Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

Plant Communications
Research article

CellPress
Partner Journal

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

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] and 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 and Yunnan Key Laboratory for the Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China

***Correspondence: Jie Liu (liujie@mail.kib.ac.cn), Ming-Bao Luan (luanmingbao@caas.cn), De-Zhu Li (dzl@mail.kib.ac.cn)**

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

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 feralization. We first produced a chromosome-scale *de novo* genome assembly of feral ramie and investigated structural variations between feral and domesticated ramie genomes. Next, we gathered 915 accessions from 23 countries, comprising cultivars, major landraces, feral populations, and the wild progenitor. Based on whole-genome resequencing of these accessions, we constructed the most comprehensive ramie genomic variation map to date. Phylogenetic, demographic, and admixture signal detection analyses indicated that feral ramie is of exoferal or exo-endo origin, i.e., descended from hybridization between domesticated ramie and the wild progenitor or ancient landraces. Feral ramie has higher genetic diversity than wild or domesticated ramie, and genomic regions affected by natural selection during feralization differ from those under selection during domestication. Ecological analyses showed that feral and domesticated ramie have similar ecological niches that differ substantially from the niche of the wild progenitor, and three environmental variables are associated with habitat-specific adaptation in feral ramie. These 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.

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., and Li D.-Z. (2024). Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop. Plant Comm. **5**, 100942.

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; Pisias et al.,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

2022; Mabry et al., 2023). 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 (Wu et al., 2021; Martin Cerezo et al., 2023). 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 (Zsögön et al., 2022; Gutaker and Purugganan, 2024), 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 an excellent model for studying the evolutionary mechanism of feralization. It was one of the first fiber crops to be domesticated and has been 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, after 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 the evolution of feralization (Ellstrand et al., 2010). *B. nivea* is separated into three morphologically distinct varieties: var. *nivea*, only known from cultivated or naturalized populations, and var. *tenacissima* and var. *strigosa* (Zhao and Milne, 2024), which both occur in apparently natural populations. Previous attempts to understand ramie domestication used limited numbers of molecular markers (Liu et al., 2009; Liao et al., 2014) and narrow population sampling, giving an incomplete picture of the location and timing of domestication. To overcome these shortcomings, 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 that led to the origin of feral ramie and to 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 toward a comprehensive understanding of the evolutionary mechanisms underlying feralization. A high-quality feral ramie genome was assembled (Figure 1A) from 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 (Figure 1A and Supplemental Figure 1; Supplemental Tables 1 and 2), considerably smaller than the estimated genome size of ~380 Mb determined by the k-mer method and flow cytometry (Supplemental Figure 2). 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 owing to effects of 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 the k-mer method is used (Pflug et al., 2020). After genome annotation, we obtained 22 312 annotated protein-coding genes and 2164 noncoding RNA genes, and we determined that more than half (54.85%) of the feral ramie genome was composed of repetitive elements (Supplemental Table 1). Over 95% of the predicted genes showed homology to genes with known functional annotations in public databases (Supplemental Table 3), and the BUSCO analysis revealed 1546 out of 1614 (95.8%) complete BUSCOs, 22 (1.4%) of which were duplicated (Supplemental Table 4). These two results indicate that the newly assembled genome is of high quality, and we are confident that our genome is well assembled.

Alignment of the new feral and existing cultivated reference genomes revealed high collinearity (Figure 1B and Supplemental Figure 3), as well as a considerable number of genomic variants between them (Figure 1C and 1D; Supplemental Table 5). The distribution of variants was not uniform along the chromosomes. Among all classes of structural variants

