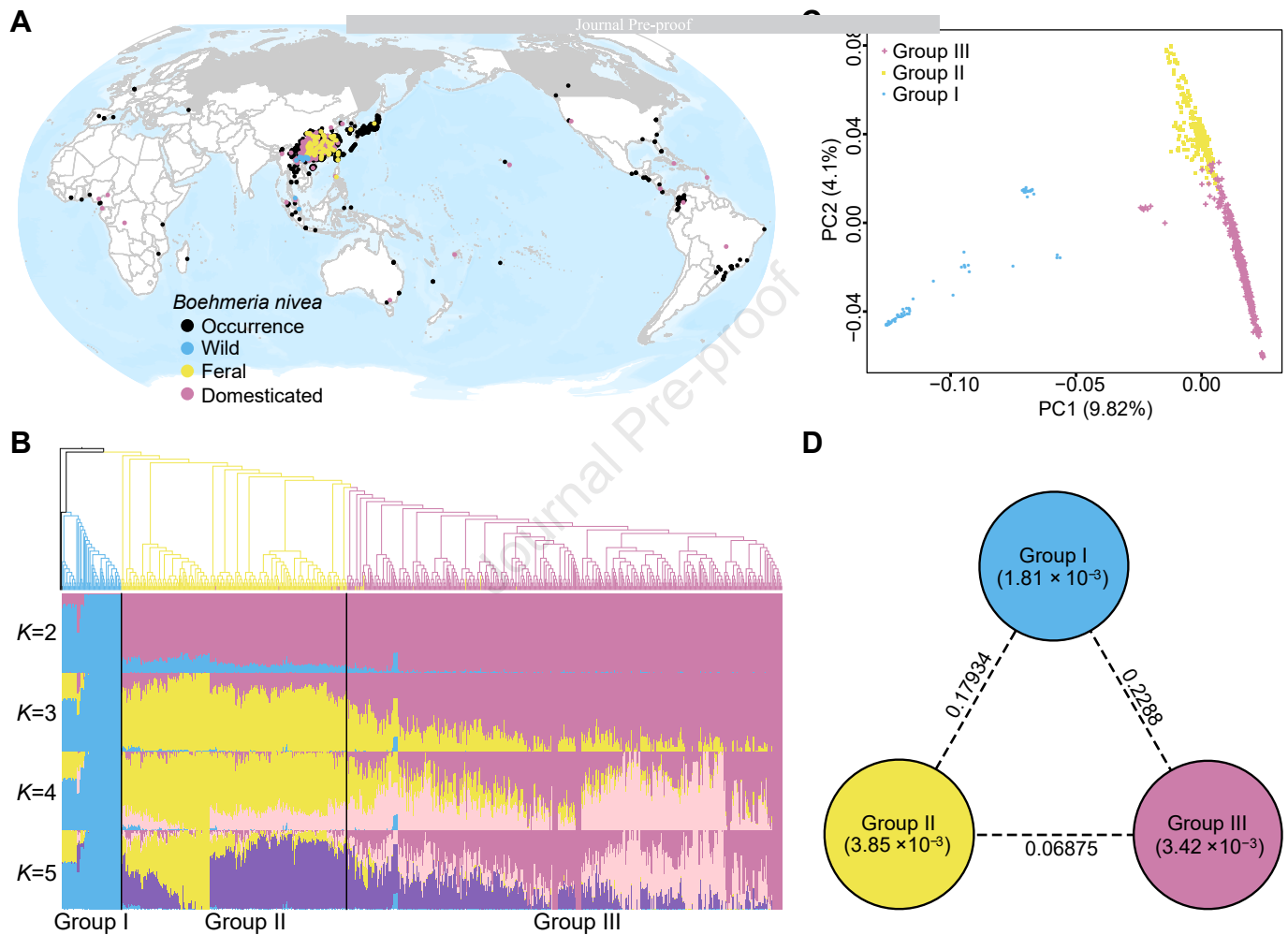
Table S23. List of AUC, the thresholds selected in the ecological niche modeling (ENM), and environmental variables with VIF<5 for wild and feral ramies.

