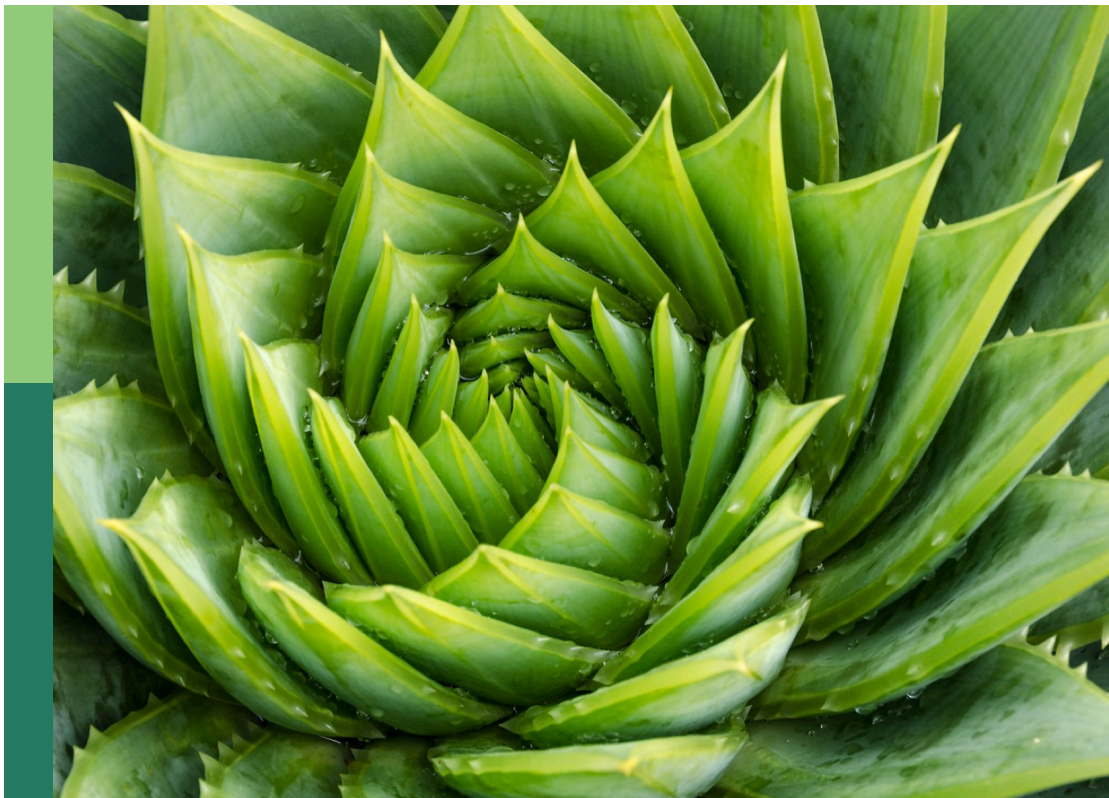
frontiers | Research Topics

Wild plant genetic resources: A hope for tomorrow

Edited by
Mohd. Kamran Khan, Francesco Di Gioia, Tofazzal Islam, Sait Gezgin and Hannah Schneider

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2737-5
DOI 10.3389/978-2-8325-2737-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Wild plant genetic resources: A hope for tomorrow

Topic editors

Mohd. Kamran Khan — Selçuk University, Türkiye

Francesco Di Gioia — The Pennsylvania State University (PSU), United States

Tofazzal Islam — Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

Sait Gezgin — Other Konya, Türkiye

Hannah Schneider — Wageningen University and Research, Netherlands

Citation

Khan,M. K., Di Gioia, F., Islam, T., Gezgin, S., Schneider, H., eds. (2023). *Wild plant genetic resources: A hope for tomorrow*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-2737-5

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 31 May 2023
DOI 10.3389/fpls.2023.1217547

OPEN ACCESS

EDITED AND REVIEWED BY
Roger Deal,
Emory University, United States

*CORRESPONDENCE
Mohd. Kamran Khan
mohdkamran.biotech@gmail.com;
mkkhan@selcuk.edu.tr

RECEIVED 05 May 2023
ACCEPTED 22 May 2023
PUBLISHED 31 May 2023

CITATION
Khan MK, Islam T, Gezgin S and Di Gioia F (2023) Editorial: Wild plant genetic resources: a hope for tomorrow. *Front. Plant Sci.* 14:1217547. doi: 10.3389/fpls.2023.1217547

Editorial: Wild plant genetic resources: a hope for tomorrow

Mohd. Kamran Khan[1]*, Tofazzal Islam[2], Sait Gezgin[1] and Francesco Di Gioia[3]

[1]Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Selcuk University, Konya, Türkiye, [2]Institute of Biotechnology and Genetic Engineering, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh, [3]Department of Plant Science, The Pennsylvania State University, University Park, PA, United States

KEYWORDS

climate change, crop wild relatives, endangered wild species, environmental stresses, nutritional development, molecular strategies, genetic variation, omics approach

Editorial on the Research Topic
Wild plant genetic resources: a hope for tomorrow

The demand of food is expected to increase significantly in the coming decades with growing population and shifting dietary patterns (Tilman et al., 2011). Crop yield is largely affected by various factors including biotic and abiotic stresses, reduced biodiversity, degraded soil, and the changing climate (Lobell and Gourdji, 2012; Islam et al., 2016; Zhu et al., 2022). Plant genetic resources including crop wild relatives (CWRs), landraces and underutilized crop species are a significant source of traits of agronomic interest (Khan et al., 2021; Renzi et al., 2022). Conserving and utilizing such genetic resources may be key for the development of climate resilient crop varieties and to ensure global food and nutrition security (Figure 1). The conservation of plant genetic resources and the availability of their omics data is critical for the improvement of crop varieties using advanced molecular breeding including genome editing (Islam, 2019; Salgotra and Chauhan, 2023). For the efficient utilization, genetic resources should be comprehensively examined, however, there has been a dearth in the assessment of their heritable traits and full characterization (Figure 1). Moreover, more studies are required on the assessment of genetic variability, level of tolerance against individual and combined biotic and abiotic stresses, yield performance and nutritional profile of wild plant genetic resources (Panwar et al., 2022; Pandey et al., 2023). Therefore, this Research Topic aimed to compile research updates on successful monitoring and utilization of wild plant genetic resources for modern crop improvement.

The research findings and critical reviews described in eight articles included in this Research Topic are briefly summarized and discussed in the following sections.

1 Steps toward efficient conservation of CWRs

Being an important source of traits for biotic and abiotic stress tolerance, and nutrients, crop wild relatives (CWRs) can largely contribute to the improvement of crops (Kahraman et al., 2017; Rajpal et al., 2023). However, these natural genetic resources are under continuous threat due to global climate change and anthropogenic activities. Therefore,

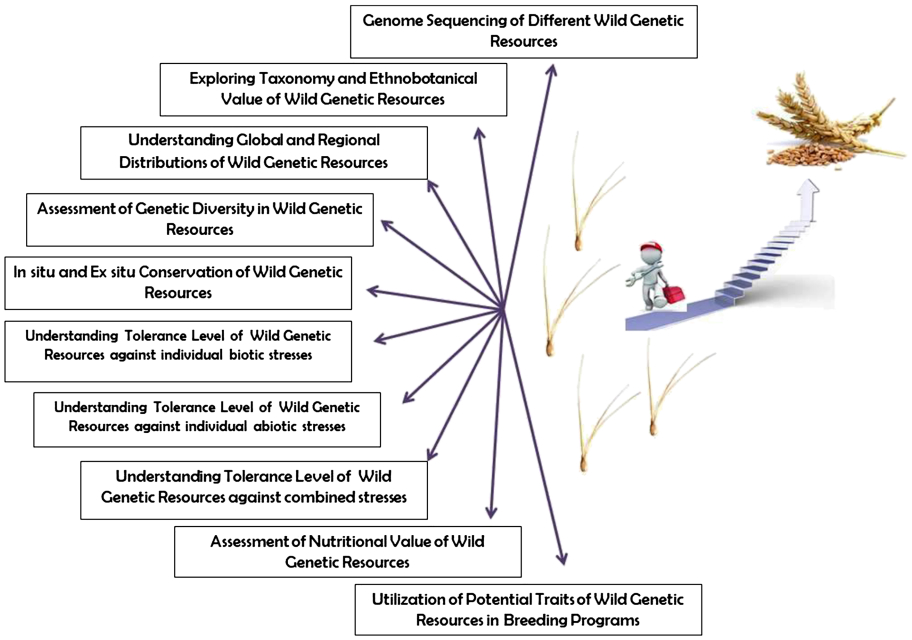


FIGURE 1
Outline of different aspects of wild genetic resources that need to be investigated further for the development of modern cultivars.

conservation of these valuable genetic resources must be prioritized. Brilhante et al. catalogued the diversity of CWRs of genus *Vigna* in Mozambique and provided conservation strategies for preferred targets and regions. This inclusive dataset containing information on species diversity, dissemination, and threats will facilitate the sustainable usage of *Vigna* CWRs in crop improvement. They suggested that climatic fluctuations and changes in agricultural practices in that region destruct natural habitats and threaten the existence of wild populations. Chen et al. explored the influence of land use change in China from 2001 to 2019 on the survival rate of three wild rice species (*Oryza rufipogon, O. officinalis*, and *O. granulata*) employing satellite-based Earth observations. Although the land use change had a suppressive effect on the population of three wild rice species, it has been suggested that vegetation surrounding the wild population acts as a biological barrier and protects the plants from the destruction due to land use change. The findings of their study emphasize the need of a modified conservation strategy for wild species of rice. Other than this, the loss of wild species because of land use change emphasizes the *in situ* conservation of wild relatives (Thingnam et al., 2023). However, genetic characterization of individual samples of such large populations is difficult due to requirement for the large resources. Thus, tolerance level of large populations of CWRs towards different stress conditions at difficult sites can be easily identified using predictive characterization techniques based on ecogeographic information. Civantos-Gomez et al. employed a machine learning-based predictive strategy to characterize the resistance toward rust disease caused by fungus *Uromyces viciae-fabae* in CWR populations of lentils in the Mediterranean basin. They concluded that rust resistant CWR populations are suitable candidates for *in situ* conservation, and specific environmental conditions have a role in developing resistance in them. Davis et al. provided a contemporary review summarizing taxonomy, conservation, and potential traits including agronomic performance and biotic/abiotic stress tolerance of four indigenous species of coffee in Uganda. This review not only identified that two of the studied coffee species are at the risk of extinction but also suggested to prioritize their conservation in Uganda.

2 Utilization of genetic resources to maintain diversity

An efficient way of utilizing CWRs is to improve the strategies of their selection based on phenotypic and genomic indices. Fenstemaker et al. utilized phenotypic, genomic, and combined strategies to select water-deficit stress tolerant lines developed from the crosses of tolerant wild tomato relative *Solanum galapagense* accession LA1141 and susceptible *Solanum lycopersicum* L. OH8245. Thermal images showed a greater phenotypic variance for canopy temperature trait in the progenies and quantitative trait loci (QTLs) contributing to water deficit tolerance which were mapped in LA1141. These findings suggest the opportunity to introgress water deficit tolerance trait from wild relatives to modern tomato cultivars. Moreover, understanding the material developed from the breeding of CWRs can facilitate its utilization for crop improvement. Sequencing of whole genomes and transcriptomes of wild genetic resources will be advantageous in this direction (Brozynska et al., 2016; Pandey et al., 2022; Khan et al., 2023). Jackfruit (*Artocarpus heterophyllus* Lam, Moraceae family) has attracted the attention of food experts and technologists due to its nutritional health benefits. However, it is an underutilized

and less explored tropical fruit crop, but a highly potential source of food and nutritional security. Islam et al. for the first time sequenced the whole genome of a year-round fruiting and high yielding jackfruit variety, BARI Kanthal-3, which was developed from a wild accession. Bioinformatics analysis of the sequence data identified the distribution of a large collection of nucleotide variation across the genome of jackfruit that can be used to identify new functional genes and their regulatory activities specific to BARI Kanthal-3. They also demonstrated that BARI Kanthal-3 has a higher number of genes related to flowering time. Their findings not only facilitate marker development for different traits in jackfruit crop to be utilized in breeding programs, but also increase the chances of their utilization to ensuring food supply by understanding their evolution and domestication process.

Reproductive isolation limits the utilization of wild species for the improvement of cultivated forms. Thus, other existing forms that are closer to widely utilized cultivated genotypes and that possess potential stress tolerance and yield related traits should be preferred for crop improvement. The Waxy (Wx) gene is found to be responsible for waxy trait in waxy rice that is high-quality rice with less than 2% of apparent amylose content (AAC) of the starch. Fu et al. developed wx mutants of rice varieties employing CRISPR-Cas9 gene-editing system that showed significant decrease in AAC. However, AAC content of rice genotypes with low initial AAC was further decreased on mutation of wx genes and thus, preference can be given to such genotypes in breeding programs. Li et al. studied a semi-domesticated rice called weedy rice (*Oryza sativa* f. *spontanea*) for the development of early heading trait in japonica rice. Four genes, two major (*Hd1* and *OsCCT22*) and two minor (*Dth7* and *Hd16*) were found to be regulating the early heading trait of weedy rice.

The articles published in this Research Topic address different aspects of the conservation and utilization of wild plant genetic resources, however, a consistent research effort is needed for the efficient conservation and utilization of the invaluable patrimony of wild plant genetic resources (Figure 1).

Author contributions

MKK conceived the idea of the Research Topic, wrote the editorial, sketched the figure and acted as editor of some of the manuscripts of the Research Topic. TI, SG, and FG, acted as editor of some of the manuscripts of the Research Topic. TI, SG, FG along with MKK edited and reviewed this editorial. All authors contributed to the article and approved the submitted version.

Funding

MKK and SG are thankful to TÜBİTAK (The Scientific and Technological Research Council of Türkiye) for supporting the research work related to this topic under the project no. 119O455. TI is thankful to Krishi Gobeshona Foundation, Bangladesh for funding to OFANS project. FG contribution was supported by the USDA NIFA and Hatch Appropriations under Project No. PEN04723 and Accession No. 1020664.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

Islam, T. (2019). CRISPR-cas technology in modifying food crops. *CAB Rev.* 14, 1–16. doi: 10.1079/PAVSNNR201914050

Islam, M. T., Croll, D., Gladieux, P., Soanes, D.M., Persoons, A., Bhattacharjee, P., et al. (2016). Emergence of wheat blast in Bangladesh was caused by a south American lineage of *Magnaporthe oryzae*. *BMC Biol.* 14, 84. doi: 10.1186/s12915-016-0309-7

Kahraman, A., Pandey, A., Khan, M. K., Lindsay, D., Moenga, S., Vance, L., et al. (2017). Distinct subgroups of *Cicer echinospermum* are associated with hybrid sterility and breakdown in interspecific crosses with cultivated chickpea. *Crop Sci.* 57, 3101–3111. doi: 10.2135/cropsci2017.06.0335

Khan, M. K., Pandey, A., Hamurcu, M., Avsaroglu, Z. Z., Ozbek, M., Omay, A. H., et al. (2021). Variability in physiological traits reveals boron toxicity tolerance in aegilops species. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.736614

Khan, M. K., Pandey, A., Hamurcu, M., Rajpal, V. R., Vyhnanek, T., Topal, A., et al. (2023). Insight into the boron toxicity stress-responsive genes in boron-tolerant triticum dicoccum shoots using RNA sequencing. *Agronomy* 13, 631. doi: 10.3390/agronomy13030631

Lobell, D. B., and Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiol.* 160, 1686–1697. doi: 10.1104/pp.112.208298

Pandey, A., Khan, M. K., Athar, T., Hamurcu, M., Germ, M., and Gezgin, S. (2023). "Chapter 17 - combined abiotic stresses in wheat species," in *Abiotic stresses in wheat*. Eds. M. K. Khan, A. Pandey, M. Hamurcu, O. P. Gupta and S. Gezgin (San Diego, CA, United States: Academic Press), 273–282.

Pandey, A., Khan, M. K., Hamurcu, M., Brestic, M., Topal, A., and Gezgin, S. (2022). Insight into the root transcriptome of a boron-tolerant triticum zhukovskyi genotype grown under boron toxicity. *Agronomy* 12, 2421. doi: 10.3390/agronomy12102421

Panwar, R., Chaudhry, B., Kumar, D., Prakash, G., Khan, M. K., Pandey, A., et al. (2022). Harnessing stress-tolerant wild bananas for crop improvement. *Crop Pasture Sci.* doi: 10.1071/CP22294

Rajpal, V. R., Singh, A., Kathpalia, R., Thakur, R. K., Khan, M. K., Pandey, A., et al. (2023). The prospects of gene introgression from crop wild relatives into cultivated lentil for climate change mitigation. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1127239

Renzi, J. P., Coyne, C. J., Berger, J., Von Wettberg, E., Nelson, M., Ureta, S., et al. (2022). How could the use of crop wild relatives in breeding increase the adaptation of crops to marginal environments? *Front. Plant Sci.* 13, 886162. doi: 10.3389/fpls.2022.886162

Salgotra, R. K., and Chauhan, B. S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* 14, 174. doi: 10.3390/genes14010174

Thingnam, S. S., Lourembam, D. S., Tongbram, P. S., Lokya, V., Tiwari, S., Khan, M. K., et al. (2023). A perspective review on understanding drought stress tolerance in wild banana genetic resources of northeast India. *Genes* 14, 370. doi: 10.3390/genes14020370

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

Zhu, P., Burney, J., Chang, J., Jin, Z., Mueller, N.D., Xin, Q., et al. (2022). Warming reduces global agricultural production by decreasing cropping frequency and yields. *Nat. Clim. Change* 12, 1016–1023. doi: 10.1038/s41558-022-01492-5

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 June 2022
doi: 10.3389/fpls.2022.881964

Assessment of the Characteristics of Waxy Rice Mutants Generated by CRISPR/Cas9

Yuhao Fu[1,2,3†], Tingting Luo[1†], Yonghuan Hua[1], Xuehai Yan[4], Xu Liu[1], Ying Liu[1], Yiping Liu[1], Baoli Zhang[1], Rui Liu[1], Zizhong Zhu[1] and Jun Zhu[1,2]*

[1] Rice Research Institute of Sichuan Agricultural University, Chengdu, China, [2] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, China, [3] Key Laboratory of Southwestern Chinese Medicine Resources, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [4] Leshan Municipal Bureau of Agriculture and Rural Affairs, Leshan, China

OPEN ACCESS

Edited by:
Tofazzal Islam, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

Reviewed by:
Mohd. Kamran Khan, Selcuk University, Turkey
Sulaiman Ahmed, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences (CAS), China

***Correspondence:**
Jun Zhu
Zhujun987@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *24 February 2022*
Accepted: *13 May 2022*
Published: *10 June 2022*

Citation:
Fu Y, Luo T, Hua Y, Yan X, Liu X, Liu Y, Liu Y, Zhang B, Liu R, Zhu Z and Zhu J (2022) Assessment of the Characteristics of Waxy Rice Mutants Generated by CRISPR/Cas9. Front. Plant Sci. 13:881964. doi: 10.3389/fpls.2022.881964

The cooking and eating quality of rice grains is a major focus from a consumer's perspective and is mainly determined by the apparent amylose content (AAC) of the starch. Waxy rice, a type of rice with an AAC of less than 2%, is an important goal for the breeding of high-quality rice. In recent years, the cloning of the *Waxy* (*Wx*) gene has revealed the molecular mechanism of the formation of waxy traits in rice. However, there have been limited studies on the physicochemical properties, such as gelatinization temperature, rapid viscosity analyzer profile, and amylopectin fine structure of *wx* mutants. In the current study, a rapid and highly efficient strategy was developed through the CRISPR/Cas9 gene-editing system for generating *wx* mutants in the background of five different rice varieties. The *wx* mutation significantly reduced the AAC and starch viscosity but did not affect the major agronomic traits (such as plant height, panicle number per plant, grain number per panicle, and seed-setting frequency). Incorporation of the *wx* mutation into varieties with low initial AAC levels resulted in further reduction in AAC, but without significantly affecting the original, desirable gelatinization traits and amylopectin structure types, suggesting that parents with low initial AAC should be preferred in breeding programs.