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

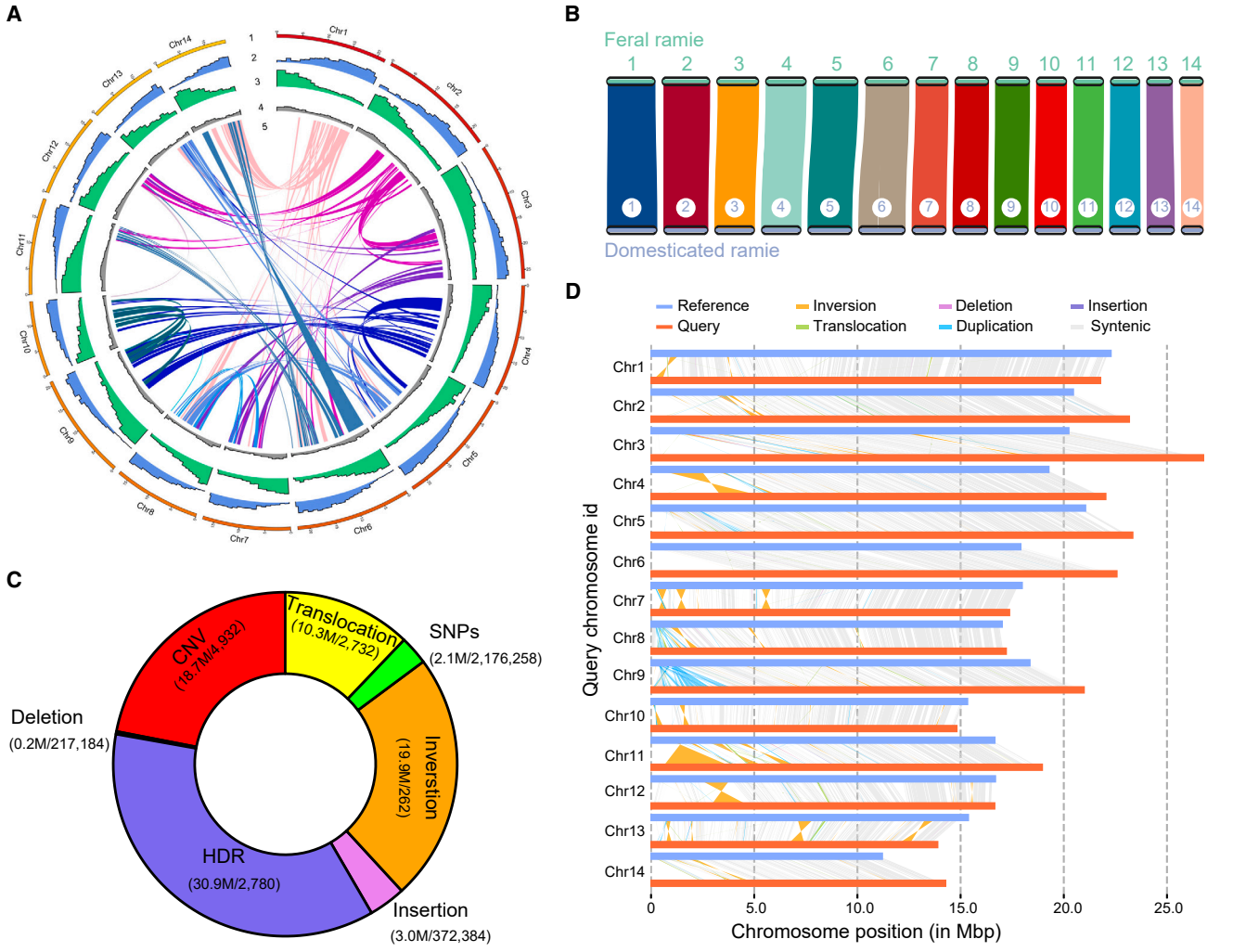

Figure 1. Genomic landscape of feral ramie and comparative genomic analyses between feral and domesticated ramie genomes.
(A) Genomic features of feral ramie. (1) Pseudochromosomes, (2) gene density per 100-kb window, (3) distribution of repetitive sequences, (4) GC content, and (5) inner lines showing syntenic blocks within the feral genome.
(B) Genome collinearity between the feral and domesticated ramie assemblies.
(C) Donut chart showing the distribution of variants between feral and domestic ramies. Numbers in parentheses (x/y) indicate x, the total length of each type of variation, and y, the number of events.
(D) Syntenic analyses between the assemblies of domesticated (reference) and feral (query) ramies; syntenic regions and SVs are highlighted with different colors.

(SVs) examined, highly diverged regions (HDRs) affected the largest portion 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) (Figure 1C; Supplemental Table 5).

Genome-wide variation and population structure

We sequenced 915 ramie individuals (Figure 2A and Supplemental Figure 4), with an average sequencing depth of 31.4× (Supplemental Table 6). 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 (Supplemental Table 7). After filtering (see materials and methods), 1 260 336 SNPs were retained.

Maximum likelihood and neighbor-joining approaches produced similar topologies (Figure 2B and Supplemental Figures 5 and 6). 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 comprised 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 (Supplemental Figure 5). These two accessions were morphologically similar to *Archiboehmeria*, a monotypic genus dubiously distinct from *Boehmeria* (Chen, 1980), so we removed them from subsequent analyses. Group II comprised accessions genetically more similar to domesticated

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

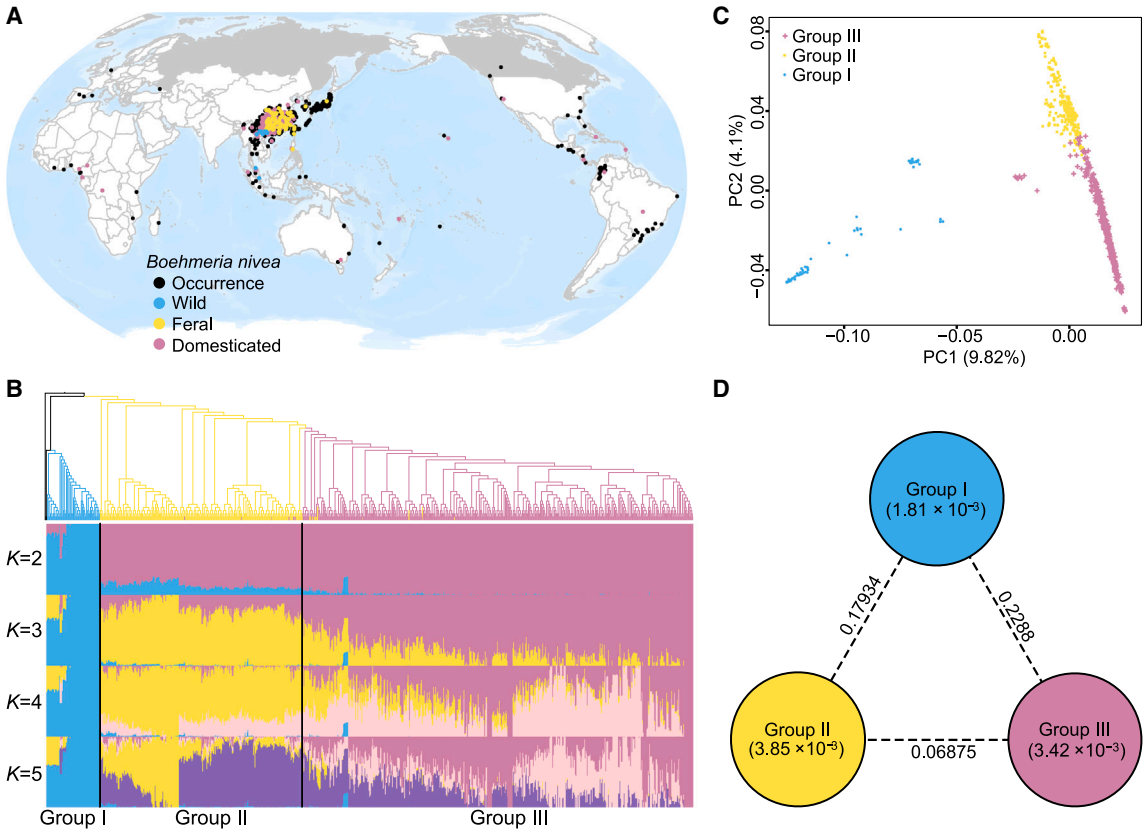

Figure 2. Population structure and genetic diversity of ramie.

(A) Geographic distribution of *Boehmeria nivea* based on occurrence points from GBIF (black circles). Sampling sites for this study are shown as blue (wild), yellow (feral), and purple (domesticated) circles.

(B) Admixture analyses with different numbers of groups (K = 2–5). Each vertical bar represents one ramie accession, and the x axis shows the three genetic groups. Each color represents one putative ancestral background, and the y axis quantifies ancestry membership.

(C) Two-dimensional PCA plot showing the clustering of accessions color coded as in **(B)**.

(D) Nucleotide diversity ($\theta\pi$) within and genetic differentiation (F_{ST}) between the groups.

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 11 feral accessions from around the world. Group II was paraphyletic with respect to group III (Figure 2B and Supplemental Figure 5).

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 (Figure 2C). These results were concordant with the phylogenetic results and indicate a relatively deep divergence between wild ramie and the other accessions, whereas feral and domesticated ramie grade into one another.

In the admixture analysis, cross-validation error decreased continuously as the number of subpopulations, K, increased, with no clear optimal K (up to K = 10; Supplemental Figure 7). 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 (Figure 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) (Figure 2D). Genetic differentiation (F_{ST}) was greatest between the wild and domesticated groups, intermediate between the feral and wild groups, and lowest between the domesticated and feral groups (Figure 2D).