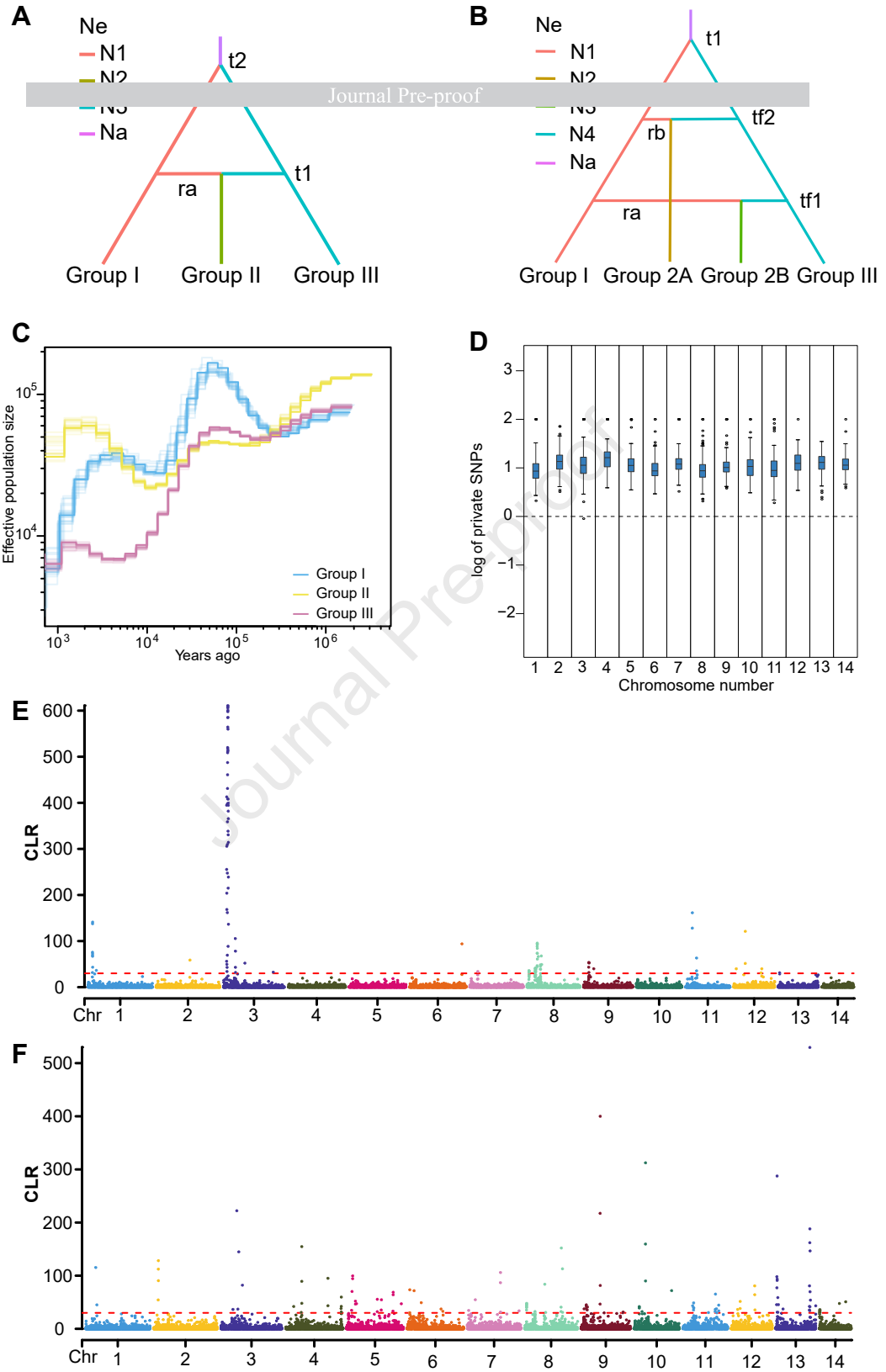
Table S24. List of 522 individuals in the non-admixed dataset.