Keywords: apparent amylose content, waxy rice, amylopectin structure, Waxy gene, CRISPR/Cas9

INTRODUCTION

Starch consists of two classes of α-polyglucans, amylose and amylopectin. Amylose is primarily a linear polymer, consisting of 1,4-D-glucopyranosyl units, while amylopectin is a highly branched polymer (Hizukuri et al., 1981; Hizukuri, 1986), with the amylose: amylopectin ratio in conventional rice being approximately 20: 80. The amylose content (AC) of the starch is arguably the most important quality indicator in rice grains, particularly with respect to cooking, processing, and eating qualities (Juliano, 1998). Rice cultivars can be divided into five categories according to AC: waxy (0–2%), very low (5–12%), low (13–20%), intermediate (21–25%), and high amylose (26–33%) (Juliano, 1992). Waxy rice (or glutinous rice) is regarded as high-quality rice, which is usually sticky when cooked (Juliano, 1998). Because of its unique, functional characteristics, waxy

rice is widely used in processed food, medicines, and cosmetics (Bao et al., 2004; Puchongkavarin et al., 2005; Chun et al., 2010).

Amylose is proposed to have a significant influence on the physicochemical properties of waxy rice flour (Jane et al., 1999). Waxy rice starch with a certain amount of amylose as the donor substrate results in increases in the levels of crystallinity, solubility and paste clarity, gelatinization temperature (GT), enthalpy (ΔH), gel strength, and storage modulus (Lu et al., 2008; Guo et al., 2019). There is a reduction in the retrogradation of flour, a process in which disaggregated amylose and amylopectin chains in a gelatinized starch paste reassociate to form more ordered structures, when waxy rice flour is used, with an increase in AC (Sasaki et al., 2000). On the other hand, the fine structure of amylopectin is considered to affect the gelatinization, retrogradation, and rheology of waxy rice starch (Kalichevsky et al., 1990; Chung et al., 2008). During retrogradation of waxy rice starch, GT is positively correlated with the branch chain length of amylopectin (Jane et al., 1999; Singh et al., 2012), with the greater branch density of amylopectin being associated with increasing ΔH and solubility but with decreasing viscosity of waxy rice starch (Sorndech et al., 2015; Ren et al., 2017). These studies revealed the vital role played by amylopectin fine structure and AC in the physicochemical properties of waxy rice. However, there are no effective and rapid means for breeding waxy rice cultivars by targeting low AC or the fine structure of amylopectin, and this limitation has constrained the larger-scale production and consumption of waxy rice cultivars.

The *Wx* gene is located on rice chromosome 6 and encodes granule-bound starch synthase I (GBSSI), which mainly controls amylose synthesis in the seed endosperm and thus directly affects the quality of rice grains (Sano, 1984; Wang et al., 1995). The use of numerous allelic variants of *Wx* (e.g., Wx^{lv}, Wx^{a}, Wx^{in}, Wx^{b}, Wx^{op}, Wx^{mp}, and *wx*) in the rice breeding programs has led to regional differences in the AC of rice (Sano, 1984; Larkin and Park, 2003; Chen et al., 2008; Mikami et al., 2008; Dobo et al., 2010; Teng et al., 2011; Zhang et al., 2019) and allows rice breeders to control the expression of the *Wx* gene to obtain rice varieties with desired starch quality traits (Satoh and Omura, 1981; Liu et al., 2003; Liu et al., 2005; Kong et al., 2015). In recent years, the CRISPR/Cas9 gene-editing system has been used to knock out or fine-tune the expression of the *Wx* gene (Ma et al., 2015). The first exon of the *Wx* gene was edited to produce a null mutation *via* CRISPR/Cas9, to generate four *japonica* rice *wx* mutant lines with significantly reduced AC, but with an unaltered expression of agronomic traits (Zhang et al., 2018; Yunyan et al., 2019). Editing the TATA box of the Wx^{b} promoter downregulated *Wx* expression and thus fine-tuned grain AC, which could improve the grain quality (Huang et al., 2020). However, the relationship between the physicochemical properties of rice starch in each *wx* mutant and its corresponding wild type (WT) remains unclear, and more information is needed before breeding for high-quality waxy starch by gene editing can be carried out effectively.

In the current study, we systematically analyzed the physicochemical properties of *wx* mutants generated in different genetic backgrounds by the CRISPR/Cas9 system. The results showed that the physicochemical properties of *wx* mutants are highly correlated with the properties of the WT parent, suggesting that parents with low initial AAC should be preferred for generating waxy rice mutants with desirable starch characteristics. Our work provides new insights into the breeding of waxy rice with high-quality starch and could contribute significantly to expanding the waxy rice germplasm resources.

RESULTS

Mutant Isolation and Agronomic Trait Assessment

We first selected two cultivated rice varieties with different major *Wx* alleles, namely, QLD (Wx^{a}) and YSZ (Wx^{b}), which are both widely grown in the rice-producing region of Southwest China. We designed a CRISPR/Cas9 (**Figure 1A**) construct that accurately targeted the second exon (87–109 bp) of the *Wx* gene with the expectation of generating a null mutation (**Figure 1B**). In the genetic background of QLD (**Figure 1C**), we obtained four homozygous mutant alleles for the target site, with each of these mutations shifting the reading frame of the candidate gene in the 5' coding region, indicating that they were most likely to be null alleles. Five homozygous mutant alleles of Wx^{b} were also generated in the background of YSZ (**Figure 1D**). We observed that 80% of T_0 transformants were mutants in QLD plants and 82.35% in YSZ plants, indicating that the CRISPR/Cas9 system exhibited a high mutagenesis efficiency (**Supplementary Table 1**). Meanwhile, our results showed that the *wx* mutations did not change the main agronomic traits of the T_1 generation lines, including plant height, panicle number per plant, grain number per plant, and seed-setting frequency (**Supplementary Figure 1**).

Assessment of the Grain Appearance and Cooking Quality of the *wx* Mutants

In addition, we identified two single-base homozygous mutants, YSZ*wx1* and QLD*wx4* (T_5 generation), and performed quality analysis on them (**Supplementary Table 2**). The results showed no significant difference in the major grain quality traits of the two *wx* mutants, except that the color of endosperm changed from transparent to milky white (**Figure 1E**) and the kernel weight decreased (**Supplementary Figure 2**). The cooking and eating quality of YSZ*wx1* rice was soft but not sticky, while that of the QLD*wx4* rice was very sticky (**Figure 1F**), and the grains of both mutants exhibited a significantly higher gel consistency (GC) than that of flour of the corresponding WT parents (**Figure 1G**).

Analyses of Apparent Amylose Content, Total Protein, Total Starch, and Grain Yield in *wx* Mutants

We observed that the AAC values of YSZ and QLD were significantly decreased to 1.16 and 2.36%, respectively (**Figure 2A**) (reduced by 91.01 and 89.80%, respectively) in the corresponding mutants, leading to increased GC (**Figure 2B**). The total protein content of YSZ and QLD was increased by 26.01 and 6.37%, respectively, in the mutants (**Figure 2C**), but the total starch content showed no consistent effect in

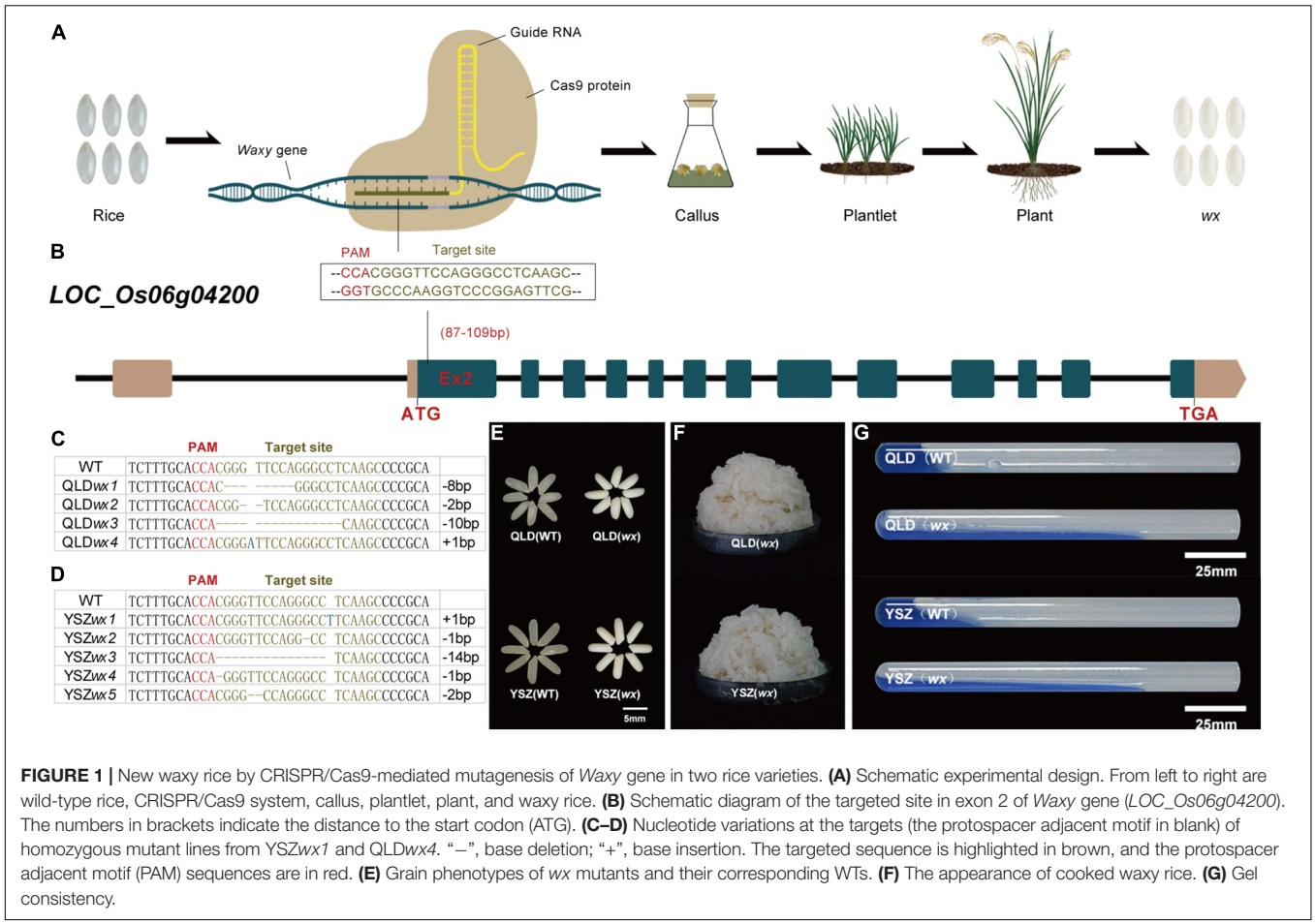


FIGURE 1 | New waxy rice by CRISPR/Cas9-mediated mutagenesis of *Waxy* gene in two rice varieties. **(A)** Schematic experimental design. From left to right are wild-type rice, CRISPR/Cas9 system, callus, plantlet, plant, and waxy rice. **(B)** Schematic diagram of the targeted site in exon 2 of *Waxy* gene (*LOC_Os06g04200*). The numbers in brackets indicate the distance to the start codon (ATG). **(C–D)** Nucleotide variations at the targets (the protospacer adjacent motif in blank) of homozygous mutant lines from YSZ*wx1* and QLD*wx4*. "−", base deletion; "+", base insertion. The targeted sequence is highlighted in brown, and the protospacer adjacent motif (PAM) sequences are in red. **(E)** Grain phenotypes of *wx* mutants and their corresponding WTs. **(F)** The appearance of cooked waxy rice. **(G)** Gel consistency.

the mutants, increasing by 8.11% in the YSZ mutant and decreasing by 6.19% in the QLD mutant, compared with the corresponding WT (**Figure 2D**). Moreover, the rice yield values indicated that the grain yields of YSZ*wx1* and QLD*wx4* were 8,195 and 9,050 kg/hm^2 (= kg/ha), respectively, which were significantly higher than the yields of the corresponding wild types (**Figure 2E**). These results suggest that editing the *Wx* gene can significantly improve both the quantity and the quality of the rice produced by the mutants.

Analyses of Starch Thermal Properties of the Mutants

The results of analyses of the thermal properties of the starch from the *wx* mutants suggested that editing of the *Wx* gene did not significantly change the GT, although the two *wx* mutants showed a slightly higher GT range than the WT lines (**Figure 2F**). For example, the onset temperature (*To*), peak temperature *(Tp)*, conclusion temperature (*Tc*), and ΔH of YSZ and YSZ *wx1* were 66.28 and 66.65°C, 70.70 and 71.66°C, 76.27 and 78.74°C, and ΔH of 11.6 and 12.55 J/g, respectively (**Supplementary Table 3**). Similar results were also obtained for QLD and QLD*wx4* (**Supplementary Table 3**). The results suggest that *To*, *Tp*, *Tc*, and ΔH from the *wx* mutants were all higher, compared with the corresponding WT, which led to the mutant starch being gelatinized later than the WT starch. However, the mutant starch still belonged to the same GT type as the WT starch.

Analyses of Starch Pasting Properties of the Mutants

The pasting properties of all *wx* mutants were significantly different from those of the corresponding WT lines ($P < 0.05$) (**Figure 3A**). For example, the mean peak viscosity (PKV), hot paste viscosity (HPV), cool paste viscosity (CPV), breakdown viscosity (BDV = PKV-HPV), setback viscosity (SBV = CPV-HPV), peak time, and pasting temperature (PT) of QLD and QLD*wx4* were 2,662 and 2,841 cp, 1,760 and 1,432 cp, 4,210 and 1,497 cp, 939 and 1,488 cp, 1,548 and -1,344 cp, 6.38 and 4.57 s, and 80.7 and 81.5°C, respectively (**Supplementary Table 4**). The results show that *wx* mutants have softer pasting properties than their corresponding WT lines (**Figure 3B**).