Demographic and divergence histories

We used a supervised machine learning algorithm (DIYABC Random Forest) (see materials and methods) to test different hypotheses concerning the origin of feral ramie. Whether we considered feral ramie as a whole (Supplemental Table 8) or treated the two largest monophyletic subclades of feral ramie as discrete populations (supplemental Table 9), under the best scenarios, feral ramie was shown to be a product of hybridization between wild and domesticated ramie (Figure 3A and 3B and Supplemental Figure 8). We describe the results

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

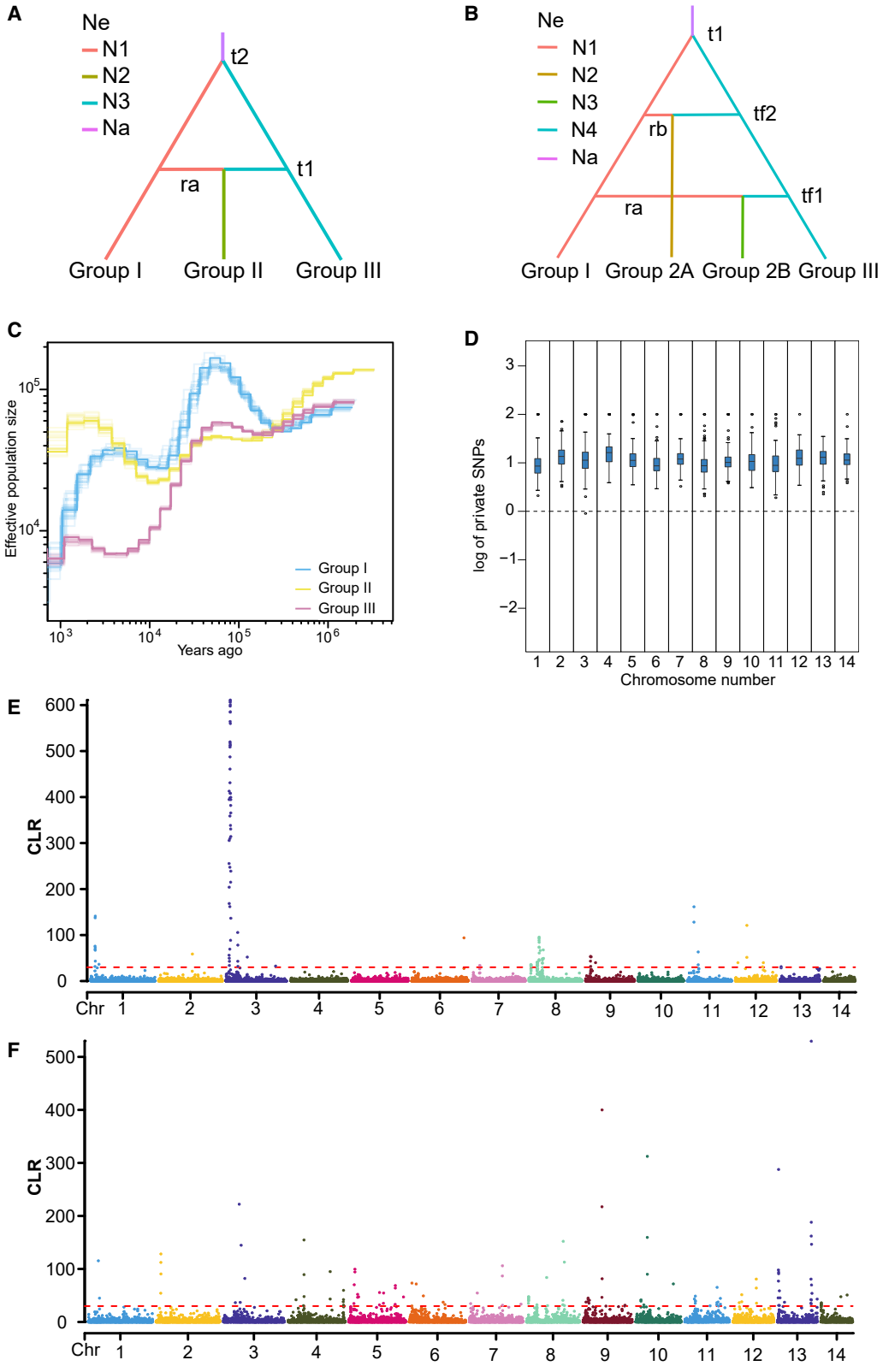

Figure 3. Demographic history and candidate genomic regions with evidence for selective sweeps between groups.

(A) Best scenario when feral ramie is considered as a whole.

(B) Best scenario when the two largest monophyletic subclades of feral ramie are treated as discrete populations.

(legend continued on next page)

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

here entirely based on division into three groups (Figure 3A; Supplemental Table 10). Groups I and III are estimated to have diverged 8678 years before present (95% quantile: 4181–10 800), indicating the initial stages of ramie domestication. Group II is estimated to have originated as a product of admixture between groups I and III (Figure 3; Supplemental Table 10) 5095 years before present (95% quantile: 1677–8967), with a smaller portion of the admixture being from group I (the wild group; 0.24; 95% quantile: 0.03–0.88) than from group III (the domesticated group; 0.76).

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

We also used "*GONE*" to examine very recent demographic history and obtained very different demographic trajectories for group III (domesticated) and II (feral) lineages but a relatively stable trend for group I (wild) populations (Supplemental Figure 9B). The population sizes of the domesticated and feral lineages started to decline ~150 generations ago; the feral lineage exhibited a very gradual decline, whereas for the domesticated lineage this was sharp and occurred in two steps. This is probably related to the continuous reduction in ramie cultivation over this period, especially in China.

Admixture signal detection

Admixture signal 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 (Supplemental Table 11).

To assess the ancestry of feral ramie compared with the ancestral populations (domesticated and wild ramie), we identified SNPs that were present in one or more accessions of domesticated ramie but not 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 with 9.3% with the wild accessions. This pattern was evident across all 14 chromosomes (Figure 3D; Supplemental Table 12), apart from one genomic region that had more wild than domesticated SNPs. Thus, both the DIYABC analysis and the admixture analysis indicate 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 (Figure 3E and 3F; Supplemental Tables 13 and 14), and we performed GO and KEGG enrichment analysis for the genes in these putative regions. In feral ramie, GO enrichment analysis revealed 72 enriched terms, including terms related to metabolic processes, cellular processes, and binding (Supplemental Table 15), whereas KEGG enrichment analysis identified 17 terms (Supplemental Table 16). Most of these items have relationships with stress tolerance. For example, ABC transporters (ko02010) are related to resistance to heavy metal pollution (Wang et al., 2015; Xu et al., 2020). In domesticated ramie, GO enrichment analysis revealed 97 enriched terms (Supplemental Table 17), and KEGG enrichment analysis revealed 12 significant terms (Supplemental Table 18). Most of these terms were 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) may 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 (Figure 3E and 3F), indicating 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 niches of ramie groups and to identify candidate ecological factors associated with habitat-specific adaptation during feralization. Empirically observed values of Hellinger's *I* and Schoener's *D* were significantly lower than those expected from pseudoreplicated datasets in paired analyses between groups I and II (wild and feral) and between groups I and III (wild and domesticated) (Figure 4A and 4B), 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 (Figure 4C). Niche overlap was greatest

(C) Demographic history of wild (group I), feral (II), and domesticated (III) ramies using MSMC2. The y axis represents inferred effective population size over time, and the x axis represents time.
(D) Distribution of wild- and cultivar-specific SNPs for each chromosome in feral material based on log10 of the ratio of crop-specific to wild-specific SNPs in 100-kb regions. The box shows the 95% confidence interval, and the black bar within each box is the mean. The horizontal dotted line represents zero, and positive and negative values represent excesses of domestic-like and wild-like SNPs, respectively.
(E) Distribution of the regions under selection in feral ramie.
(F) Distribution of the regions under selection in feral ramie, with horizontal dotted lines representing the cutoff fulfilling the requirement for the selected regions.