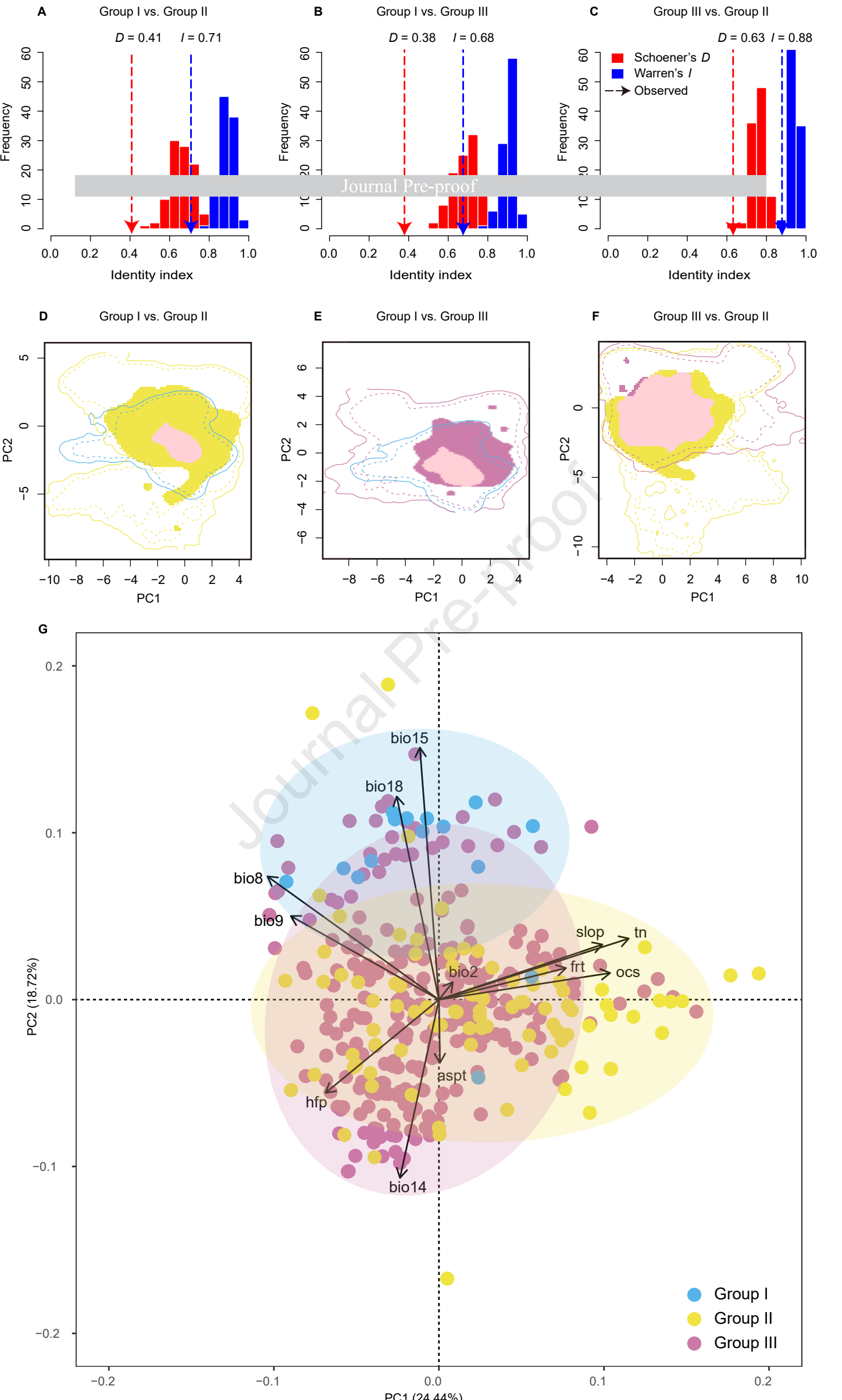
Table S25. Environmental variables considered in this study, taken from four databases, Y represents those selected after reducing collinearity.