Chain-Length Distribution of Amylopectin

Subsequently, we compared the chain-length distribution of amylopectin between the two *wx* mutants and their corresponding WT parents. The chain-length distributions of amylopectin in *wx* mutants were markedly different from those of the corresponding WT lines (**Figure 4A**). *Wx* mutants

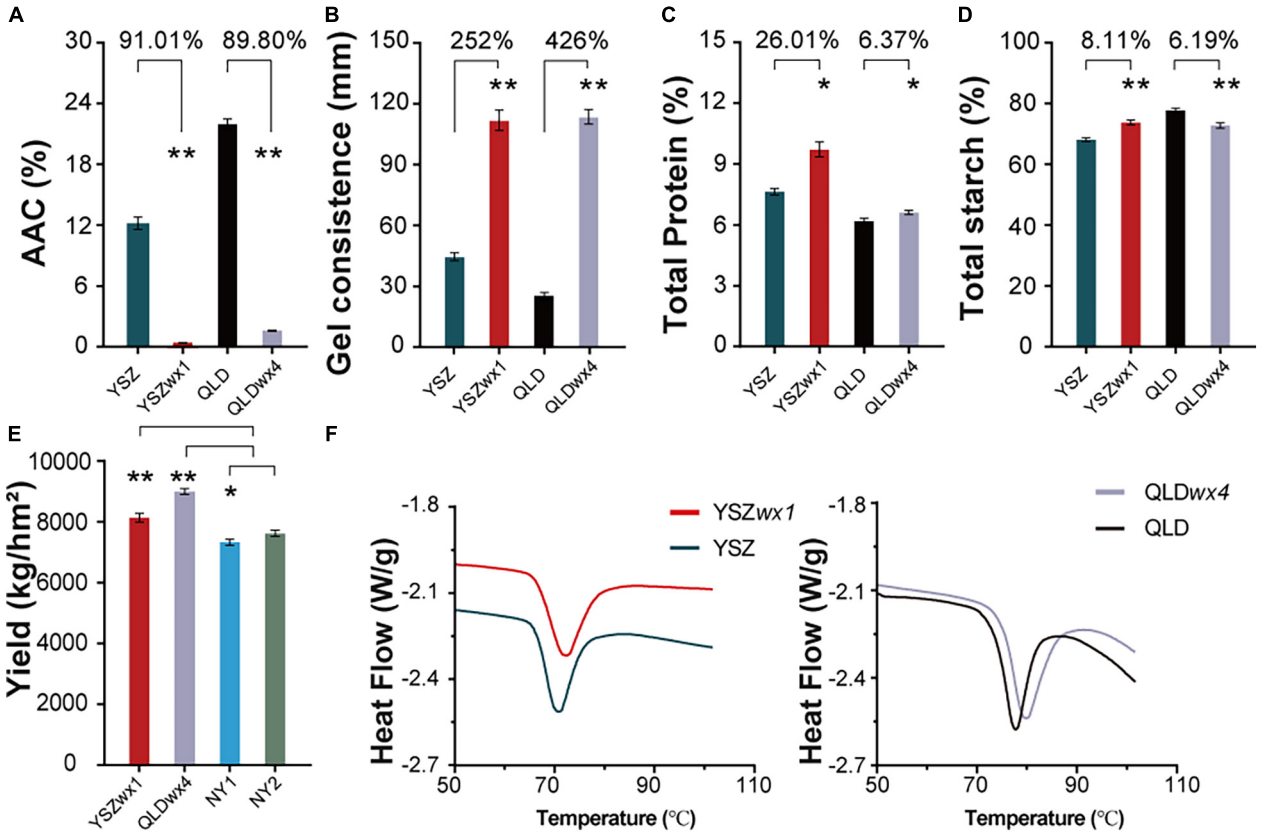


FIGURE 2 | Apparent amylose content (AAC), gel consistency (GC), total protein, total starch, grain yield, and gelatinization properties of *wx* mutants and their corresponding WT lines. **(A)** AAC of *wx* mutants and their corresponding WTs. **(B)** GC of *wx* mutants and their corresponding WTs. **(C)** Total protein of *wx* mutants and their corresponding WTs. **(D)** Total starch of *wx* mutants and their corresponding WTs. **(E)** The grain yield of *wx* mutants of NY1 and NY2. NY1 is a *japonica* waxy rice (in blue), NY2 is an *indica* waxy rice (in green). **(F)** Gelatinization properties curve in *wx* mutants and their corresponding WTs. Error bars are mean ± SD ($n = 3$). Significant differences were determined by Student's *t*-test (*$P < 0.05$, **$P < 0.01$).

exhibited a higher percentage of short-chain amylopectin (degree of polymerization, DP 7–20) and a lower percentage of medium- to long-chain amylopectin (DP > 20) than the WT lines (**Figure 4B**). Based on the amylopectin chain ratio (ACR) [ACR = (ΣDP ≤ 10)/(ΣDP ≤ 24)], the amylopectin structure of the starch from cultivated rice can be classified into one of three types: L-type (ACR ≤ 0.200), M-type (ACR 0.201–0.239), and S-type (ACR ≥ 0.240) (Nakamura et al., 2002). Interestingly, even though the ACR values of the *wx* mutants were significantly lower than those of the corresponding WT lines, they still belonged to the same amylopectin structural type. The amylopectin structure of QLD*wx4* and QLD belonged to the L-type, whereas YSZ*wx1* and YSZ belonged to the M-type (**Figure 4C**).

Correlation Analysis of Starch Samples

We used the same method for generating *wx* mutants with three other rice cultivars to find a potential link between gene-edited *wx* mutants and wild types (WTs) (**Supplementary Figure 3** and **Supplementary Table 5**). These three rice cultivars all belonged to the Wx^b genotype (**Supplementary Table 6**). The results showed that the cultivar (**Figure 5A**) with a low AAC would generate a *wx* mutant with a lower AAC (**Figure 5B**). Furthermore, the AAC of the wx^a mutants was higher than that of the three wx^b mutants, and the AAC values of QLD*wx4* and SH789*wx* were greater than 2% (**Supplementary Tables 2**, **6**). The *Tc* and Δ*H* values of the *wx* mutants were greater than those of the corresponding WTs, although their GT types were unchanged (**Table 1**). Similarly, the ACR values were altered by mutation but stayed within the same amylopectin structure type as the WT (**Figure 5B**). Further data analysis showed a positive linear correlation between ACR and GT (**Figure 5C**). In addition, two new starch types from the mutants were proposed, based on the amylopectin structure and the GT, namely, the HGT-type (high gelatinization temperature type, ACR < 0.18) and LGT-type (low gelatinization temperature type, ACR > 0.18). We demonstrated that the GT type and ACR type were not affected by the editing of the *Wx* gene, and that, to achieve *wx* mutants with a lower AAC, the WT parent exhibiting a lower AAC should be selected.

In this study, we also separately analyzed the relationship between the physicochemical properties of the starch obtained from the WT and the *wx* mutants. Pearson's linear correlation analysis indicated that the correlations between rice grain quality, amylopectin structure, and starch physicochemical properties differed between WTs and the *wx* mutants (**Figure 6**). The WT

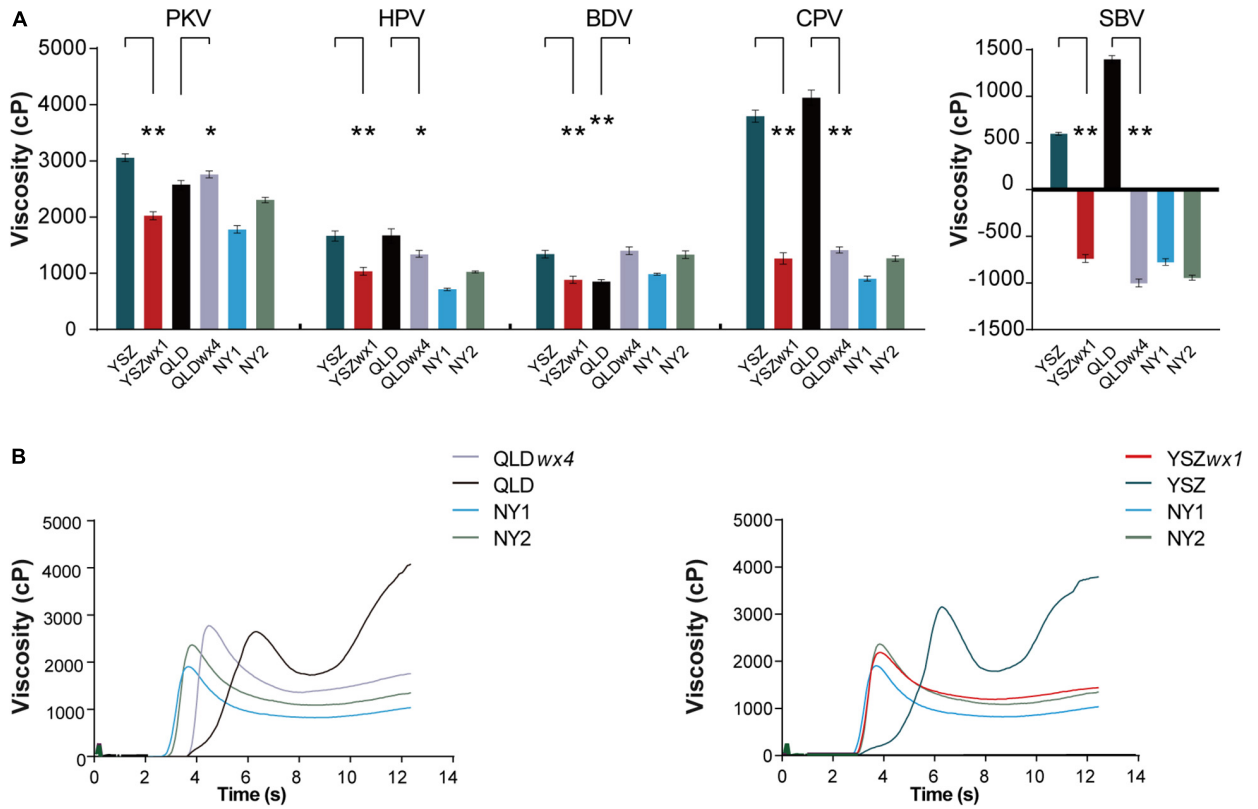


FIGURE 3 | Rapid viscosity analyzer profiles of rice starch in *wx* mutants and their corresponding WTs. **(A)** Rapid viscosity analyzer (RVA) profiles of starch in *wx* mutants and their corresponding WTs. **(B)** Rapid viscosity analyzer profile curve of *wx* mutants and their corresponding WT flour samples. The date of peak viscosity (PKV), hot paste viscosity (HPV), cool paste viscosity (CPV), breakdown viscosity (BDV = PKV-HPV), and setback viscosity (SBV = CPV-HPV) were derived from the RVA profiles. NY1 is a *japonica* waxy rice variety (in blue), and NY2 is an *indica* waxy rice variety (in green). Error bars are mean ± SD ($n = 3$). Significant differences were determined by Student's *t*-test (*$P < 0.05$, **$P < 0.01$).

data revealed that GT was negatively correlated with both DP 6–12 and ACR but was positively correlated with DP 13–24. The AAC was positively correlated with peak time, CPV, HPV, and SBV, but was negatively correlated with BDV and GC. On the other hand, the data from *wx* mutants indicated that GT was negatively correlated with DP 6–12, DP 6–24, and ACR, but was positively correlated with DP 13–24, peak time, CL, and DP 25–100. The AAC of *wx* mutants showed a negative correlation with GC and a positive correlation with peak time, HPV, and CPV. The results indicated that amylopectin structure, AAC, GT, and pasting properties in different rice cultivars reflected different physicochemical properties.

DISCUSSION

Over the past few decades, many waxy rice cultivars have been developed by conventional plant breeding techniques and mutagenesis. However, these methods have the disadvantages of unpredictable starch quality, as well as being time-consuming and laborious (Toda, 1980; Deng, 1992; Olsen and Purugganan, 2002). Recently, the development of genome-editing technology, particularly the discovery and application of the CRISPR/Cas9 gene-editing system, has led to its widespread use in studying plant genome functions (Ran et al., 2013; Belhaj et al., 2015; Bortesi and Fischer, 2015). Several studies have reported that the *Wx* gene can be successfully edited with high efficiency in rice through the use of the CRISPR/Cas9 system, which shows a high average mutation rate of up to 80%, without affecting the major agronomic traits in the transgenic lines (Zhang et al., 2018; Yunyan et al., 2019; Huang et al., 2020). In agreement with previous studies, the *wx* mutation rate in the present study showed a high value of 80% and above, and *wx* mutants exhibited major agronomic traits similar to those of their corresponding WT lines. However, in this study, we identified the *Wx* allele and the starch quality of the WT lines before editing and compared the major starch quality traits of the *wx* mutants from parents with different *Wx* alleles, an approach that is different from that undertaken in previous studies.

Our results showed that the AAC values of three wx^b mutants were lower than that of wx^a mutants (**Supplementary Tables 2**, **6**), and WT parents with a lower level of AAC will generate *wx* mutants with lower AAC values. The AAC of wx^a mutants was higher than the 2% upper threshold, a characteristic that is widely believed to be unacceptable for the AAC of elite waxy rice grains (Juliano, 1992). Therefore, our

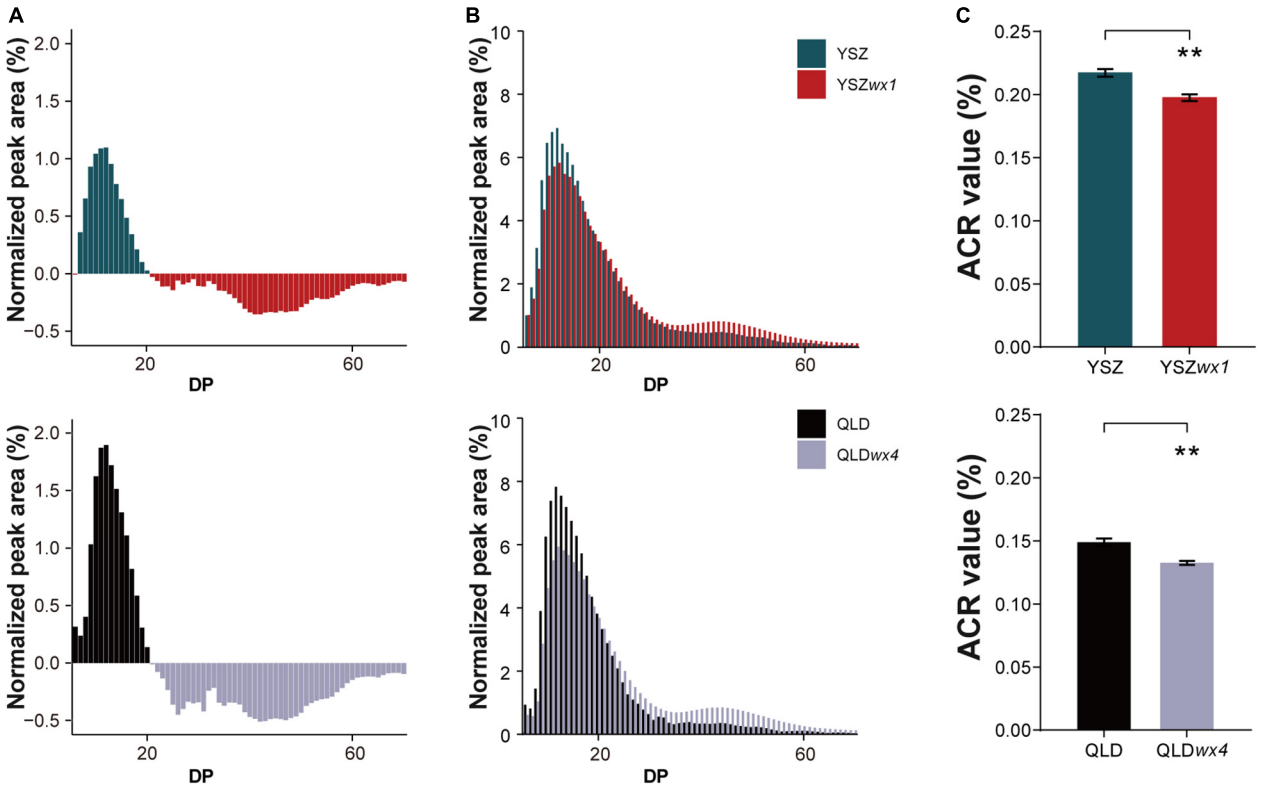


FIGURE 4 | The fine structure of amylopectin in *wx* mutants and their corresponding WTs. **(A)** Comparison of percentage values of high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) chromatograms of amylopectin chain length between *wx* mutants and their corresponding WTs. The plus values represent the WTs and the minus values represent the *wx* mutants. **(B)** The difference in the chain-length distribution of amylopectin between *wx* mutants and their corresponding WTs. **(C)** The $\Sigma DP \leq 10/\Sigma DP \leq 24$ values of *wx* mutants and their corresponding WTs. ACR: amylopectin chain ratio. Error bars are mean ± SD ($n = 3$). Significant differences were determined by Student's *t*-test: $^{**}P < 0.01$.

results indicate that it is necessary to select parental rice varieties with a lower level of AAC to obtain high-quality waxy rice grains in the breeding program (conventional or mutagenesis breeding). A recent study reported that silencing the expression of the *Wx* gene regulates genes related to amylose synthesis, with upregulation of granule-bound starch synthase II (GBSSII) compensating for some loss of amylose in the rice endosperm (Pérez et al., 2019). This may explain why the same *Wx* parental genotype results in different AAC levels after gene editing (**Figure 5A**), and further studies are warranted to study the contribution of different GBSSII types to the amylose content in endosperm in order to develop new waxy rice varieties with more precise and desirable AAC levels.

Amylopectin is also proposed to have a significant influence on the physicochemical properties of waxy rice flour (Man et al., 2013; Huang and Lai, 2014). Previous studies had shown that some starch-synthesizing enzymes influence the structure of amylopectin, such as soluble starch synthases (SSs), branching enzymes (BEs), debranching enzymes (DBEs), and isoamylase 1 (ISAI) (Umemoto et al., 1995; Man et al., 2013; Shufen et al., 2019; Zhu et al., 2020). These observations may rationally explain the differences in chain-length distributions of amylopectin in different *wx* mutants obtained from the same *Wx* allele (**Figures 4, 5** and **Supplementary Figure 4**). The expression of the major starch synthesis-related genes is upregulated following the silencing of the expression of the *Wx* gene (Pérez et al., 2019), leading to the differences in chain-length distributions of amylopectin observed in *wx* mutants and their corresponding WTs (**Figure 4** and **Supplementary Figure 4**). Although the chain-length distributions of amylopectin were changed, the range of ACR values was not significantly different (**Figure 5**). Based on the ACR values, three types of amylopectin could be distinguished, namely L-type (ACR < 0.2), M-type (0.2–0.24), and S-type (ACR > 0.24). The GT of rice can also be divided into three types: low-type (<70°C), intermediate-type (70–74°C), or high-type (>74°C), and is negatively correlated with ACR (Kongseree and Juliano, 1972; Nakamura et al., 2002; Vandeputte et al., 2003). In the current study, we found that *wx* mutants exhibited ACR values and amylopectin types similar to those of their corresponding WTs, and that the ACR value was linearly correlated with GT, which means that the breeder can obtain a waxy rice genotype with a GT and an amylopectin structural type similar to those of the corresponding WT. We also distinguished two amylopectin structure types based on ACR values: HGT-type (ACR < 0.18) and LGT-type (ACR > 0.18). The GT of the HGT-type is more than 77°C, in contrast to that of the LGT-type,

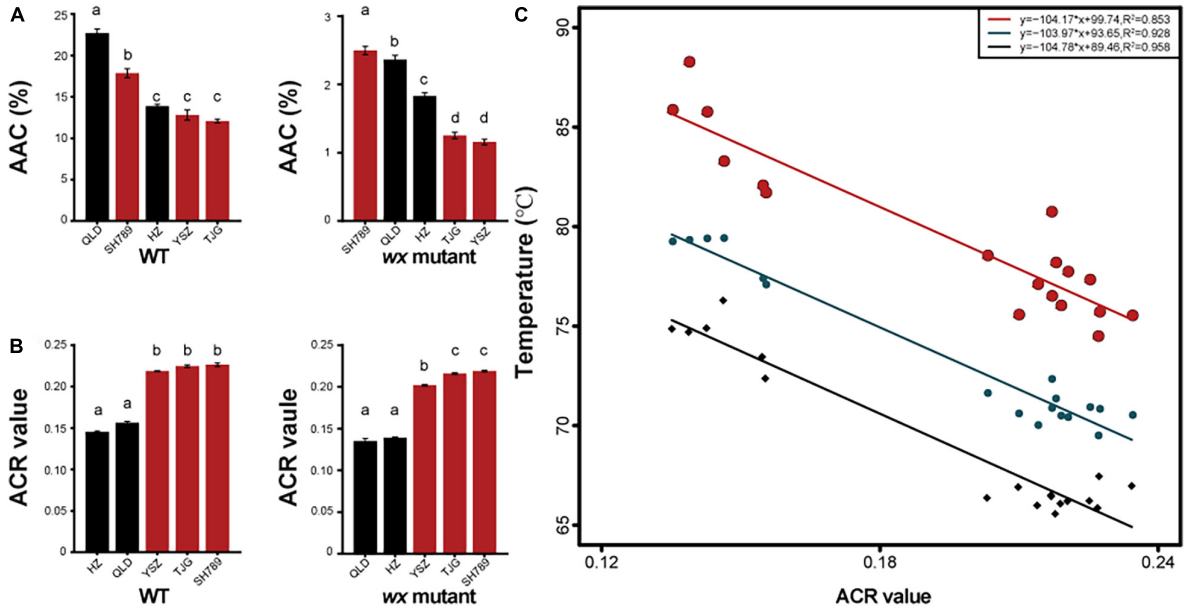


FIGURE 5 | The apparent amylose content (AAC) and amylopectin chain ratio (ACR) value of five wx mutants and their corresponding WTs **(A)** AAC of five *wx* mutants and their corresponding WTs. **(B)** ACR value of five *wx* mutants and their corresponding WTs. AAC and ACR value of WTs (left-hand bar graph) and *wx* mutants (middle bar graph). **(C)** Relationships between ACR value and gelatinization temperature; each point represents the ACR value of a variety of rice or waxy rice. Black points represent the onset temperature (*To*), bluish-black points represent the peak temperature (*Tp*), and red points represent the conclusion temperature (*Tc*). Error bars are mean ± SD ($n = 3$). Values followed by different superscripts in the same column are significantly different ($P < 0.05$).

which is less than 72°C (**Table 1**). Collectively, it can be easier to discriminate between amylopectin structure types selected on the basis of GT, which could benefit breeding for rice with specific waxy starch characteristics.