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

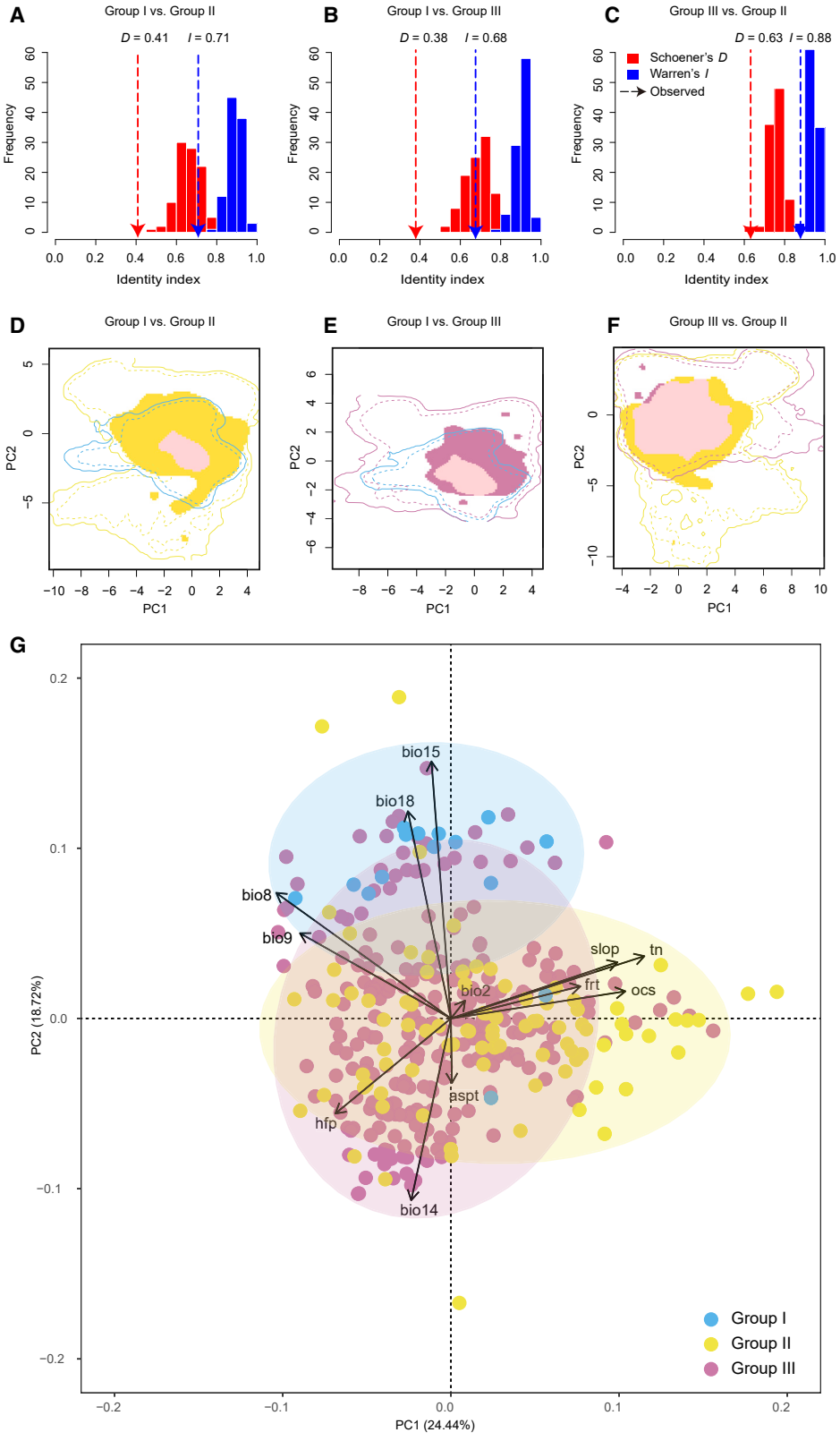

Figure 4. Results of ecological analyses.

(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 ($p < 0.01$).

(legend continued on next page)

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

between groups II and III (D = 0.63), with niches shared between groups accounting for 86.8%, whereas overlap was lowest between groups I and III (D = 0.38), with shared niches accounting for only 37% (Figure 4D and 4F; Supplemental Table 19). 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), whereas PCA2 was correlated with precipitation variables, including the precipitation of the driest month (bio14) and warmest quarter (bio18) and precipitation seasonality (bio15) (Figure 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 that in group II or III (Supplemental Table 20). All 12 environmental variables investigated had a statistically significant phylogenetic signal (Supplemental Table 21), 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 (Manel et al., 2018; Grummer et al., 2019). The results revealed 8 regions (at $-\log_{10}(p) > 7.83$) significantly associated with 3 of the 12 environmental variables in feral ramie: mean temperature of the wettest quarter (bio16), precipitation of the driest month (bio14), and total nitrogen (Figure 5A and Supplemental Figure 10; Supplemental Table 22). In total, 13 genes were identified, and the largest number were related to temperature (bio8) (Supplemental Table 22). For example, Bnt01G001074 on chromosome 1 is involved in the blue-light signaling pathway (GO:0009785) and plant circadian rhythms (KEGG: ko04712), which are proposed to be associated with temperature adaptation (Ben Michael et al., 2020). Other examples include Bnt12G017285 on chromosome 12 with transmembrane transporter activity (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 the response to 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 scenarios, we carried out ecological niche modeling (ENM). The results showed that both wild and feral ramies had an area under the receiver operating characteristic curve (AUC) value of $\geq$0.9 (Supplemental Table 23), indicating a better-than-random prediction. The suitable area was greatly influenced by climate change (Figure 5B–5E). The potential suitable area for wild ramie was greater in the Last Interglacial than in the Last Glacial Maximum (LGM) and the present, and it was predicted to increase in the future (2090). The area suitable for feral ramie was predicted to remain stable through 2090 (Figure 5F and 5G).

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, southwestern Guangxi, and the Indo-Chinese Peninsula. These are all places where ramie is cultivated, so it seems likely that ramie 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 (Xie et al., 2021; He et al., 2023). This might explain why wild ramie has lower genetic diversity than the other two groups. Our data show that wild ramie is genetically distinct from feral and domesticated 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) (Figure 2B and Supplemental Figure 5). 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 (Supplemental Figure 5). *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, not the ancestor of, domesticated ramie (Supplemental Figure 5).