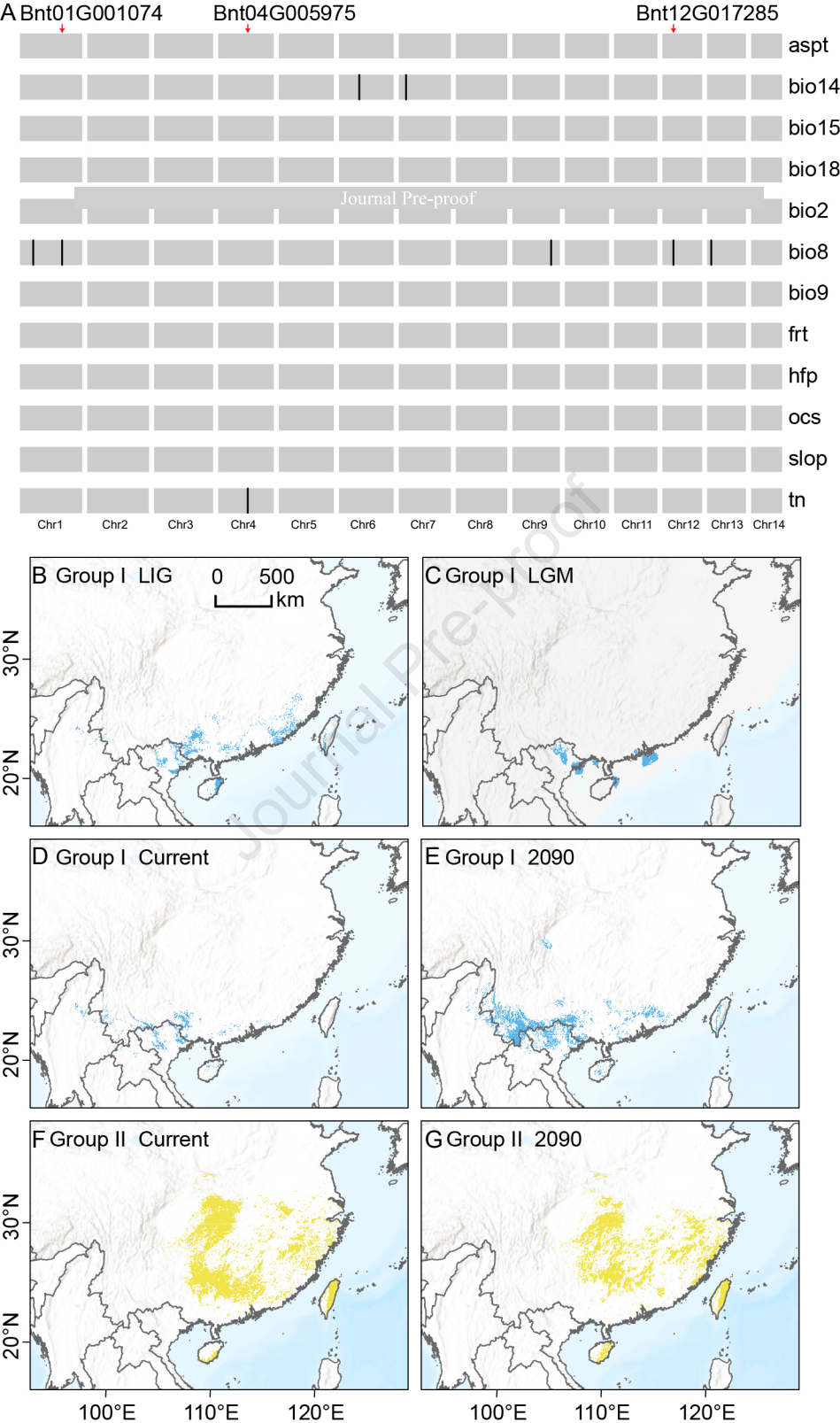
Table S26. Environmental data for feral ramie samples used for genome-environment association analysis.








A
Group I vs. Group II
D = 0.41
I = 0.71
B
Group I vs. Group III
D = 0.38
I = 0.68
C
Group III vs. Group II
D = 0.63
I = 0.88
Schoener's D
Warren's I
Observed
Frequency
Identity index
Journal Pre-proof
D
Group I vs. Group II
E
Group I vs. Group III
F
Group III vs. Group II
PC1
PC2
G
bio15
bio18
bio8
bio9
bio2
slop
tn
frt
ocs
aspt
hfp
bio14
PC1 (24.44%)
PC2 (18.72%)
Group I
Group II
Group III
Journal Pre-proof


A Bnt01G001074
Bnt04G005975
Bnt12G017285
aspt
bio14
bio15
bio18
bio2
bio8
bio9
frt
hfp
ocs
slop
tn
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Journal Pre-proof
B Group I LIG
0
500
km
C Group I LGM
D Group I Current
E Group I 2090
F Group II Current
G Group II 2090
30°N
20°N
100°E
110°E
120°E
Journal Pre-Proof