Previous studies had tended to indicate that, in all rice cultivars, the GT was negatively correlated with short amylopectin chains (DP 6–12) and ACR values, but positively correlated with the medium ($12 \leq$ DP ≤ 24) and long amylopectin chains (DP > 37) (Wang and Wang, 2002; Vandeputte et al., 2003; Wong et al., 2003). However, our study indicated that GT was only significantly negatively correlated with ACR and DP 6–12 when the genotypes were WTs, but was significantly positively correlated with DP 25–100 when all were *wx* mutants or waxy rice genotypes (**Figure 6**). Therefore, carrying out a statistical analysis of starch quality data combined from both non-waxy and waxy genotypes may not be the best way to identify a relationship between the physicochemical properties. On the basis of the above-mentioned results, we suggest that analyzing the relationship between the physicochemical properties of the same variety (waxy or non-waxy) can provide greater accuracy, and that selection on the basis of ACR value can more effectively identify genotypes with specific GT values.

CONCLUSION

In the current study, five new waxy rice mutants were generated, using the CRISPR/Cas9 system to edit the *Wx* gene. Detailed analysis of the two *wx* mutants from QLD and YSZ showed a significant decrease in AAC and viscosity compared with the corresponding WT, and a significant increase in GC and total protein content, with the grain yields of both mutants being higher than those achieved by the two parental elite cultivated rice WTs. Based on the ACR value, two types of amylopectin structure are proposed: HGT-type (ACR < 0.18; GT > 77°C) and LGT-type (ACR > 0.18; GT < 72°C). Our study also found that WTs with a low level of AAC generated *wx* mutants with lower AAC, with the *wx* mutants also having values of ACR, GT, and amylopectin structure types similar to those of their corresponding WTs, suggesting that parents with targets of specific GT values and low initial AAC should be selected for in breeding programs. In conclusion, we propose a novel strategy for generating waxy rice genotypes directly, and provide new insights into expanding waxy rice germplasm resources for use in breeding programs.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

We selected five elite rice varieties with different genetic backgrounds, including four *indica* varieties (YSZ, QLD, SH789, and HZ) and one *japonica* variety (TJG), which were mainly planted in eastern Asia, central China, and western China. The seed was supplied by the Rice Research Institute of Sichuan Agricultural University. The *japonica* waxy line NY1 and the *indica* waxy line NY2 were used as controls. Unless indicated, all

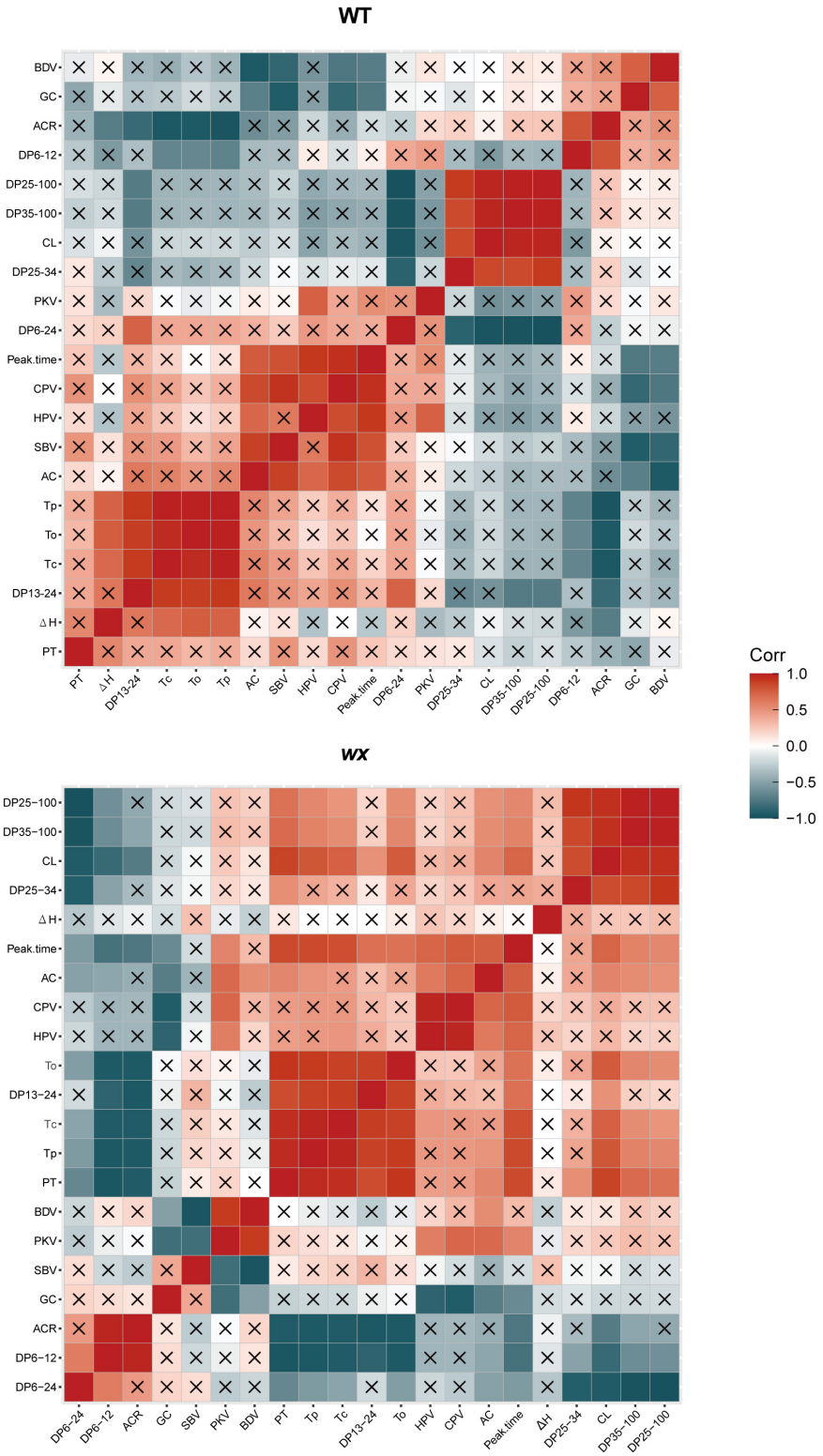


FIGURE 6 | Pearson's correlation analysis of the relationship between amylopectin structure of *wx* mutants and their corresponding WTs. The stronger the red color, the more significant the positive correlation; the stronger the bluish-black color, the more significant the negative correlation. "×" means no correlation. Δ*H*, gelatinization enthalpy; *To*, onset temperature; *Tp*, peak temperature; *Tc*, conclusion temperature; AC, amylose content; PKV, peak viscosity; HPV, hot paste viscosity; CPV, cool paste viscosity; BDV, breakdown viscosity; SBV, setback viscosity; PT, pasting temperature; ACR value, amylopectin chain ratio; CL, chain length; GC, gel consistency.

TABLE 1 | Differential scanning calorimetry (DSC) and enthalpy (Δ*H*) values of *wx* mutants and their corresponding WTs.

Cultivar	*To* (Δ)	*Tp* (°C)	*Tc* (°C)	Δ *H* (J/g)	*Tc* – *To* (°C)	GT type
YSZ	66.0 ± 0.1	70.3 ± 0.2	76.2 ± 0.2	11.9 ± 0.2	10.2 ± 0.2	*I*
YSZ*wx1*	66.4 ± 0.1**	71.8 ± 0.2**	78.6 ± 0.4**	13.3 ± 0.3**	12.2 ± 0.2**	
QLD	73.5 ± 0.5	77.3 ± 0.3	81.9 ± 0.3	12.8 ± 0.2	8.4 ± 0.3	*H*
QLD*wx4*	74.8 ± 0.5*	79.6 ± 0.4**	86.0 ± 0.3**	15.2 ± 0.4**	11.2 ± 0.8**	
SH789	65.9 ± 0.1	70.1 ± 0.7	74.6 ± 0.2	10.8 ± 0.2	8.8 ± 0.2	*I*
SH789*wx*	65.5 ± 0.2*	71.4 ± 0.2*	78.2 ± 0.2**	12.7 ± 0.3**	12.7 ± 0.1**	
TJG	66.3 ± 0.1	71.1 ± 0.2	77.6 ± 0.3	10.3 ± 0.2	11.3 ± 0.3	*H*
TJG*wx*	66.6 ± 0.1*	72.3 ± 0.2**	80.5 ± 0.2**	14.1 ± 0.2**	14.0 ± 0.3**	
HZ	76.3 ± 0.2	79.3 ± 0.3	83.6 ± 0.3	14.0 ± 0.2	7.2 ± 0.2	*H*
HZ*wx*	74.9 ± 0.2**	79.6 ± 0.4	88.5 ± 0.5**	15.4 ± 0.2**	13.6 ± 0.3	

*To, onset gelatinization temperature; Tp, peak gelatinization temperature; Tc, final gelatinization temperature; ΔH, gelatinization enthalpy (J/g); I, intermediate gelatinization temperature type; H, high gelatinization temperature type. Data are presented as means ± sd. *P < 0.05, **P < 0.01.*

rice lines were grown in paddy fields in Chengdu, China, during the normal rice-growing seasons.

CRISPR/Cas9 Gene-Editing and Gene Cloning

The CRISPR/Cas9-targeted genome editing tool was constructed and operated as previously described (Feng et al., 2013). The primer sequences used to construct the vector are shown in **Supplementary Table 7**.

The polymerase chain reaction (PCR) amplification conditions were as follows: initial denaturation at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30 s, extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min, with the PCR mix being based on the Foregene PCR Fast Mix. The PCR products were sequenced by the Sanger method. The primers used for *Wx* sequences, the *Wx* allele genotype, target sequences, and reporter gene detection are listed in **Supplementary Table 7**.

Tasting Properties of Waxy Rice

A sample (30 g) of milled waxy rice grains was cooked in an electric rice cooker to determine the tasting properties (model MDFBZ02ACM, Xiaomi, China). The tasting quality of milled waxy rice was determined with an RCTA-11A Taste Analyzer (Satake, Japan) (Zhang et al., 2016).

Measurements of Waxy Rice Starch

Grains of a single *wx* mutant with stable inheritance in the T_5 generation were dried at 37°C for 2 weeks. The dried grains were shelled, polished, and milled by a pearling rice mill, and finally screened through a 74-μ mesh. Starch was prepared by the method of Precha-Atsawanan et al. (2018). The waxy flour was steeped in 0.35% sodium hydroxide solution and incubated at 4°C for 48 h. The precipitants in the suspension were collected by centrifugation at 3,000 g for 15 min. The starch layer was re-mixed with water to which was added 1% alkaline protease (Sigma), and the pH was adjusted to 9 with sodium hydroxide and incubated at 37°C for 8 h. The washing step was repeated at least three times or until the pH value reached 7. The final precipitant was collected by centrifuging at 3,000 g for 30 min, and the starch was dried in a warm air oven at 40°C for 48 h. The starch was ground with a mortar and pestle, using a 0.08-mm sieve to collect the starch powder.

Quantification of Starch and Amylose

The rice flour was equilibrated in a constant temperature and humidity cabinet for 7 days. The starch content was determined according to the method of Smith and Zeeman (2006). D-glucose was stained with the GOPOD reagent (glucose oxidase plus peroxide and 4-aminoantipyrin) and determined by absorbance at 510 nm. Amylose content was measured as described in GB/T 15683-2008/ISO 6647-1: 2007. The amylase-iodine blue color was determined at 720 nm. Standard curves were plotted with standardized samples of rice amylose and amylopectin (China National Rice Research Institute, Zhejiang, China).

Quantification of Gel Consistency and Total Protein

The gel consistency (GC) of grains was evaluated as described previously (Cagampang et al., 1973). A sample (100 ± 1 mg) of the starch powder was placed into a 13 × 100 mm tube and wetted with 0.2 ml of 95% ethanol containing 0.025% thymol blue, and then 2.0 ml of 0.2 M KOH was added and allowed to boil for 3 min. GC was measured in ml as the length (mm) of the gel spreading in the tubes when laid flat on the graph for 1 h. The total protein content of the grain was assayed by the Kjeldahl method according to the Comin amylose assay procedure (Suzhou Comin Biotechnology Co., Ltd).[1]

Waxy Rice Powder Pasting Properties

To the white rice flour (3 g at 14% moisture content) obtained from the T_5 transformant generation was added 25 ml of ddH_2O_2. The setting procedure included the following steps: incubation for 1 min at 50°C, increasing the temperature at a rate of 12°C/min to 95°C (over 3.75 min), maintaining the temperature at 95·C (for 2.5 min), then decreasing it to 50°C (over a period of 3.75 min) at 12°C/min, and maintaining the temperature at 50°C (for 2 min). The initial speed was

[1]www.cominbio.com

set to 960 r/min for 10 s and then adjusted to 160 r/min. The pasting parameters of waxy rice flour, such as the peak viscosity (PKV), hot paste viscosity (HPV), cool paste viscosity (CPV), breakdown viscosity (BDV = PKV − HPV), setback viscosity (SBV = CPV − HPV), and pasting temperature (PT), were evaluated using a rapid viscosity analyzer (RVA, Newport Scientific, Australia) according to the methods of Chung et al. (2011).

Thermal Properties of Waxy Rice Starch

The thermal properties of waxy rice starch were measured using differential scanning calorimetry (DSC Q2000; TA Instruments Ltd., Crawley, United Kingdom) (Yang et al., 2018). The onset temperature (*To*), peak temperature (*Tp*), conclusion temperature (*Tc*), and gelatinization enthalpy (ΔH) were recorded using a Universal Analysis 2000 (TA Instruments Ltd., Crawley, United Kingdom). A sample (3 mg) of starch was mixed with 6 μl of distilled water, and the paste was added to an aluminum pan and incubated at 4°C for 24 h. The specimens were then heated from 30 to 100°C at a rate of increase of 10°C/min.

Amylopectin Chain-Length Distribution

The chain-length distribution of amylopectin was analyzed according to the method of Kowittaya and Lumdubwong (2014). Waxy rice starch was suspended in 95% (v/v) ethanol and boiled for 60 min. Hydrolyzed polyglucan was treated with *Pseudomonas amyloderamosa* isoamylase (1,000 U/μl, Sigma). The chain-length distribution of waxy rice starch was analyzed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). There were three mobile phases: A (aqueous solution), B (100 mM NaOH and 1 M NaAC), and C (100 mM NaOH). The flow rate was controlled at 0.4 ml/min, and the column temperature was controlled at 30°C.

Data Analysis

All experiments were independently repeated three times. The results were expressed as the mean ± standard deviation. The correlation was assessed using Pearson's correlation coefficient. Data analysis was performed with SPSS 25.0, and the significance was defined as $P < 0.05$ (* or lowercase letters) and $P < 0.01$ (** or uppercase letters). SnapGene 5.2 was used to visualize PCR and cloning.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YF, TL, and JZ designed the strategy. YH, XY, XL, YL, YPL, BZ, RL, and ZZ completed part of the experiments. YF and JZ organized the figures and article modification. YF, TL, and JZ analyzed the data and wrote the manuscript. All authors commented on the manuscript.

FUNDING

We acknowledge funding from the Sichuan Science and Technology Program (Grant Nos. 2021YFN0043, 2020ZHFP0017, and 2020YFN0006) and the Chengdu Science and Technology Program (Grant No. 2021-YF05-02115-SN).

ACKNOWLEDGMENTS

We thank the Sichuan Science and Technology Program and the Chengdu Science and Technology Program for supporting this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.881964/full#supplementary-material

REFERENCES

Bao, J., Corke, H., and Sun, M. (2004). Genetic diversity in the physicochemical properties of waxy rice (*Oryza sativa* L) starch. *J. Sci. Food Agric.* 84, 1299–1306.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* 32, 76–84.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006

Cagampang, G. B., Perez, C. M., and Juliano, B. O. (1973). A gel consistency test for eating quality of rice. *J. Sci. Food Agric.* 24, 1589–1594. doi: 10.1002/jsfa.2740241214

Chen, M. H., Bergman, C., Pinson, S. R. M., and Fjellstrom, R. G. (2008). Waxy gene haplotypes associations with pasting properties in an international rice germplasm collection. *J. Cereal Sci.* 48, 781–788.