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 reasoning, it appears that basal populations in group II (Supplemental Figure 5), which are mainly from Jiangxi, Guangdong, and Guangxi provinces and form a subgroup at K = 5 (Figure 2B and Supplemental Figure 7), may represent the earliest ramie feralization events. This suggests that these regions were the first places where ramie was cultivated and are the likely region of its domestication, largely consistent with a previous study based on nuclear SSR marker analysis, which suggested 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 (Larson et al., 2014; Li et al., 2019; Groppi et al., 2021), 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 (Supplemental Figure 5).

(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 groups I, II, and III, respectively. Pink in each figure indicates 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. For details of the variables, refer to Supplemental Table 25.

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

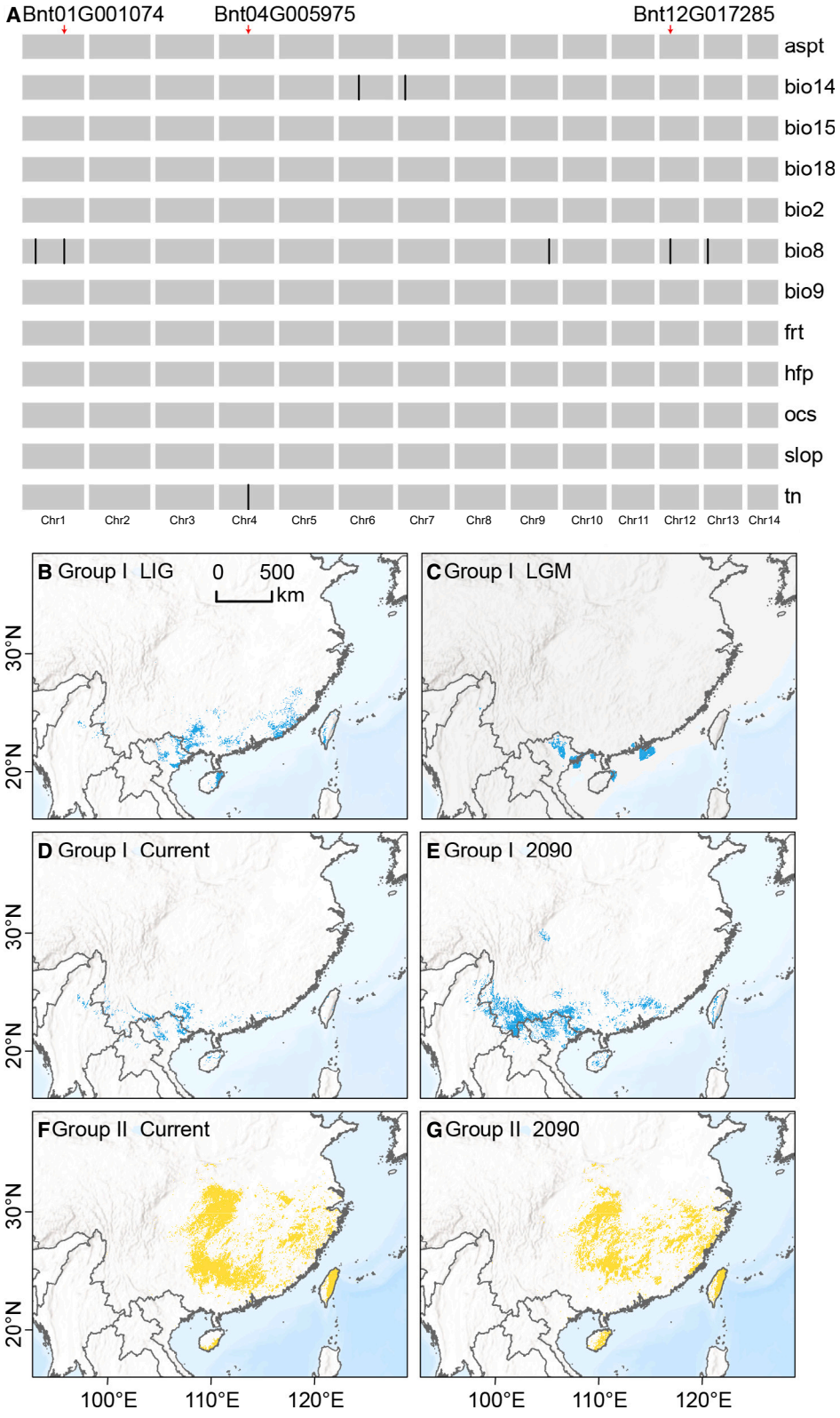

Figure 5. Potential range shifts and genome–environment associations.

(A) Results of genome–environment association analysis. Genomic locations of SNPs associated with environmental factors are shown, and 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)** estimated by ENM using 12 environmental variables and species occurrence points.

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

All feral accessions from Japan and Korea were grouped with feral populations from Zhejiang, Anhui, and northern Jiangxi (Supplemental Figure 5), suggesting that these locations were the source of the Japanese and Korean accessions, likely driven by human migration and maritime trade. This disjunct 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 a large amount of recent trade likely occurred between these regions around that time (Roy and Lutfar, 2012). Taiwanese indigenous people 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 accessions from Taiwan (Supplemental Figure 5), 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 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 (Thurber et al., 2010; Dwivedi et al., 2023). 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 (Supplemental Figure 5). *B. nivea* var. *tenacissima* shares with var. *strigosa* a branched stem, a partly connate stipule, and mostly green abaxial surfaces of the leaf blades (Figure 6), and these characteristics in var. *tenacissima* may be atavistic and/or due to crossing with var. *strigosa*. Other features, such as an assurgent or appressed strigose stem, differentiate var. *tenacissima* from var. *strigosa*. Most accessions identified as var. *nivea* in group II are more similar to 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, by contrast, some feral populations of both corn and rice were found to have lower genetic diversity than crop populations (Vigueira et al., 2013; Qiu et al., 2017). Our observation is best explained by feral ramie populations expanding their gene pools via hybridization from wild material and/or landraces. Admixture signal detection revealed 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 (Figure 3).

However, there is little overlap between the geographical ranges of var. *strigosa* and feral material (Supplemental Figure 4), 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 (Figure 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, and feral ramie is therefore most probably of exoferal or exo-endoferal origin (Wu et al., 2021). Crucially, feral ramie may contain genetic diversity that may be of use in ramie breeding going forward.

Demographic analysis revealed that the ancestors of wild, feral, and domesticated ramie lineages all exhibited a parallel reduction in *Ne* from 48 to 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 onward, is consistent with anthropogenic destruction of its habitat (Xiao et al., 2018; Xie et al., 2021). The dramatic reduction in *Ne* for domesticated ramie ~4.2–3 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 (Figure 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 of feral ramie.

Investigations identifying loci involved in domestication and their significance for feralization have been carried out in many animal taxa, but such studies are limited in plants. Genome scans have become routine and offer a potential means 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 determine whether any of these SVs demonstrate fixed differences among wild, feral, and/or domesticated populations. Analysis of selective sweeps revealed that genomic regions targeted by the domestication and 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, we generated and explored the largest genomic resource for ramie to date, revealing its domestication

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

Figure 6. Morphological comparisons among three varieties.
(A–C) *Boehmeria nivea* var. *nivea*; **(A)** habit with unbranched stem, **(B)** white abaxial leaf blade, and **(C)** free and patent hirsute stem.
(D–F) *B. nivea* var. *strigosa*; **(D)** habit with branched stem, **(E)** green abaxial leaf blade, **(F)** and 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, and **(I)** appressed hirsute and partly connate stipule.

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

and feralization history and the genetic basis of environmental adaptation for feral ramie. Our results not only provide evidence that feral ramie can be a source material for improvement of 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, and more work is needed to examine the molecular genetic basis of fitness-related phenotypes in feral settings. Likewise, the universality of evolutionary mechanisms during feralization needs to be examined in more plant species.

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 its cultivation history is most ancient, was extensively sampled across 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, and all three varieties of *B. nivea* (vars. *nivea*, *tenacissima*, and *strigosa*) were comprehensively sampled (Figure 2A and Supplemental Figure 4). Among sampled material, the term "wild" is used exclusively to refer to a 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 stages (Wu et al., 2021), all ramie referable to var. *tenacissima* or var. *nivea* growing in the wild without human control are considered feral in our study. Moreover, the term "landrace" encompasses a range of different concepts that have varied over time (Casañas et al., 2017). Our study followed the landrace definition of Villa et al. (2005), Zeven (1998), and Dwivedi et al. (2016), i.e., a dynamic population of a cultivated species that has a historic origin and distinct identity, lacks formal crop improvement, and is often genetically diverse, locally adapted, and associated with traditional farming systems or a low input agriculture system. Because most farmers in China have given up growing ramie (see introduction), there is little domesticated ramie on farms, and samples of all cultivars and most landraces were acquired from the National Infrastructure for Bast Fiber Crop Germplasm Resources of China (Supplemental Table 6).

No feral ramie genome has been reported to date, although three whole genomes of cultivated ramie have been reported (Wang et al., 2021; Chen et al., 2023). In this study, we collected for this purpose fresh material from a feral adult (lab no. HZS10, Supplemental Table 6) in Shennong Valley National Forest Park, Hunan Province, China (26.503°N, 114.001°E). Living collections and seeds of this individual are preserved in the Germplasm Bank of Wild Species, Kunming Institute of Botany, CAS.

Genomic DNA extraction and sequencing

Genomic DNA was extracted from the leaves of feral ramie HZS10 using a modified CTAB method. The quality of the extracted DNA was examined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and its quantity determined by electrophoresis on a 0.8% agarose gel. Illumina sequencing libraries were generated using the VAHTS Universal DNA Library Prep Kit for MGI (Vazyme, Nanjing, China) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The library was quantified using a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA) and a Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). Finally, the MGI-SEQ 2000 platform was used to generate 12.7 Gb of paired-end sequencing data. 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 19.74 Gb of PacBio long-read data. A Hi-C library was constructed and sequenced on the MGI-SEQ 2000 platform for chromosome-level scaffolding, generating 156.73 million paired-end reads and 46.33 Gb of sequencing data.

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

Genome *de novo* assembly and annotation

To estimate the genome size of individual HZS10, the raw Illumina short reads were pre-processed to remove adaptors and low-quality bases using SOAPnuke (Chen et al., 2018b) with default settings, and the clean data were used to determine k-mer distributions with 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 paired-end short reads were used to correct systematic errors in the PacBio contigs using Pilon (Walker et al., 2014). To anchor the corrected contigs onto 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 using 3D-DNA (Dudchenko et al., 2017). The completeness and accuracy of genome assembly were quantitatively assessed using BUSCO (Simão et al., 2015) with the eudicotyledons_odb10 gene set.

For annotation of repetitive sequences, two methods were used 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 performed using

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

the TE protein database as a query library. Second, we used *de novo* prediction; 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). We also 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 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: *ab initio*, homology-based, and RNA-seq-aided gene prediction. For *ab initio* prediction, we used the gene predictor software Augustus (version 3.3.1) (Stanke et al., 2006) and Genescan (Burge and Karlin, 1997). Models used for each gene predictor were trained on a set of high-quality proteins generated from the RNA-seq dataset. Homology-based gene prediction was performed 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 with SOAPnuke (Chen et al., 2018b), then assembled the clean RNA-seq reads into transcripts using Trinity (Grabherr et al., 2011), and finally defined the gene structures using PASA (Haas et al., 2003). 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 annotation of non-coding RNAs (rRNAs, small nuclear RNAs, 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 genes associated with tRNAs.

For functional annotation of protein-coding genes, BLASTP was used to align the feral ramie protein sequences with those in public databases, including the NCBI, NR, TrEMBL, InterPro, Swiss-Prot, and KEGG databases, using 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 using PFAM databases. GO terms for each gene were obtained using Blast2GO (Conesa and Götz, 2008).

Synteny analysis and comparative genomics

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

To identify structural variants (SVs) between the feral and domesticated assemblies, we performed comparative genomics analysis. 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 the parameter setting "-ax asm20 -eqx". SyRI (Goel et al., 2019) (-k -F S) was used to identify structural rearrangements and local variants between the two genomes. All identified variants were annotated using the SnpEff program (Cingolani et al., 2012) with the parameter -ud 2000, and a dot plot was drawn using plotsr software (Goel and Schneeberger, 2022) with parameters -m 20000 -x -q 500000 -s -t.