Chun, A. R., Kim, D. J., Yoon, M. R., Oh, S. K., and Ju, S. C. (2010). Variation in quality and preference of sogokju (Korean traditional rice wine) from waxy rice varieties. *Korean J. Cropence* 55, 177–186.

Chung, H. J., Liu, Q. A., Lee, L., and Wei, D. Z. (2011). Relationship between the structure, physicochemical properties and in vitro digestibility of rice starches with different amylose contents. *Food Hydrocoll.* 25, 968–975. doi: 10.1111/jpn.13512

Chung, J. H., Han, J. A., Yoo, B., Seib, P. A., and Lim, S. T. (2008). Effects of molecular size and chain profile of waxy cereal amylopectins on paste rheology during retrogradation. *Carbohydr. Polym.* 71, 365–371.

Deng, X. (1992). Selection of two-line hybrid glutinous rice Deng Xiaolin. *Hybrid Rice* 7, 33–34.

Dobo, M., Ayres, N., Walker, G., and Park, W. D. (2010). Polymorphism in the GBSS gene affects amylose content in US and European rice germplasm. *J. Cereal Sci.* 52, 450–456.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D. L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* 23, 1229–1232. doi: 10.1038/cr.2013.114

Guo, L., Li, J., Gui, Y., Zhu, Y., and Cui, B. (2019). Improving waxy rice starch functionality through branching enzyme and glucoamylase: role of amylose as

a viable substrate. *Carbohydr. Polym.* 230:115712. doi: 10.1016/j.carbpol.2019.115712

Hizukuri, S. (1986). Polymodal distribution of the chain lengths of amylopectins, and its significance. *Carbohydr. Res.* 147, 342–347.

Hizukuri, S., Takeda, Y., Yasuda, M., and Suzuki, A. (1981). Multi-branched nature of amylose and the action of debranching enzymes. *Carbohydr. Res.* 94, 205–213.

Huang, L. C., Li, Q. F., Zhang, C. Q., Chu, R., Gu, Z. W., Tan, H. Y., et al. (2020). Creating novel Wx alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* 18, 2164–2166. doi: 10.1111/pbi.13391

Huang, Y. C., and Lai, H. M. (2014). Characteristics of the starch fine structure and pasting properties of waxy rice during storage. *Food Chem.* 152, 432–439. doi: 10.1016/j.foodchem.2013.11.144

Jane, J., Chen, Y. Y., Lee, L. F., McPherson, A. E., Wong, K. S., Radosavljevic, M., et al. (1999). Effects of amylopectin Branch Chain length and amylose content on the gelatinization and pasting properties of starch. *Cereal Chem. J.* 76, 629–637.

Juliano, B. O. (1992). Structure, chemistry, and function of the rice grain and its fractions. *Cereal foods world* 37, 772–779.

Juliano, B. O. (1998). Varietal impact on rice quality. *Cereal Foods World* 43, 207–211,214–216,218–222. doi: 10.1007/s11306-020-01670-6

Kalichevsky, M. T., Orford, P. D., and Ring, S. G. (1990). The retrogradation and gelation of amylopectins from various botanical sources. *Carbohydr. Res.* 198, 49–55.

Kong, X., Kasapis, S., and Bao, J. (2015). Viscoelastic properties of starches and flours from two novel rice mutants induced by gamma irradiation. *LWT Food Sci. Technol.* 60, 578–582.

Kongseree, N., and Juliano, B. O. (1972). Physicochemical properties of rice grain and starch from lines differing in amylose content and gelatinization temperature. *J. Agric. Food Chem.* 20, 714–718.

Kowittaya, C., and Lumdubwong, N. (2014). Molecular weight, chain profile of rice amylopectin and starch pasting properties. *Carbohydr. Polym.* 108, 216–223. doi: 10.1016/j.carbpol.2014.02.081

Larkin, P., and Park, W. (2003). Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). *Molecular Breeding* 12, 335–339.

Liu, Q., Wang, Z., Chen, X., Cai, X., Tang, S., Yu, H., et al. (2003). Stable inheritance of the antisense Waxy gene in transgenic rice with reduced amylose level and improved quality. *Transgenic Res.* 12, 71–82. doi: 10.1023/a:1022148824018

Liu, Q. Q., Yu, H. X., Chen, X. H., Cai, X. L., Tang, S. Z., Wang, Z. Y., et al. (2005). Field performance of transgenicindicahybrid rice with improved cooking and eating quality by down-regulation ofWxGene expression. *Mol. Breed.* 16, 199–208.

Lu, T.-J., Duh, C.-S., Lin, J.-H., and Chang, Y.-H. (2008). Effect of granular characteristics on the viscoelastic properties of composites of amylose and waxy starches. *Food Hydrocoll.* 22, 164–173.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007

Man, J., Yang, Y., Huang, J., Zhang, C., Chen, Y., Wang, Y., et al. (2013). Effect of simultaneous inhibition of starch branching enzymes I and IIb on the crystalline structure of rice starches with different amylose contents. *J. Agric. Food Chem.* 61, 9930–9937. doi: 10.1021/jf4030773

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H. Y., Suzuki, Y., et al. (2008). Allelic diversification at thewxlocus in landraces of Asian rice. *Theor. Appl. Genet.* 116, 979–989.

Nakamura, Y., Sakurai, A., Inaba, Y., Kimura, K., Iwasawa, N., and Nagamine, T. (2002). The fine structure of amylopectin in endosperm from Asian cultivated rice can be largely classified into two classes. *Starch Stärke* 54, 117–131.

Olsen, K. M., and Purugganan, M. D. (2002). Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162, 941–950. doi: 10.1093/genetics/162.2.941

Pérez, L., Soto, E., Farré, G., Juanos, J., Villorbina, G., Bassie, L., et al. (2019). CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis. *Plant Cell Rep.* 38, 417–433. doi: 10.1007/s00299-019-02388-z

Precha-Atsawanan, S., Puncha arnon, S., Wandee, Y., Uttapap, D., Puttanlek, C., and Rungsardthong, V. (2018). Physicochemical properties of partially debranched waxy rice starch. *Food Hydrocoll.* 79, 71–80.

Puchongkavarin, H., Varavinit, S., and Bergthaller, W. (2005). Comparative study of pilot scale rice starch production by an alkaline and an enzymatic process. *Starch* 57, 134–144.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.

Ren, J., Li, Y., Li, C., Gu, Z., Cheng, L., Hong, Y., et al. (2017). Pasting and thermal properties of waxy corn starch modified by 1,4-alpha-glucan branching enzyme. *Int. J. Biol. Macromol.* 97, 679–687. doi: 10.1016/j.ijbiomac.2017.01.087

Sano, Y. (1984). Differential regulation of waxy gene expression in rice endosperm. *Theor. Appl. Genet.* 68, 467–473. doi: 10.1007/BF00254822

Sasaki, T., Yasui, T., and Matsuki, J. (2000). Effect of amylose content on gelatinization, retrogradation, and pasting properties of starches from Waxy and Nonwaxy Wheat and their F1 seeds. *Cereal Chem.* 77, 58–63.

Satoh, H., and Omura, T. (1981). New endosperm mutations induced by chemical mutagens in rice *Oryza sativa* L. *Jap. J. Breed.* 31, 316–326.

Shufen, C., Yicong, C., Baobing, F., Guiai, J., Zhonghua, S., Ju, L., et al. (2019). Editing of rice isoamylase gene ISA1 provides insights into its function in starch formation. *Rice Sci.* 26, 77–87.

Singh, H., Lin, J. H., Huang, W. H., and Chang, Y. H. (2012). Influence of amylopectin structure on rheological and retrogradation properties of waxy rice starches. *J. Cereal Sci.* 56, 367–373. doi: 10.1016/j.carbpol.2015.12.025

Smith, A. M., and Zeeman, S. C. (2006). Quantification of starch in plant tissues. *Nat. Protoc.* 1, 1342–1345. doi: 10.1038/nprot.2006.232

Sorndech, W., Meier, S., and Jansson, A. (2015). Synergistic amylomaltase and branching enzyme catalysis to suppress cassava starch digestibility. *Carbohydr. Polym.* 132, 409–418. doi: 10.1016/j.carbpol.2015.05.084

Teng, B., Zeng, R-z, Wang, Y., Liu, Z., Zhang, Z., Zhu, H., et al. (2011). Detection of allelic variation at the Wx locus with single-segment substitution lines in rice (*Oryza sativa* L.). *Mol. Breed.* 30, 583–595.

Toda, M. (1980). Breeding of new rice varieties by gamma-rays. *Gamma Field Symp.* 18, 73–82.

Umemoto, T., Nakamura, Y., and Ishikura, N. (1995). Activity of starch synthase and the amylose content in rice endosperm. *Phytochemistry* 40, 1613–1616.

Vandeputte, G. E., Derycke, V., Geeroms, J., and Delcour, J. A. (2003). Rice starches. II. Structural aspects provide insight into swelling and pasting properties. *J. Cereal Sci.* 38, 53–59.

Wang, Y.-J., and Wang, L. (2002). Structures of four waxy rice starches in relation to thermal, pasting, and textural properties. *Cereal Chem.* 79, 252–256.

Wang, Z. Y., Zheng, F. Q., Shen, G. Z., Gao, J. P., Snustad, D. P., Li, M. G., et al. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. *Plant J.* 7, 613–622. doi: 10.1046/j.1365-313x.1995.7040613.x

Wong, K.-S., Kubo, A., Jane, J-l, Harada, K., Satoh, H., and Nakamura, Y. (2003). Structures and properties of amylopectin and phytoglycogen in the endosperm of sugary-1 mutants of rice. *J. Cereal Sci.* 37, 139–149.

Yang, T., Tan, X., Huang, S., Pan, X., Shi, Q., Zeng, Y., et al. (2018). Effects of experimental warming on physicochemical properties of indica rice starch in a double rice cropping system. *Food Chem.* 310:125981.

Yunyan, F., Jie, Y., Fangquan, W., Fangjun, F., Wenqi, L., Jun, W., et al. (2019). Production of two elite glutinous rice varieties by editing Wx gene. *Rice Sci.* 26, 118–124.

Zhang, C., Zhou, L., Zhu, Z., Lu, H., Zhou, X., Qian, Y., et al. (2016). Characterization of grain quality and starch fine structure of two japonica rice (*Oryza sativa*) cultivars with good sensory properties at different temperatures during the filling stage. *J. Agric. Food Chem.* 64, 4048–4057. doi: 10.1021/acs.jafc.6b00083

Zhang, C., Zhu, J., Chen, S., Fan, X., Li, Q., Lu, Y., et al. (2019). Wx(lv), the ancestral allele of rice Waxy gene. *Mol. Plant* 12, 1157–1166. doi: 10.1016/j.molp.2019.05.011

Zhang, J., Zhang, H., Botella, J. R., and Zhu, J.-K. (2018). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. *J. Integr. Plant Biol.* 60, 369–375. doi: 10.1111/jipb.12620

Zhu, J., Yu, W., Zhang, C., Zhu, Y., Xu, J., Li, E., et al. (2020). New insights into amylose and amylopectin biosynthesis in rice endosperm. *Carbohydr. Polym.* 230:115656. doi: 10.1016/j.carbpol.2019.115656

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 July 2022
DOI 10.3389/fpls.2022.947538

OPEN ACCESS

EDITED BY
Francesco Di Gioia,
The Pennsylvania State University (PSU), United States

REVIEWED BY
Anamika Pandey,
Selçuk University, Turkey
Mohd. Kamran Khan,
Selçuk University, Turkey

*CORRESPONDENCE
David M. Francis
Francis.77@osu.edu

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 18 May 2022
ACCEPTED 27 June 2022
PUBLISHED 28 July 2022

CITATION
Fenstemaker S, Cho J, McCoy JE, Mercer KL and Francis DM (2022) Selection strategies to introgress water deficit tolerance derived from *Solanum galapagense* accession LA1141 into cultivated tomato. *Front. Plant Sci.* 13:947538. doi: 10.3389/fpls.2022.947538

Selection strategies to introgress water deficit tolerance derived from *Solanum galapagense* accession LA1141 into cultivated tomato

Sean Fenstemaker[1], Jin Cho[1], Jack E. McCoy[2], Kristin L. Mercer[2] and David M. Francis[1]*

[1]Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH, United States, [2]Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH, United States

Crop wild relatives have been used as a source of genetic diversity for over one hundred years. The wild tomato relative *Solanum galapagense* accession LA1141 demonstrates the ability to tolerate deficit irrigation, making it a potential resource for crop improvement. Accessing traits from LA1141 through introgression may improve the response of cultivated tomatoes grown in water-limited environments. Canopy temperature is a proxy for physiological traits which are challenging to measure efficiently and may be related to water deficit tolerance. We optimized phenotypic evaluation based on variance partitioning and further show that objective phenotyping methods coupled with genomic prediction lead to gain under selection for water deficit tolerance. The objectives of this work were to improve phenotyping workflows for measuring canopy temperature, mapping quantitative trait loci (QTLs) from LA1141 that contribute to water deficit tolerance and comparing selection strategies. The phenotypic variance attributed to genetic causes for canopy temperature was higher when estimated from thermal images relative to estimates based on an infrared thermometer. Composite interval mapping using BC_2S_3 families, genotyped with single nucleotide polymorphisms, suggested that accession LA1141 contributed alleles that lower canopy temperature and increase plant turgor under water deficit. QTLs for lower canopy temperature were mapped to chromosomes 1 and 6 and explained between 6.6 and 9.5% of the total phenotypic variance. QTLs for higher leaf turgor were detected on chromosomes 5 and 7 and explained between 6.8 and 9.1% of the variance. We advanced tolerant BC_2S_3 families to the BC_2S_5 generation using selection indices based on phenotypic values and genomic estimated breeding values (GEBVs). Phenotypic, genomic, and combined selection strategies demonstrated gain under selection and improved performance compared to randomly advanced BC_2S_5 progenies. Leaf turgor, canopy temperature, stomatal conductance, and vapor pressure deficit (VPD) were evaluated and compared in BC_2S_5 progenies grown under deficit irrigation. Progenies co-selected for phenotypic values and GEBVs

wilted less, had significantly lower canopy temperature, higher stomatal conductance, and lower VPD than randomly advanced lines. The fruit size of water deficit tolerant selections was small compared to the recurrent parent. However, lines with acceptable yield, canopy width, and quality parameters were recovered. These results suggest that we can create selection indices to improve water deficit tolerance in a recurrent parent background, and additional crossing and evaluation are warranted.

KEYWORDS

thermal images, genomic selection (GS), proximal sensing, high-throughput phenotyping, inbred backcross method, canopy temperature (CT)

Introduction

Approximately 1.2 billion people worldwide reside in areas with water scarcity, and this number is growing (Food and Agriculture Organization [FAO], 2021 accessed at: https://www.fao.org/land-water/water/drought/en/). In arid regions, population growth and economic development are exhausting renewable but finite water resources (Food and Agriculture Organization [FAO], 2021). Water deficit tolerance is imparted through morphological and physiological traits in plants. Traits used to indicate plant response to water deficit include, but are not limited to, leaf rolling, flower and fruit set, water use efficiency, recovery after re-watering, stomatal conductance, plant survival, leaf water potential, leaf osmotic potential, osmoregulation, transpiration rate, photosynthetic rate, enzymatic activities, pollen viability, and seed germination (Foolad, 2007; Galmés et al., 2011). Mechanical reduction of water loss can be imparted by changes in plant morphology or mechanisms that promote stomatal closure. Morphological traits, such as leaf size, shape, thickness, orientation, reflective capabilities, trichomes, and leaf angle, can modulate physiological response to deficit irrigation (Inoue et al., 1990; Kitaya et al., 1998). Additionally, water deficit stress may influence leaf expansion, osmotic adjustment, biomass partitioning, and stomatal characteristics (Koch et al., 2019; Pardo and VanBuren, 2021). Other tolerance mechanisms may include adjustments of carbon concentration and a reduction of photorespiration, leading to increased water use efficiency (Pardo and VanBuren, 2021). Morphological and physiological responses are often challenging to measure in studies with many treatments or large population sizes because there is a limited window of time to capture a response that is physiologically meaningful and genetically relevant (Smith et al., 2021).

Optimizing strategies to select for water stress tolerance may be a key to successfully exploiting genes from exotic germplasm. Plant canopy temperature has been proposed as a proxy to indicate stress under water deficit (Perera et al., 2020) and may provide an efficient phenotyping alternative to physiological measurements (Grossiord et al., 2020). Therefore, image-based phenotyping may provide an opportunity to increase the precision and throughput of trait measurement. For example, estimates of canopy temperature obtained from images could substitute as a measure of water deficit stress and is potentially a viable trait for the efficient assessment of plant physiological responses. Quantifying the genetic variation associated with specific traits offers a framework to identify and optimize appropriate phenotyping strategies. For plant breeding applications, improving the genetic resolution of trait measurements can increase the relative efficiency of selection procedures on a per cycle, per cost, and per time basis (Heffner et al., 2010).

The foundation of crop improvement is based on genetic variation underlying traits of interest. Previous studies have shown that wild tomato species possess tolerances to predatory insects, excessive moisture, salinity, and water deficit (Rick, 1973). The introgression of genes from wild tomato species that impart tolerance to abiotic stress specifically has been reported from *Solanum pimpinellifolium*, *Solanum peruvianum*, *Solanum cheesmaniae*, *Solanum habrochaites*, *Solanum chmielewskii*, and *Solanum pennellii* (Rick, 1978; Saeed and Fatima, 2021). Introgressions from the wild relative *S. pennellii* have improved tomato response under water deficit (Dariva et al., 2020). Accessing traits from wild relatives by conventional breeding is challenging because of reproductive barriers and linkage drag (Swarup et al., 2021). A wild tomato species endemic to the Galápagos Islands, *Solanum galapagense* (Darwin et al., 2003), is a potential donor of abiotic stress tolerance (Brozynska et al., 2016). *S. galapagense* is adapted to harsh environments, including low moisture and saline coastal habitats (Rick, 1956; Rush and Epstein, 1976; Pailles et al., 2020). The relative merit of a wild accession as a source of tolerance can be determined by using the observations made by collectors in native habitats (Rick, 1973, 1978). *S. galapagense* accession LA1141 was found growing on the interior walls of a crater on the island of Santiago on the north-facing slope, which received average annual precipitation from

1991 to 2020 of 7.7 cm (World Bank Group, 2021, accessed at: climateknowledgeportal.worldbank.org). The habitat described in the collection notes suggested that LA1141 is adapted to an arid environment (Castro, 1968, accessed at: tgrc.ucdavis.edu/Data/Acc/dataframe.aspx?start=AccSearch.aspx&navstart=nav.html) and is plausibly a suitable candidate to improve cultivated tomatoes through introgression. Introgression of tolerance from wild accessions from the more closely related red-fruited tomato, such as *S. galapagense*, may have advantages, such as greater recombination and potentially less linkage drag, compared to crosses with a more distantly related, green-fruited wild tomato relative like *S. pennellii* (Hamlin et al., 2020).

The objectives of this study were to evaluate methods of measuring plant response to water deficit and compare selection strategies to identify germplasm with lower canopy temperature, higher turgidity, higher stomatal conductance, and lower vapor pressure deficit (VPD). Using an inbred backcross strategy (Kabelka et al., 2002; Robbins et al., 2009), we created tomato populations derived from the tolerant donor parent, *S. galapagense* accession LA1141 (Fenstemaker et al., 2022). This population was used to do the following: (1) Improve and compare phenotyping workflows for measuring canopy temperature based on high throughput image analysis to indicate water deficit tolerance. (2) Associate measures of water deficit stress with single nucleotide polymorphisms (SNPs) to describe the genetic basis of low canopy temperature and high turgidity under water deficit in LA1141. (3) Discover and introgress chromosomal regions contributing to canopy temperature and severity of wilt in regions of the genome derived from accession LA1141. (4) Evaluate phenotypic and genomic-based selection strategies to identify water deficit tolerant progenies and evaluate the recovery of horticultural traits from an elite commercial parent in tolerant selections. Results from this study can be directly used to enhance the introgression of water deficit tolerance from wild relatives of tomato.

Materials and methods

Plant materials, crosses, and growing conditions

An inbred backcross (IBC) population was created using *S. galapagense* S.C. Darwin and Peralta (formerly *Lycopersicon cheesmaniae* f. minor) (Hook. f) C.H.Mull.) accession LA1141, as a source of traits for tomato improvement, and *Solanum lycopersicum* L. (formerly *Lycopersicon esculentum* Mill) OH8245 (Berry et al., 1991), as the recurrent parent (Fenstemaker et al., 2022). The IBC population was derived from an initial hybridization of *S. galapagense* LA1141 as the female parent and OH8245 as the male. BC_1 progenies were used as females for further crosses to OH8245. The C.M. Rick Tomato Genetic Resources Center, University of California, Davis, provided the seed for LA1141. During these studies, IBC selections were further inbred to BC_2S_5. Selected BC_2S_5 progenies were advanced based on BC_2S_3 phenotypic values, genomic estimated breeding values (GEBVs), and random selection.

Seedling care for greenhouse and field trials followed the same protocol. Seeds were sown in 288-cell trays with a cell volume of 13 mL. Greenhouse temperatures were set to 27°C during the day and 25°C at night with a 16 h photoperiod. Photosynthetically active radiation (PAR) was supplied by natural sunlight, 1,000-W metal-halide lamps (Multi-Vapor® GE Lighting, East Cleveland, OH, United States), and 1,000-W high-pressure sodium lamps (Ultra Sun® Sunlight Supply, Vancouver, WA, United States) with a target threshold of 250 m^{-2} s^{-1} for natural sunlight before initiating artificial lighting to maintain light levels. PAR values in the greenhouse ranged from 250 to 637 μmol m^{-2} s^{-1} with an average of 391.4 μmol m^{-2} s^{-1}. Fertilization was applied using a 20-20-20 fertilizer (20% N, 20% P_2O_5, and 20% K_2O) (Jack's professional All-Purpose Fertilizer, JR Peters INC., Allentown, PA, United States) and delivered at a concentration of 200 mg L^{-1} twice per week. For greenhouse evaluations, plants with three to five expanded leaves were transferred to 3.7-L plastic containers, filled with PRO-MIX (Premier Horticulture, Quakerstown, PA, United States), and spaced 30 cm apart on a greenhouse bench. Transplants in 3.7-L plastic containers were automatically irrigated four times per day, for 6 min, at a volume of 220 mL per irrigation cycle.

Greenhouse evaluations of plants under water deficit

Greenhouse trials were conducted to evaluate water-deficit stress in the LA1141 and OH8245 parents, the segregating LA1141 × OH8245 BC_2S_3 families, and selections were advanced to the BC_2S_5 generation using phenotypic, genomic, and combined selection strategies or randomly chosen BC_2S_3 families advanced to the BC_2S_5 generation. The LA1141 and OH8245 parents were evaluated using a randomized complete block design with three blocks and three replicates within each block. The segregating BC_2S_3 families were evaluated in augmented designs as single replicates, with 36 replicates of each LA1141 and OH8245 parent distributed in a grid-like pattern corresponding to the row and column blocks across environmental gradients established between cooling pads, fans, and differences in solar radiation. The LA1141 and OH8245 parents were both used as over-replicated controls to correct for environmental variation in the greenhouse (Federer and Raghavarao, 1975; Bojacá et al., 2009). The BC_2S_3 families were evaluated in the greenhouse twice, in a summer environment and a fall environment. The BC_2S_5 advanced lines

selected randomly or using phenotypic, genomic, and combined strategies were evaluated in an augmented design with three blocks, and a single replicate within each block. In the BC_2S_5 evaluation, each LA1141 and OH8245 parent was replicated nine times in each unique row by column spatial designation to account for environmental variation in the greenhouse as described above.

Water deficit treatments followed the same protocol in all greenhouse evaluations. When the plants reached the growth stage of six to eight expanded leaves in 3.7-L containers, irrigation was simultaneously withheld on all genetic treatments, and plants were evaluated daily, and the first evaluation happened at saturation. Pots were weighed at saturation, after 72 h of deficit irrigation, and at 144 h to estimate evapotranspiration. A Decagon GS3 sensor (Decagon Devices, Pullman, WA, United States) was calibrated for soilless potting media and used to measure volumetric water content (VWC) at saturation. The Decagon GS3 sensor utilized capacitance probes that measure real dielectric permittivity values and are generally preferred for water deficit evaluations compared to instruments that use time-domain reflectometry to measure soil water content changes in greenhouse pot experiments (Sharma et al., 2017). Estimates of VWC and gravimetric evaluations confirmed that starting saturation was even and consistent at the beginning of each experiment.

The leaf turgor status of individual genetic treatments was assessed daily during the dry-down phase on a whole plant basis using a visual turgor index as an estimation of whole plant wilt severity. Turgor index ratings ranged from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead), consistent with previous studies (Waterland et al., 2010). The turgor index was used to rate whole plant wilt status with higher scores indicating higher turgidity. Canopy temperature was measured with a handheld infrared thermometer (IRT) (Zhuhai JiDa Huapu Instrument Co., Hong Kong) and with an image-based methodology (see section High-throughoutput thermal image analysis). Canopy temperature was estimated with IRT by measuring the surface temperature (°C) of two fully expanded leaves per plant. IRT-estimated canopy temperature was monitored daily under ambient environmental conditions in the greenhouse between 10:00 and 12:00. The IRT was calibrated against a standard laboratory thermometer using a water bath calibration method (Horwitz, 1999). Additionally, the IRT was set to a constant emissivity of 0.97 for measurements of plant canopy consistent with previous studies (Kustas et al., 2012). Image-based whole plant canopy temperature was estimated using a FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR, United States) and calibrated relative to water baths at a known temperature (Horwitz, 1999). The thermal camera was also set to a constant emissivity of 0.97. Images were captured against a standardized background using a 50 cm × 76 cm black polystyrene core foam board (Elmer's Westerville, OH, United States) at 1 m; extraction of temperature data from images is further detailed in section "High-throughput thermal image analysis." The IRT measurements and thermal images were recorded simultaneously.

The physiological response of LA1141, OH8245, and BC_2S_5 advanced selections were assessed daily during the dry-down phase and measured on a single, upper, and fully expanded leaf using the LI-600 porometer/fluorometer (LI-COR Biosciences, Lincoln, NE, United States). Measurements of stomatal conductance (g_{sw}), vapor pressure deficit (VPD), and light-adapted chlorophyll fluorescence (PhiPS2) were taken under ambient conditions in the greenhouse between 10:00 and 12:00.

Phenotypic data collected on LA1141 and OH8245 plants grown under water deficit were analyzed as a fixed-effects model using the function "lm" in the R core package version 4.0.3 (R Core Team, 2020). The model $Y_{ij} = \mu + genotype_i + Block_j + \varepsilon_{ij}$: where Y_{ij} was the response variable, μ was the mean response of the parents, $genotype_i$ represented the replicated LA1141 donor parent ($N = 9$) and OH8245 recurrent parent ($N = 9$), $Block_j$ was used to estimate variation in the greenhouse across the air movement gradient established between cooling pads and fans, and ε_{ij} was the associated experimental error. Factors with significant p-values ($p < 0.05$) were analyzed using Tukey's Honest Significant Difference test, with the "HSD.test" function in the R package Agricolae (De Mendiburu, 2017).

The random-effects model, $Y_{ij} = \mu + genotype_i + Row_{ij} + Column_{ij} + \varepsilon_{ij}$, was used to evaluate the BC_2S_3 families and over-replicated parental controls. The analysis was conducted using the "lmer" function in the R package lme4 (Bates et al., 2015). For this analysis, Y_{ij} was the response variable, $genotype_i$ represented 160 individuals from the population as single replicates, and 36 replicates each of OH8245 and LA1141 as over-replicated controls. The environmental terms Row_{ij} and $Column_{ij}$ were used to capture spatial variation within the greenhouse as described above, and ε_{ij} was the associated error. Data were analyzed in the R core package version 4.0.3 (R Core Team, 2020). The significance of the main effects in these tests was determined by comparing a fully parameterized model to a model with a single term dropped using a likelihood ratio test based on a chi-square distribution (Snijdgers and Bosker, 2012). A significant p-value was interpreted as evidence that the parameter dropped was important to the fit of the model, and Bayesian Information Content (BIC) values were used to confirm that the full model provided a better fit to confirm the significance of genetic and environmental terms.

The germplasm screen of BC_2S_3 families was repeated twice, once in July and once in November. Best Linear Unbiased Predictors (BLUPs) were extracted for genetic treatments in the model using the "ranef" function in lme4 (Bates et al., 2015). BLUP values for canopy temperature and turgor response variables represented the spatially adjusted values for each BC_2S_3 family in each germplasm screen. The BLUPs estimated for canopy temperature and plant turgidity and were then averaged across the two experiments and used

for the quantitative trait loci (QTL) mapping study and for developing phenotypic and genomic selection indices. Similarly, the random model described above was used to evaluate and analyze BC_2S_5 progenies in the selection strategies validation experiment. For this analysis, Y_{ij} was the response variable, $genotype_i$ represented 30 individuals from the population replicated three times. The LA1141 and OH8245 parents were each replicated nine times. Row_{ij} and $Column_{ij}$ were used to capture environmental variation as mentioned previously.

High-throughput thermal image analysis

Thermal images were captured with a FLIRONE GEN3 iOS thermal camera as described above. Images were analyzed in a workflow that entailed extracting radiometric data as an integer matrix using the "readflirJPG" function in the R package Thermimage (Tattersall, 2021). The function "raw2temp" was used to convert the raw values obtained from the binary thermal image into estimated temperature with the equation (Eq. 1):

$$temperature = PB/log(PR1/(PR2 * (raw + PO)) + PF) - 273.15 \quad (1)$$

Where *PB* was Planck's constant B, the *log* was the base 10 logarithm, *PR1* was Planck's constant R1, *PR2* was Planck's constant R2, *raw* was a 16-bit encoded value associated with the radiance hitting the thermal camera sensor, *PO* was Planck's constant O, *PF* was Planck's constant F, and –273.15 was the conversion of temperature from K to °C. The Planck constant values were calibration constants that are specific to each FLIR thermal camera and were extracted using the software ExifTool (Harvey, 2020) as implemented in the R package "Thermimage" using the function "flirsettings (camvals = -*Planck*)". The emissivity of the plant object, atmospheric temperature, and temperature of the ambient surroundings were extracted from the sensor on the FLIRONE GEN3 iOS thermal camera and were used to estimate plant canopy temperature.

Raw files with pixel values corresponding to the plant canopy surface temperature were then imported into the open-source software ImageJ (Java-based distribution Fiji) (Schindelin et al., 2012) for image processing and analysis. A collection of ImageJ functions and macros called "ThermImageJ" (Tattersall, 2019) was used to import data extracted from the thermal images, isolate regions of interest (ROI), and estimate plant canopy temperature. Histogram-based thresholding algorithms were accessed in the core ImageJ package using the steps "Image" > "Adjust" > "Threshold". Image thresholding was performed to isolate the plant canopy ROI from the background of the image and employed the labeling operator *Tr* which labeled a pixel in the image only if its intensity *g* exceeded a specific threshold value *Tmin* (Eq. 2):

$$Tr\alpha\ [g(x) \geq Tmin] : g(x) \rightarrow 1(x) \quad (2)$$

Where 1 was the binary pixel label denoting the foreground value and *g* was the pixel intensity at the vector *x* (Prodanov and Verstreken, 2012). *Tmin* values can be determined using specific thresholding methods and we used "MEAN" described in Glasbey (1993). The "MEAN" method was chosen for our workflow because the proportion of total phenotypic variance attributed to genetic factors associated with canopy temperature was higher than the other tested thresholding methods (**Supplementary Table 1**). The "MEAN" thresholding algorithm was applied following these steps: "MEAN" > "Apply" > "set to NaN", resulting in a 32-bit image with the plant canopy ROI isolated from the image background. We estimated canopy temperature from our final image following the steps: "Analyze" > "Measure", which calculated the average estimated surface temperature of the plant canopy remaining in the image. Macros were written to automate and standardize the processes described above and to facilitate batch image analysis.

Comparisons of phenotyping methods

Phenotypic measurements for selected BC_2S_5 progenies were compared using linear regression and variance partitioning to identify traits with higher genetic variance. The slope of the regression line was used to estimate the relationship of thermal image estimated canopy temperature to IRT, turgor ratings, and physiological traits measured with the LI-600. The random model described in section 3.2, above, was used to estimate the proportions of variance due to genetics ($genotype_i$), environmental factors (Row_{ij} and $Column_{ij}$), and experimental error (ε_{ij}) in the models. The proportions of environmental and genetic variance provided an estimate of the broad-sense heritability and repeatability of each trait (Falconer and Mackay, 1996).

Genotyping and linkage map construction

Genomic DNA was isolated from fresh, young leaf tissue from the 160 BC_2S_3 families and parental lines using a modified CTAB method (Fenstemaker et al., 2022). The 157 polymorphic SNP markers and linkage map construction were described previously (Fenstemaker et al., 2022). Briefly, a genetic linkage map was constructed using the R/qtl package version 1.47-9 in the R statistical software environment version 4.0.3 (Broman et al., 2003; Shannon et al., 2013; R Core Team, 2020) based on the LA1141 × OH8245 BC_2S_3 population (Fenstemaker et al., 2022). Marker data corresponding to the LA1141 × OH8245 BC_2S_3 was deposited in Zenodo.[1] Map construction with the BC_2S_3 population was a compromise

1 https://doi.org/10.5281/zenodo.5650152

that provided genetic structure, captured a high percentage of elite parent background underlying important agronomic and horticultural traits, and offered opportunities to fix desirable donor alleles in further generations.

Quantitative trait loci analysis in the LA1141 x OH8245 BC_2S_3 population

The QTL analysis was conducted with composite interval mapping (Zeng, 1994) using the "cim" function in the R/qtl package. Analysis was performed using a 2 centimorgan (cM) step, one marker selected as a cofactor, and a marker window set to 40 cm. The marker cofactor number and window size was due to limited recombination in the BC_2S_3 population. The Haley Knott regression (Haley and Knott, 1992) method was used for QTL detection. A significance threshold of LOD = 3.3 for both canopy temperature and plant turgor was determined by resampling the data (α = 0.05, n = 1,000; Churchill and Doerge, 1994). A cut-off of LOD = 2.4 significance threshold corresponded to $p < 0.01$. Genetic effects were estimated as differences between phenotypic averages expressed as regression coefficients using the "fitqtl" function with the argument "get.ests = TRUE" and "dropone = FALSE" in the R/qtl package. The percentage of phenotypic variance explained was estimated using the "fitqtl" function with the argument "dropone = TRUE" in the R/qtl package.

Genomic selection models

Genomic estimated breeding values were calculated using ridge regression (RR) as implemented in the rrBLUP package in the R statistical software environment version 4.0.3 (Endelman, 2011; R Core Team, 2020). The RR computations were performed using the function "mixed.solve" in rrBLUP. Markers were considered as random effects associated with plant turgor and canopy temperature response variables. The estimated marker effects were used to calculate the GEBV of each LA1141 × OH8245 BC_2S_3 family. The equation used was: GEBV = X × MV: where GEBV was the vector of dimension (n, 1) containing the GEBVs for n families, X (n, m) was the matrix of scores for m markers and n families, and MV (m, 1) was a vector of marker effects for the m markers. GEBVs estimated from RR were used for genomic selection.

Validation and comparison of selection methods

The phenotypic selection was based on the BC_2S_3 family visual ratings corresponding to plant turgor and canopy temperature values estimated using image-based methods. Phenotypic values were expressed as BLUPs estimated from the population screens described in section "High Throughput Phenotyping Using Thermal Images." The BLUP values corresponding to each trait were sorted numerically and assigned a rank. The BC_2S_3 family with the highest value associated with plant turgor was ranked 1 (best), and the family with the lowest value associated with canopy temperature was ranked 1 (best). Plant turgor and canopy temperature BLUP ranks were summed into a single value that resulted in a rank-sum list. Top-ranking progenies (N = 10) were chosen according to this simplified multi-trait index (MTI). The selection intensity K was defined in standard deviation (SD) units relative to the mean. All phenotypic selections for plant turgor were made at a selection intensity of K = 1. Three canopy temperature phenotypic selections were made at a selection intensity of K = 2, and seven were made at a selection intensity of K = 1.