Variant calling and filtering

Genome resequencing was performed for 915 ramie accessions (Supplemental Table 6) and an outgroup using the methods described above, but with the Illumina NovaSeq platform. Raw data were quality checked and then filtered using fastp (version 0.20.0) (Chen et al., 2018a). Clean paired-end reads from each accession were mapped to the latest reference genome of domesticated ramie (Qingyezhuma) (Wang et al., 2021) using the Burrows–Wheeler Aligner (Li and Durbin, 2010) with default parameters. After alignment, Picard (version 2.18.17, http://broadinstitute.github.io/picard/) was used to mark duplicate reads, and SAMtools (Li et al., 2009) was used to convert the 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 parameters QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, and indels were filtered with the parameters QD < 2.0, FS > 200.0, MQ < 40.0, SOR > 10.0, ReadPosRankSum < −20.0. We then 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 a 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 using the parameters *indep-pairwise 50 5 0.5*. In total we retained 1 260 336 SNPs. ADMIXTURE (Alexander et al., 2009) was then used to infer the optimum number of clusters (*K*) among all ramie accessions. *K* values from 2 to 10 were examined, and the cross-validation error was calculated to identify the most likely number of clusters. A PCA was performed using EIGENSOFT (Price et al., 2006). To infer relationships among accessions, two kinds of rooted phylogenetic trees were reconstructed. First, using the same 1 260 336 SNPs, a neighbor-joining phylogenetic tree was obtained by calculating the pairwise genetic distances using PLINK (Purcell et al., 2007) and constructing the tree using PHYLIP (Retief, 2000). Second, a maximum likelihood tree was constructed based on four-fold-degenerate sites in the 915 ramie accessions. SNPs were extracted and compared with the 7 460 735 four-fold degenerate sites identified in the ramie genome using iTools (20180520) (Dinov et al., 2008). SNPs from each individual were merged into one file using mafft (version 7.407) (Katoh and Standley, 2013), and low-quality regions were trimmed with trimAl (version 1.4.rev22) (Capella-Gutiérrez et al., 2009). The 120 201 SNPs were then used to construct a rooted maximum likelihood tree using IQ-TREE (version 1.6.12) (Nguyen et al.,

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

2015) with the parameters -alrt 1000 -bb 1000 (ultrafast bootstrap). *Girardinia diversifolia* (sample ID W1000) was used as an outgroup.

On the basis of population structure and each individual's habitat (see results), we defined three groups of individuals. Group I included only wild individuals and was distinct from all feral and domesticated material. Group II comprised all but 11 feral individuals plus 9 domesticated accessions; this group was genetically similar to domesticated material, but with an apparently admixed genomic composition. Group III comprised the vast majority of cultivated landraces and modern cultivar accessions from the National Infrastructure for Bast Fiber Crop Germplasm Resources of China, as well as 11 feral individuals from around the world. Overall, our dataset comprised 552 group III accessions (primarily domesticated), 286 group II accessions (primarily feral), and 77 group I accessions (all wild).

Diversity statistics, 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 that included 522 accessions (51, 144, and 327 individuals, respectively, from groups I, II, and III; Supplemental Table 24).

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 demographic analyses, we first used MSMC2 (Schiffels and Wang, 2020), which has advantages for 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 we then used SHAPEIT4 (Delaneau et al., 2019) to phase each chromosome. MSMC-tools (https://github.com/stschiff/msmc-tools) were used to generate input files for MSMC2 for each chromosome. The average generation time was set to 1 year, and the mutation rate was assumed to be $\mu = 1.5 \times 10^{-8}$ mutations × bp^{-1} × $generation^{-1}$ (Koch et al., 2000). Next, demographic history was 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 has been found to be accurate for up to at least the most recent 200 generations.

To test alternative evolutionary scenarios for the origin of feral ramies and their relationships to wild and domesticated ramies, we used Approximate Bayesian Computation and supervised machine learning methods implemented in DIYABC-RF v1.0 (Collin et al., 2021). In one analysis, group II was treated as a whole, and in another, the two largest monophyletic groups of individuals were treated as separate groups (subclades 2A and 2B in Supplemental Figure 5). 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 6 models (Supplemental Figure 8A) and 8 models (Supplemental Figure 8B), respectively. For all scenarios, training sets were generated using 4000 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 (Supplemental Table 10). Following the recommendations in the manual and the RF algorithm for model choice based on linear discriminant analysis, we used 5 noise variables and generated 2000 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 were identified in the feral group (group II) and the domesticated group (group III) using SweeD (version 4.0.0) (Pavlidis et al., 2013). Genome-wide SNPs were trimmed with the parameter setting "-maf 0.05, -missing 0.1", and an empirical estimate of effective population size derived from the MSMC2 analysis described above was incorporated. Composite likelihood ratios were calculated in windows with an average size of 10 kb across the genome by setting grid numbers according to chromosome lengths (number of grids = chromosome length/10 000). Those with the top 5% highest composite likelihood ratio values were identified as potential selective sweep regions, and sweep regions with a physical distance no larger than 100 bp were merged. Biological functions were assigned to candidate genes within these genomic regions using annotations from the KEGG and GO databases, and statistically enriched terms were identified.

Admixture detection and genomic composition of feral ramie

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

To assess the genomic composition of feral ramie compared with domesticated and wild ramies, we identified a subset of SNPs that were present in one or more accessions of domesticated material but were not detected in wild material, which we termed "crop-specific private SNPs." Likewise, SNPs detected in wild but not domesticated 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 them 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 indicated that there were more wild-specific than domestic-specific private SNPs within a genomic window.

Ecological analyses

We used several ecological analyses to reveal differences in the niches of ramie groups 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 were mainly collected from herbarium specimens; we therefore used only samples collected in China for this analysis. Using the R package spThin (Aiello-Lammens et al., 2015), we only retained records of the same groups that were separated from each other by $\geq$5 km, and the final dataset

Please cite this article in press as: Wu et al., Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop, Plant Communications (2024), https://doi.org/10.1016/j.xplc.2024.100942

thus 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) databases, which together included bioclimatic, topographical, pedologic, and anthropogenic variables (Supplemental Table 25). To reduce collinearity among environmental variables, we used the R package usdm (Naimi et al., 2014) to keep only those variables with VIF < 5, which resulted in 12 environmental variables being retained (Supplemental Table 25).

Using these data, three kinds of analysis were used to study niche differentiation among the three groups: (1) using the 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*, in which a value of 0 suggests no overlap and 1 indicates complete overlap; (2) to quantify the 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 among these three groups using ANOVA.

Genome–environment association analyses were performed with PCA controlled as fixed effects using EMMAX (Zhou and Stephens, 2012), taking environmental data as phenotypes (Supplemental Table 26) and using 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).

Using the 12 variables, we predicted the potential geographic distributions of wild and feral ramie under past and future climate scenarios. For wild ramie, we studied the potential distribution during the Last Interglacial, LGM, the present, and the future. For feral ramie, ENM was carried out only for the present and the future. For the future, we took the year 2090 under the pessimistic RCP8.5 scenario (IPCC, 2013). For each sample location, ENM was performed using the biomod2 R package (Thuiller et al., 2009); we used an ensemble of six models (GBM, CTA, FDA, MARS, RF, and MAXNET) with 10 bootstrap replicates, using 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.

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 calculated Blomberg's *K* using the *multiphylosignal* function in the R package picante (Kembel et al., 2010).

DATA AND CODE AVAILABILITY

The genome sequence data for feral ramie reported in this paper have been deposited in the Genome Warehouse at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession number GWHERBU00000000 (BioProject PRJCA015489) and are publicly accessible at https://ngdc.cncb.ac.cn/gwh. The raw resequencing data for 915 individuals reported in this paper have been deposited in the Genome Sequence Archive at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA011837 and CRA010145) under project accession number PRJCA015489 and are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This study was supported by the CAS Strategic Priority Research Program (XDB31000000), the National Natural Science Foundation 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, the CAS (ZDBS-LY-7001), the CAS "Light of West China" Program (to Z.-Y.W. and J.L.), the Applied and Fundamental Research Foundation of Yunnan Province (202401AT070190), the CAS Youth Innovation Promotion Association (2019385), and the Central Public-interest Scientific Institution Basal Research Fund (Y2023PT11). R.M. and M.C. also thank the CAS President's International Fellowship Initiative for financial support (2022VBA0004 and 2020VBB0016, respectively).

AUTHOR CONTRIBUTIONS

Z.-Y.W., D.-Z.L., J.L., and M.-B.L. conceived the study. Z.-Y.W., J.L., Y.Z., and M.-B.L. did field work. A.K.M. helped collect most samples outside of China. Z.-Y.W. and Y.Z. carried out lab work. Z.-Y.W., J.L., M.A.C., Y.-H.L., G.-F.Z., P.-Z.F., and Z.-P.L. performed data analyses. Z.-Y.W. organized the data and wrote the first draft. M.A.C., R.T.C., R.I.M., and M.K.C. helped improve the focus and discussion. All authors revised and approved the final manuscript.

ACKNOWLEDGMENTS

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 contributions 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. No conflict of interest is declared.

Received: November 19, 2023
Revised: December 20, 2023
Accepted: May 6, 2024

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 Res. **19**:1655–1664.

Bao, Z., and Eddy, S.R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. **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 Syst. Sci. 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 Biol. **20**:77.

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

Blomberg, S.P., Garland, T., Jr., 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. J. Mol. Biol. **268**:78–94.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. **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. Front. Plant Sci. **8**:145.

Chen, C.J. (1980). *Archiboehmeria* C. J. Chen–a new genus of Urticaceae. Acta Phytotaxon. Sin. **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, eds. (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., Gao, G., et al. (2023). The chromosome-level assembly of ramie (*Boehmeria nivea* L.) genome provides insights into molecular regulation of fiber fineness. J. Nat. 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. Sci. Data **7**:320.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **Chapter 4**. Unit.4.10–14.10.14.

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 Sci. Chian **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. Nat. Methods **18**:170–175.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., 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}. 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. Mol. Ecol. Resour. **21**:2598–2613.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genom. **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, 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. Nat. Commun. **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., Aiden, A.P., et al. (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.P., 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 Syst. **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. Front. Genet. **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 Plant Sci. **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. Evol. Appl. **3**:494–504.

Farrant, J.M., and Hilhorst, H. (2022). Crops for dry environments. Curr. Opin. Biotechnol. **74**:84–91.

Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. **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 Ecol. Evol. **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 Biol. **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. Nat. Biotechnol. **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. Nat. Commun. **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 Ecol. Evol. **34**:641–654.

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

Gutaker, R.M., and Purugganan, M.D. (2024). Adaptation and the geographic spread of crop species. Annu. Rev. Plant Biol. **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. Curr. Opin. Plant Biol. **66**, 102169.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., 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 Res. **31**:5654–5666.

He, X., Ziegler, A.D., Elsen, P.R., Feng, Y., Baker, J.C., 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 Biol. Evol. **11**:2045–2054.

Hester, S.B., and Yuen, M.L. (1989). Ramie: Patterns of world production and trade. J. Textil. Inst. **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. Int. J. Mol. Sci. **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.). Front. Plant Sci. **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. Cytogenet. Genome Res. **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 Res. **46**:D335–D342.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. **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). Mol. Biol. Evol. **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.

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 Res. **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., Dobney, K., et al. (2014). Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. USA **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., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (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., Zhang, B., et al. (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. Nat. Genet. **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 Biol. **20**:36.

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

Liao, L., Li, T., Zhang, J., Xu, L., Deng, H., and Han, X. (2014). The domestication and dispersal of the cultivated ramie (*Boehmeria nivea* (L.) Gaud. Freyc.) determined by nuclear SSR marker analysis **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. Preprint at arXiv. https://doi.org/10.48550/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 Prod. Sci. **12**:88–96.

Liu, X., and Fu, Y.X. (2020). Stairway Plot 2: Demographic history inference with folded SNP frequency spectra. Genome Biol. **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 Res. **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 Genet. **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). Mol. Biol. Evol. **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. Mol. Ecol. **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., Daly, M., et al. (2010). The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. **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. Nat. Genet. **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 Bioinf. **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., and Minh, B.Q.; evolution (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **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., Živkovic, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. **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 (Bethesda). **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 Cell Physiol. **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. Nat. Genet. **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., Sham, P.C., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575.

Purugganan, M.D. (2022). What is domestication? Trends Ecol. Evol. **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. Nat. Commun. **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 Biol. **21**:70.

Retief, J.D. (2000). Phylogenetic analysis using phylip. In Bioinformatics Methods and Protocols, S. Misener and S.A. Krawetz, eds. (Humana Press), pp. 243–258.

Roy, S., and Lutfar, L.B. (2012). Bast fibres: Ramie. In In Handbook of Natural Fibres: Types, Properties and Factors Affecting Breeding and Cultivation, R.M. Kozłowski, ed. (Woodhead Publishing), 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. Mol. Biol. Evol. **37**:3642–3653.

Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The multiple sequentially markovian coalescent. In Statistical Population Genomics (Human Press), 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 Bioinf. **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 Res. **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 Ethnol. Stud. **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. Nat. Genet. **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. Mol. Ecol. **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. Sci. Data **3**, 160067.

Vigueira, C.C., Olsen, K.M., and Caicedo, A.L. (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 Genet. Resour. **3**:373–384.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (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 Res. **38**:e164.

Wang, L., Yang, H., Liu, R., Fan, G., and Research, P. (2015). Detoxification strategies and regulation of oxygen production and flowering of *Platanus acerifolia* under lead (Pb) stress by transcriptome analysis. Environ. Sci. Pollut. Res. Int. **22**:12747–12758.

Wang, Y., Li, F., He, Q., Bao, Z., Zeng, Z., An, D., Zhang, T., Yan, L., Wang, H., Zhu, S., et al. (2021). Genomic analyses provide comprehensive insights into the domestication of bast fiber crop ramie (*Boehmeria nivea*). Plant J. **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. Commun. Biol. **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 Plant Sci. **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. 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. Veg. Hist. Archaeobotany **30**:477–488.

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

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

Yu, H., and Li, J. (2022). Breeding future crops to feed the world through de novo domestication. Nat. Commun. **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. Nat. Commun. **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. Nat. Commun. **11**:671.

Zhao, Y., Milne, R., Li, Z., et al. (2024). *Boehmeria nivea* var. *strigosa* (Urticaceae), a new variety from Southwest China. Guihaia. https://doi.org/10.11931/guihaia.gxzw202311013.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **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. Plant J. **109**:402–414.