The genomic selection model mentioned above was used to calculate GEBVs for each trait. The GEBVs were sorted numerically and assigned a rank as described above. Canopy temperature GEBV and plant turgor GEBV ranks were summed into a single value used as an MTI as described in the section above, and top progenies were chosen based on this GS selection index. The top six selections for plant turgor GEBV were made at K = 2, with four additional selections at K = 1. All genomic selections for canopy temperature were made at K = 2. The final groups of selections consisted of phenotypic selections (Pheno, N = 9), genomic selections (GS, N = 8), randomly selected (Random, N = 10) families, and a group that was co-selected as top-ranking phenotypic and genomic selections (Pheno + GS, N = 3). These groups consisted of the top 10 ranking phenotypic selections, top 10 ranking genomic selections, and 10 randomly selected families totaling 30 progenies for further inbreeding.

The prediction abilities of the GS models were evaluated using cross-validation (theoretical accuracy) and empirical validation (realized accuracy). Cross-validation was conducted using a leave-one-out strategy (Liabeuf et al., 2018) on the BC_2S_3 families, which were considered our training population. Empirical validation was conducted using greenhouse performance data for plant turgor and canopy temperature measured on advanced BC_2S_5 lines derived from inbreeding selected BC_2S_3 families from the training population. The abilities of GS models to predict performance were estimated by two different models. First, the cross-validation prediction accuracy (r_g) was evaluated using the Pearson coefficient of the correlations between GEBVs and phenotypic BLUPs in the BC_2S_3 families. Second, empirical validation r_g was evaluated using the Pearson coefficient of the correlations between the BC_2S_3 family GEBV and phenotypic BLUPs in the advanced BC_2S_5 selections. Additionally, the percentage of co-selection (% co-selection) was calculated using the number of selected families identified as both top 10 ranking phenotypic values at a minimum selection intensity of K = 1 and top 10 ranking GEBVs at a minimum selection intensity of K = 2 divided by the total number BC_2S_3 families selected using phenotypic values and GEBVs (N = 20).

The LA1141 × OH8245 BC_2S_3 phenotypic, genomic, and randomly selected families were advanced to the BC_2S_5 generation and evaluated in an augmented design in the greenhouse as described in section "Greenhouse evaluations of plants under water deficit."

Trait BLUP values from selections were combined according to selection strategy, which was considered a treatment factor in the analysis. A fixed-effects model with selection strategy as a factor was used to determine if BC_2S_5 progenies chosen using selection strategies were significantly different from randomly advanced lines. The linear model used was $Y \sim \mu + Selection + \varepsilon$: where Y was the trait BLUP, μ was the trial mean, *Selection* was the selection strategy used (GS, Phenotypic, Pheno + GS, and Random), and ε was experimental error. Factors with a significant p-value ($p < 0.05$) were analyzed using Tukey's Honest Significant Difference with the "HSD.test" function in the R package Agricolae (De Mendiburu, 2017). The traits evaluated in the selection strategy validation experiment were turgor, canopy temperature, g_{sw}, and VPD.

Evaluation of selections for horticultural traits

Advanced genomic and phenotypic selections were also evaluated in a field trial to assess whether OH8245 horticultural traits were recovered. The field trial was designed as a randomized complete block design with two blocks and a single replicate in each block. The experimental unit was the plot. The field site was located at the Horticulture Unit 1 Research Farm in Wooster, Ohio. Each plot consisted of seven to ten plants and was spaced 30 cm apart in rows, with each row separated by 1.5 m. Maintenance of field plots followed standard practices for tomato production in the Midwest (Philips et al., 2021 accessed at: mdc.itap.purdue.edu). The genetic treatments consisted of the OH8245 recurrent parent, the BC_2S_5 phenotypic selections (N = 10), and BC_2S_5 genomic selections (N = 10). Seedlings were transplanted to the field four weeks after emergence. Plots were harvested when 80% of fruit in a plot reached the red ripe stage of maturity, which averaged 107 Julian calendar days. Before harvest, the plant canopy's width and height were measured in cm by hand. Three plants were hand-harvested from the middle of each plot. The fruit was sorted into ripe, green, and cull maturity categories, and each group was weighed separately. Cull fruits were fruit with cracks, blemishes, or disease. Total yield was measured as the combined harvested weight of the three groups. A sub-sample of 20 fruit was analyzed using color and chemical traits associated with tomato fruit quality. Fruit color was measured on a cross-section of 11 fruit and shoulder cuts of nine fruit using the image-based software Tomato Analyzer (Darrigues et al., 2008). Soluble solids (Brix°) were quantified by filtering juice through a Kimwipe™ (Kimberly-Clark Corp., Neenah, WI, United States) and measured using a handheld refractometer (PAL-1, Atago U.S.A., Bellvue, WA, United States).

Field performance of OH8245 and BC_2S_5 advanced lines that were chosen using phenotypic and genomic selection was evaluated using the fixed effects model $Y_{ij} = \mu + genotype_i + Block_j + \varepsilon_{ij}$: where Y_{ij} was the response variable, $genotype_i$ represented the BC_2S_5 selections (N = 20), and the OH8245 recurrent parent, $Block_j$ was replication within the field, μ was the trial mean associated with the yield or quality parameters, and ε_{ij} was the error. Genetic and environmental factors with a significant *p-value* ($p < 0.05$) were analyzed with Tukey's Honest Significant Difference using the "HSD.test" function in the R package Agricolae (De Mendiburu, 2017).

Results

LA1141 under water deficit stress

Based on plant turgor, canopy temperature, and physiological measurements, LA1141 is more tolerant to water deficit stress than the OH8245 recurrent parent (**Figures 1**, **2** and **Supplementary Table 2**). Experiments were conducted over 144 h of deficit irrigation, with differences in turgor and canopy temperature observed between parents at 48 h through termination of the experiment (**Supplementary Table 2**). For simplicity, turgor and canopy temperatures are reported when they reach their maximums. Significant differences in turgor between LA1141 and OH8245 (p = 1.50e-15) are shown at 72 h (**Figures 1**, **2A** and **Supplementary Table 2**). Accession LA1141 maintained a lower canopy temperature as measured

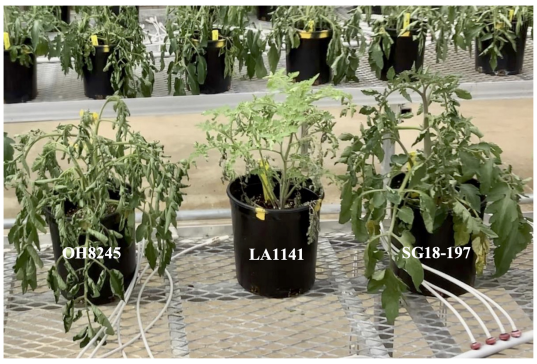


FIGURE 1
Response of, *Solanum lycopersicum* OH8245, *S. galapagense* accession LA1141 and LA1141 × OH8245 inbred backcross line SG18-197 at 72 h of water deficit. Turgor ratings ranged from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead), consistent with previous studies (Waterland et al., 2010). Accession LA1141 (labeled) **(center)** received a rating of 5, OH8245 **(left)** received a rating of 3, and LA1141 × OH8245 inbred backcross line SG18-197 **(right)** received a rating of 4.

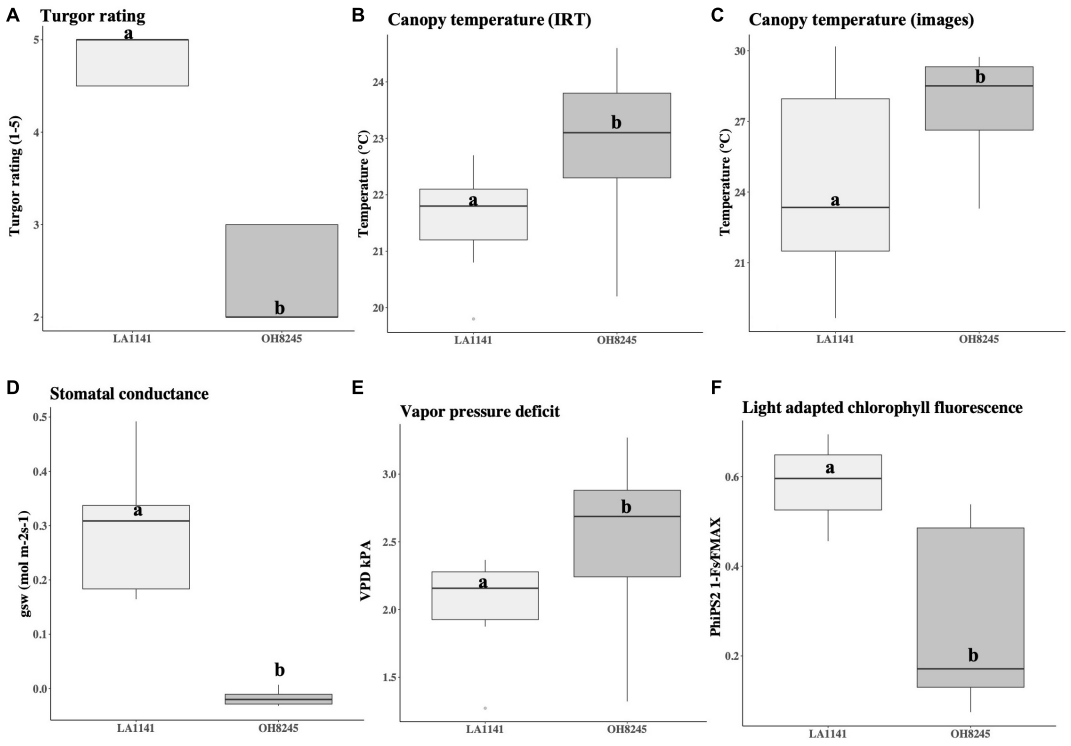


FIGURE 2
Boxplots comparing water-deficit stress response of the *Solanum galapagense* accession LA1141 donor parent (N = 9) to the *S. lycopersicum* OH8245 recurrent parent (N = 9). The traits measured were **(A)** turgor ratings ranging from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead) consistent with previous (Waterland et al., 2010). **(B)** Canopy temperature estimated using a handheld infrared thermometer (IRT) (Zhuhai JiDa Huapu Instrument Co., Hong Kong), and **(C)** from images captured with the FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR, United States). Physiological measurements were taken with the LI-600 Porometer/ Fluorometer (LI-COR Biosciences, Lincoln, NE, United States) and included **(D)** stomatal conductance (g_{sw} mol m^{-2} s^{-1}), **(E)** vapor pressure deficit (VPD kPa), and **(F)** light-adapted chlorophyll fluorescence (PhiPS2 1-Fs/FMAX). Values are reported at the time point where they reach their maximums. Different letters indicate statistically significant differences among groups (Tukey's Honest Significant Difference, $p < 0.05$).

by both an infrared thermometer (IRT) (p = 0.032) and thermal images (p = 0.049) (Figures 2B,C and Supplementary Table 2). Accession LA1141 and OH8245 exhibit different physiological responses, which become significant at 72 h of deficit irrigation (Supplementary Table 2). LA1141 maintains higher stomatal conductance (g_{sw}) (p = 1.80e-07), lower vapor pressure deficit (VPD) (p = 0.025), and higher light-adapted chlorophyll fluorescence (PhiPS2) (p = 0.009) compared to OH8245 (Figures 2D–F and Supplementary Table 2). Observable and measurable differences in response between accession LA1141, OH8245, and their progenies (Figure 1) provided the basis for genetic studies describing water deficit tolerance derived from *S. galapagense*.

High-throughput phenotyping using thermal images

Thermal image-based phenotyping and analysis detected greater differences in canopy temperatures between genotypes under water deficit compared to canopy temperature measured with the infrared thermometer (IRT) (Figure 3). To test whether the high-throughput thermal image analysis pipeline offered advantages over the IRT, we evaluated LA1141, OH8245, and LA1141 × OH8245 BC_2S_3 families and BC_2S_5 selections for canopy temperature using both phenotyping approaches. Regression of canopy temperature measured with the IRT to values estimated by thermal images in the BC_2S_3 families showed a significant linear relationship (p < 2.20e-16, R^2 = 0.30) (Supplementary Figure 3). Variance components for "Genotype" and experimental factors attributed to the environment, including "Row," "Column," and "error," were partitioned to estimate the proportion of genetic variance associated with canopy temperature (Table 1). The total phenotypic variation partitioned into genetic effects associated with canopy temperature measured with IRT was 19.23% (Table 1). In contrast, the percentage of total phenotypic variance attributed to genetic factors measured with thermal images was 22.16% (Table 1). Thermal image estimated canopy temperature, therefore, provides higher repeatability.

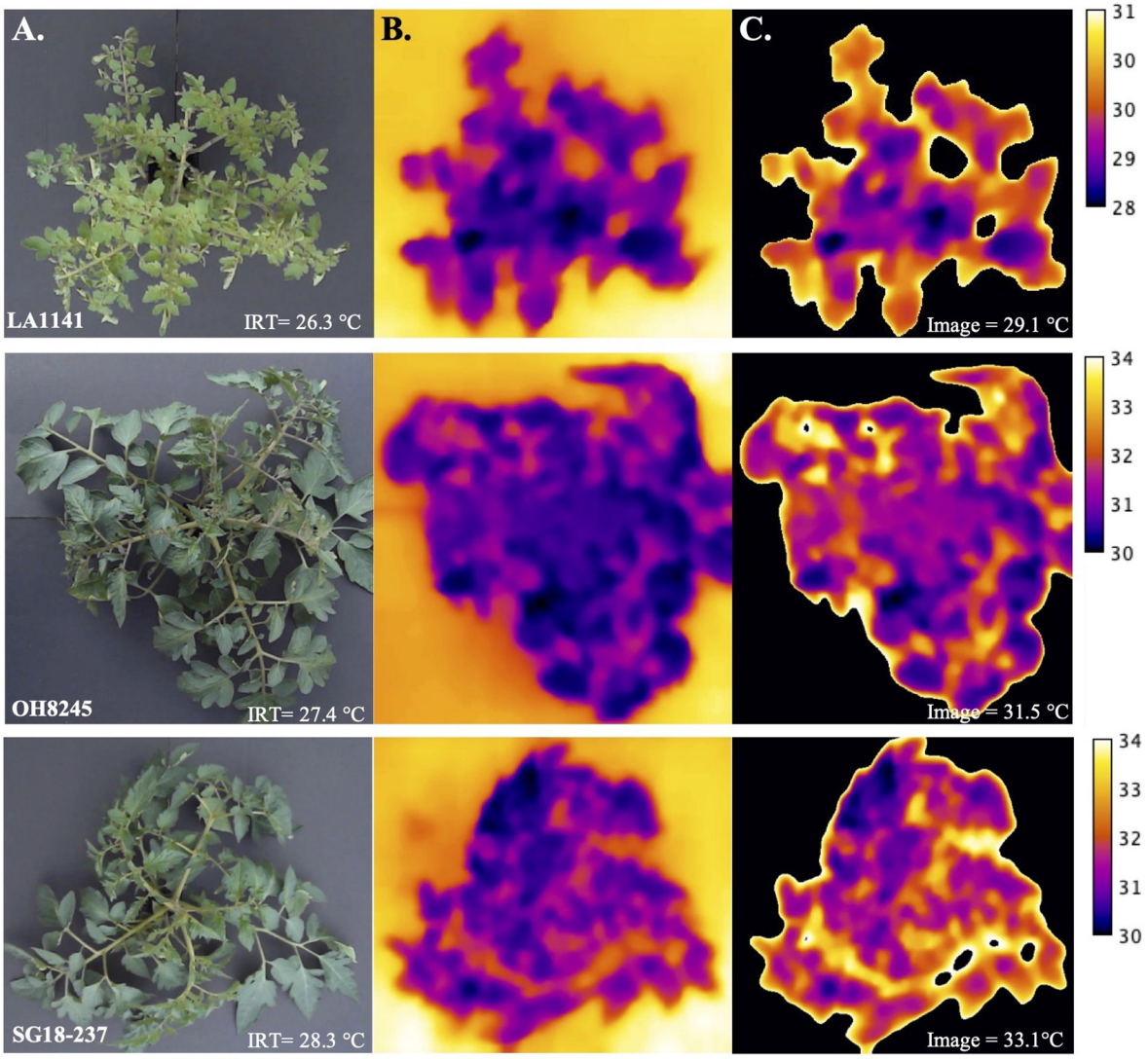


FIGURE 3
Estimating canopy temperature (°C) using images and an infrared thermometer (IRT) as a measure of plant water deficit stress. **(A)** Original images of LA1141 (top), OH8245 (middle), and inbred backcross line SG18-237 (bottom). The original image of the plant canopy was captured with the FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR) and temperature (°C) was measured using a handheld IRT (Zhuhai JiDa Huapu Instrument Co., Hong Kong) simultaneously. **(B)** Plots of the temperature data extracted from thermal images using the function "plotTherm" in the R package Thermimage (Tattersall, 2021). **(C)** Histogram-based thresholding using "MEAN" (Glasbey, 1993) in the JAVA-based distribution of ImageJ "Fiji" (Schindelin et al., 2012). Images were analyzed in a workflow that entailed extracting radiometric data using the "readflirJPG" function in the R package Thermimage (Tattersall, 2021). Raw files with pixel values corresponding to the plant canopy surface temperature were then imported into the open-source software ImageJ (Schindelin et al., 2012) for image processing and analysis. A collection of ImageJ functions and macros called "ThermImageJ" (Tattersall, 2019) was used to import the temperature data extracted from the thermal images, isolate regions of interest (ROI), and estimate temperature. The image calibration bar was added to each individual image in panel C using the ImageJ macro "ThermImageJ" (Tattersall, 2019).

Comparisons between image-based canopy temperature at 48 h and turgor ratings at 72 h show a significant correlation (p = 0.031), suggesting that canopy temperature can predict the onset of wilt (**Supplementary Figure 4A**). Image-based canopy temperature measurements at 48 h are also significantly correlated to g_{sw} (p = 0.023) and VPD (p = 0.002), but not to PhiPS2 (p = 0.583) at 72 h, suggesting that canopy temperature is predictive of g_{sw} and VPD (**Supplementary Figures 4B–D**). Notably, a higher percentage of phenotypic variance is partitioned into genetic effects associated with canopy temperature than g_{sw}, VPD, and PhiPS2 (**Table 1**). However, the percentage of phenotypic variance partitioned into genetic effects associated with turgor is higher than canopy temperature and physiological measurements (**Table 1**). Still,

TABLE 1 Percentage of total variance estimates for turgor ratings, canopy temperature, and LI-600 physiological measurements in *S. galapagense* LA1141, *S. lycopersicum* OH8245, and LA1141 × OH8245 BC_2S_5 progenies.

Sources of variation[a]	Turgor[b]	Image (°C)[c]	IRT (°C)[d]	g_{sw} mol $m^{-2}s^{-1}$[e]	VPD kPa[f]
Genotype	45.55	22.16	19.23	16.99	10.87
Row	11.40	0.00	0.04	2.59	2.19
Column	5.64	2.80	25.56	17.99	12.37
Residual	37.41	75.04	55.17	62.43	74.57

[a]Genotype is represented by *S. galapagense* accession LA1141 (N=9) and *S. lycopersicum* OH8245 (N=9), and advanced inbred progenies that have been backcrossed to OH8245 two times and self-pollinated five times (BC_2S_5) (N = 30, replicated three times). Row and Column were used as environmental terms to capture spatial variation across the greenhouse and each row by column location contained both replicated parental controls.
[b]Plant turgor ratings ranged from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead) consistent with previous studies (Waterland et al., 2010).
[c]Whole plant canopy temperature (°C) measured using a FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR, United States).
[d]Leaf surface temperatures (°C) of two fully expanded leaves per plant measured with infrared thermometer (Zhuhai JiDa Huapu Instrument Co., Hong Kong).
[e]Stomatal conductance to H2O (mol $m^{-2}s^{-1}$) measured with the LI-600 porometer/fluorometer (LI-COR Biosciences, Lincoln, NE, United States).
[f]Vapor pressure deficit kPa at leaf temperature measured with the LI-600 porometer/fluorometer (LI-COR Biosciences, Lincoln, NE, United States).

an image-based canopy temperature appears to be a suitable proxy for more intensive physiological measurements, such as g_{sw} and VPD in our water deficit germplasm screens. Finally, the ability of images to predict the onset of wilt demonstrates that image-based measurements can improve the efficiency of germplasm screens.

LA1141 × OH8245 BC_2S_3 families under water deficit stress

The BC_2S_3 families that differed from the trial mean by a selection intensity of K = 1 indicated tolerance or susceptibility to water deficit stress. Tolerance in specific LA1141 × OH8245 BC_2S_3 families for canopy temperature and turgor was recovered. Germplasm evaluations were conducted in the greenhouse in two seasonal environments (July and November). A summary of greenhouse conditions in summer and fall environments is provided (**Supplementary Table 5**). Genetic effects for turgor (p < 2.20e-16) and canopy temperature (p = 8.17e-07) were significantly different (**Supplementary Table 6**). Additionally, summer and fall environments were significantly different for turgor (p < 2.20e-16) but not significantly different for canopy temperature (p = 0.736) (**Supplementary Table 6**). The environmental term "Column" corresponds to a solar radiation gradient in these experiments. The term "Row" corresponds to an air movement gradient between the greenhouse cooling pad and fans. The interaction between the seasonal environment and row, or column, terms represent a unique greenhouse position within each environment. The "Environment × Row" term was significantly different for turgor (p = 0.007) but not canopy temperature (p = 0.371) (**Supplementary Table 6**). However, Environment × Column was significantly different for both turgor (p = .0002) and canopy temperature (p = 5.05e-14) (**Supplementary Table 6**). The experimental design used over-replicated checks to estimate the best linear unbiased predictors (BLUPs) as described in the methods section "Greenhouse evaluations of plants under water deficit" to account for the variation in the greenhouse described above. Trait values expressed as BLUPs exhibit shrinkage around the mean and provided conservative estimates of turgor and canopy temperature adjusted to environmental differences based on the over-replicated LA1141 and OH8245 parental checks.

Additionally, estimates of evapotranspiration were not significantly different based on genotype (p = 0.461) but were significantly different between experimental environments (p = 3.42e-08) (**Supplementary Table 6**). However, the environmental interaction factors Environment × Row (p = 0.089) and Environment × Column (p = 0.732) were not significantly different (**Supplementary Table 6**). These results suggested that the position in the greenhouse and genetic differences for estimated evapotranspiration within an experimental environment did not explain significant differences in water-deficit stress response. Differences detected between the screening environments are not surprising because the average temperature in the greenhouse was higher in the summer compared to the fall (**Supplementary Table 5**). Consequently, observed genetic variation in the germplasm appears to be independent of estimated evapotranspiration, both for canopy temperature and plant turgor. These traits were subsequently used in interval mapping studies, genomic selection models, and the development selection indices.

Quantitative trait loci analysis in the LA1141 × OH8245 BC_2S_3 population

Four putative QTLs were identified, and explained between 6.6 and 9.49% of the phenotypic variation for canopy temperature and turgor (**Figure 4** and **Table 2**). Phenotypic values were expressed as an average of trait BLUPs across environments with a mean of 0, and the effect of an allele substitution was expressed relative to the mean. All QTLs

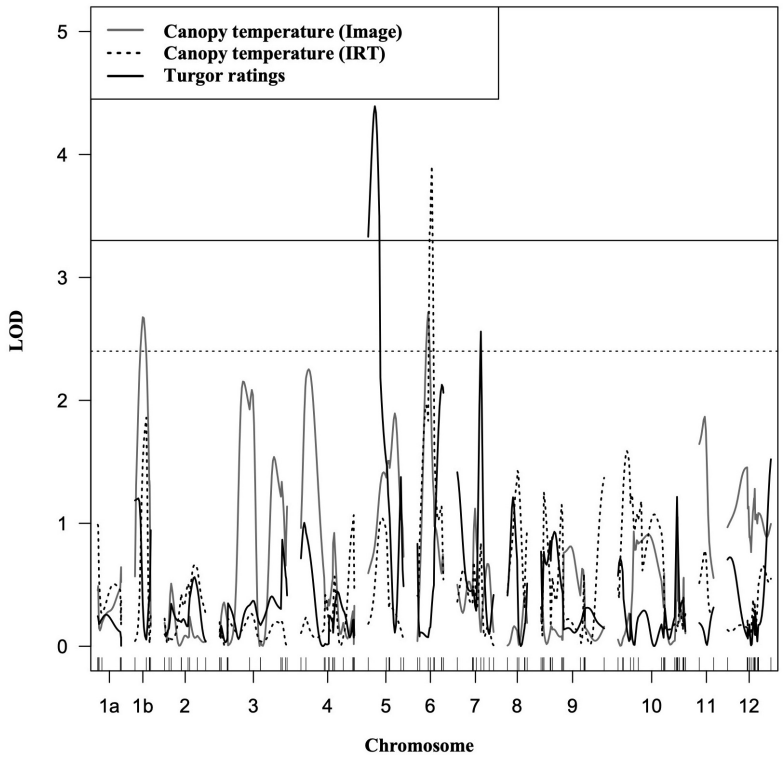


FIGURE 4
Composite interval mapping of LA1141-derived tolerance to water deficit stress. The traits measured were canopy temperature estimated using images (**grey**), canopy temperature estimated with an infrared thermometer (dotted), and plant turgor ratings (black) ranging from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead) consistent with previous studies (Waterland et al., 2010). Traits were measured in the LA1141 × OH8245 BC_2S_3 inbred backcross families (N = 160). The y-axis is the logarithm of the odds (LOD). The solid (black) horizontal line (LOD = 3.3) is the resampled LOD significance cutoff (α = 0.05, N = 1,000 permutations). The dotted (gray) horizontal line (LOD = 2.4) represents a significance level of $p < 0.01$. The x-axis represents linkage groups corresponding to the 12 chromosomes in tomato and chromosome distance in centimorgans (cM). The genetic distance was calculated using the Kosambi function Kosambi (1943) to correct for multiple crossovers.

contributing to water deficit tolerance were derived from the LA1141 donor parent (Table 2). A region on the distal arm of chromosome 1 (linkage group 1b) had a LOD score of 2.66, explained 6.6 % of the total phenotypic variation, and lowered canopy temperature by 0.02°C (Table 2). A region on the proximal arm of chromosome 6 had a LOD score of 3.46, explained between 6.02 and 9.09 % of the total phenotypic variation, and lowered canopy temperature by 0.03°C (Table 2). A region on the proximal arm of chromosome 5 had a LOD score of 3.33, explained 9.14 % of the total phenotypic variation, and increased turgor ratings between 0.33 and 0.36 units (Table 2). A region on the distal arm of chromosome 7 had a LOD score of 2.5, explained 6.78% of the phenotypic variation, and increased ratings associated with higher turgor by 0.42 units (Table 2). The QTL found on chromosomes 1, 5, and 7 were detected in both individual screens and the combined dataset (Table 2). The QTL associated with canopy temperature detected on chromosome 6 was detected in the summer and combined environments but not the fall environment (Table 2).

Validation of selection strategies

Prediction accuracy (r_g) was evaluated with cross validation in the LA1141 × OH8245 BC_2S_3 training population and empirically in BC_2S_5 lines derived from further inbreeding. Cross-validation correlations between GEBVs and phenotype were significant for canopy temperature (p = 1.53e-15), and accuracy was r_g = 0.57. Similarly, correlations for turgor were significant (p = 4.61e-09) with an accuracy of r_g = 0.44 (Table 3). Correlations between GEBVs based on the BC_2S_3 training population and observed values of BC_2S_5 lines were significant for canopy temperature (p = 0.009, r_g = 0.57) and turgor (p = 0.021 r_g = 0.31) (Table 3).

Improvement in plant performance in our experiments is demonstrated by high plant turgor, low canopy temperature, high g_{sw}, and low VPD. Selection strategies based on phenotype, genomic selection models and a combination of the two were compared to randomly advanced lines. The selection strategy was significantly different from random selections for turgor (p = 0.004), canopy temperature (p = 0.006), g_{sw} (p = 0.026),

TABLE 2 Quantitative trait loci (QTL) and flanking single nucleotide polymorphism (SNP) markers associated with canopy temperature and turgor ratings in *S. galapagense* LA1141 x *S. lycopersicum* OH8245 BC_2S_3 families.

	LA1141 X OH8245 BC_2S_3									
Trait[a]	SNP marker	LOD[b]	*p*[c]	QTL × seasonal environments[d]	Donor allele	Allele substitution effect[e]	Percent phenotypic variance explained[f]	Chromosome	Physical position (bp)[g]	Genetic position (cM)[h]
Thermal image canopy temperature	solcap_snp_sl_2234	0.00	0.998	Summer, fall, combined	LA1141	0.00	0.001	1b	79025804	00.00
	solcap_snp_sl_14323	2.66	0.004	Summer, fall, combined	LA1141	−0.02	6.60	1b	87223580	20.37
	solcap_snp_sl_13404	1.41	0.041	Summer, fall, combined	LA1141	−0.02	3.98	1b	88561836	25.81
	solcap_snp_sl_14458	2.71	0.002	Summer, combined	LA1141	−0.03	7.18	6	36520866	19.42
	solcap_snp_sl_1337	2.19	0.007	Summer, combined	LA1141	−0.03	6.02	6	37305722	23.28
	solcap_snp_sl_12757	1.00	0.090	Summer, combined	LA1141	−0.01	2.8	6	38186675	29.74
IRT canopy temperature	solcap_snp_sl_14458	1.97	0.010	Summer, combined	LA1141	−0.02	5.53	6	36520866	19.42
	solcap_snp_sl_1337	3.46	0.000	Summer, combined	LA1141	−0.03	9.49	6	37305722	23.28
	solcap_snp_sl_12757	2.15	0.007	Summer, combined	LA1141	−0.03	6.01	6	38186675	29.74
Turgor ratings	solcap_snp_sl_19102	3.33	0.000	Summer, fall, combined	LA1141	0.33	9.14	5	1909149	00.00
	solcap_snp_sl_5050	1.84	0.015	Summer, fall, combined	LA1141	0.36	5.15	5	6045160	32.00
	solcap_snp_sl_22065	0.26	0.54	Summer, fall, combined	LA1141	0.07	0.76	7	3718124	34.18
	solcap_snp_sl_5861	2.5	0.003	Summer, fall, combined	LA1141	0.42	6.78	7	59688274	42.51
	solcap_snp_sl_7025	0.104	0.29	Summer, fall, combined	LA1141	0.08	0.97	7	63561726	57.45

[a]Tolerance to water deficit measured as canopy temperature and plant turgor ratings in the OH8245 × LA1141 families that were backcrossed twice to OH8245 and self-pollinated three times (BC_2S_3). Thermal image canopy temperature and Infrared thermometer (IRT) both represent maximum canopy temperature values. Trait values were expressed as Best Linear Unbiased Predictors (BLUPs).

[b]Logarithm to base 10 (LOD) scores. A significance threshold of LOD = 3.3 for both canopy temperature and plant turgor was determined by resampling the data ($\alpha = 0.05$, $n = 1{,}000$; Churchill and Doerge, 1994). A cut-off of LOD = 2.4 significance threshold corresponded to $p < 0.01$.

[c]The *p-value* retrieved using the Haley-Knott regression formula: y ~ Q1, where y is the response variable and Q1 is the marker.

[d]Significant (p=<.01) QTLs detected in summer environments, fall environments, and in the combined dataset.

[e]Genetic effects evaluated as differences between phenotype averages expressed as regression coefficients.

[f]Percent variance explained estimated by $1 - 10^{-2\ LOD/n}$, where n is the sample size and LOD is the LOD score for the marker.

[g]Physical position in base pairs corresponds to the Tomato Genome version SL4.0 (Hosmani et al., 2019).

[h]Genetic position corresponds to the LA1141 x OH8245 BC_2S_3 linkage map previously developed (Fenstemaker et al., 2021).

10.3389/fpls.2022.947538

TABLE 3 Evaluation of accuracy and genetic gain for selection strategies during inbreeding of the *S. galapagense* LA1141 × *S. lycopersicum* OH8245 population.

Population[a]		Canopy temperature °C[b]	Turgor ratings[c]
BC_2S_3 families	Minimum	18.50	1
	Maximum	28.90	5.00
	Mean	24.50	3.14
	s.d.[d]	1.56	1.18
	Cross validation (r_g)[e]	**0.57 (p = 1.53e-15)**	**0.44 (p = 4.61e-09)**
BC2S5 families	Minimum	18.05	2.16
	Maximum	27.74	4.07
	Mean	22.13	3.05
	s.d.	2.44	0.53
	Empirical validation (r_g)[f]	**0.47 (p = 0.009)**	**0.31 (p = 0.021)**
BC_2S_5 GS	Minimum	22.01	2.61
	Maximum	25.97	3.49
	Mean	23.71	3.15
	s.d.	1.51	0.38
	Genetic gain[g]	**-0.79**	**0.01**
BC_2S_5 Pheno + GS	Minimum	20.66	2.76
	Maximum	25.37	3.93
	Mean	23.07	3.44
	s.d.	2.35	0.61
	Genetic gain	**-1.43**	**0.31**
BC_2S_5 Phenotype	Minimum	22.38	2.76
	Maximum	25.06	4.08
	Mean	23.93	3.41
	s.d.	0.86	0.47
	Genetic gain	**-0.57**	**0.27**
BC_2S_5 Random	Minimum	22.41	2.17
	Maximum	30.05	3.35
	Mean	26.81	2.64
	s.d.	2.87	0.39
	Genetic gain	**2.31**	**-0.5**

[a]Population represents LA1141 × OH8245 were backcrossed twice to OH8245 and self-pollinated three tiems (BC_2S_3) (N = 160) and advanced selections that underwent additional self-pollination (BC_2S_5) based on genomic estimated breeding values (GEBVs) (GS, N = 8), LA1141 × OH8245 BC_2S_3 canopy temperature and turgor best linear unbiased predictors (BLUPs) (Pheno, N = 9), a combination of the two (GS + Pheno, N = 3), and randomly advanced lines (Random, N = 10).

[b]Canopy temperature measured as whole plant canopy temperature (°C) using a FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR, United States).

[c]Plant turgor based on a rating scale ranged from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1=dead) consistent with previous studies (Waterland et al., 2010).

[d]Standard deviation.

[e]Correlation coefficient between genomic estimated breeding values (GEBVs) and phenotypic values in BC_2S_3 progenies (training population).

[f]Correlation coefficient between genomic estimated breeding values (GEBVs) and phenotypic values in the advanced BC_2S_5 progenies.

[g]Increase in performance through selection (lower canopy temperature and higher turgor scores). Bold values are self-evident.

and VPD (p = 0.046) (Supplementary Table 7). Phenotypic, genomic, and combined strategies resulted in positive gain under selection, with selected lines showing higher turgor and lower canopy temperatures under water deficit. Differences between the BC_2S_3 families and BC_2S_5 lines chosen based on GEBV resulted in gain under selection by increasing turgor by 0.01 units (Table 3). Lines selected using combined strategies resulted in gain under selection by increasing turgor ratings by 0.31 on the five-point scale (Table 3). Phenotypically selected BC_2S_5 advanced lines resulted in gain under the selection of 0.27 units (Table 3). In contrast, differences between the BC_2S_3 training population and the randomly advanced BC_2S_5 progenies did not result in gain under this selection and lowered turgor ratings by 0.50 on the five-point scale (Table 3). Importantly, phenotypic, and combined strategies had higher turgor compared to randomly advanced progenies (Figure 5A). However, chosen progenies using genomic selection alone did not have different turgor ratings when compared to phenotypic, combined, or random selection (Figure 5A).

Similarly, selection for lower canopy temperature resulted in observed genetic gains. Differences between the BC_2S_3 families and BC_2S_5 lines chosen based on GEBVs for canopy temperature lowered the canopy temperature to 0.79°C (Table 3). Lines selected using combined strategies lowered the canopy temperature to 1.43°C (Table 3). Phenotypically selected BC_2S_5 progenies resulted in gain under selection by lowering canopy temperature to 0.57°C (Table 3). In contrast, random selection raised canopy temperature by 2.31°C (Table 3). In conclusion, all selection strategies lowered canopy temperature compared to random selection (Figure 5B).

Selected water deficit tolerance based on high turgor and low canopy temperature also had the result of improving physiological measurements. Statistical differences for physiological measurements were observed for advanced lines chosen using combined selection strategies. Selected lines based on combined strategies maintained higher g_{sw} and lower VPD compared to randomly advanced lines (Figures 5C,D) suggesting that genomic selection can be used to augment phenotypic selection and indirectly select for improved plant physiological status under water deficit. Therefore, genomic, and phenotypic selection for turgor and canopy temperature are indirect methods to improve plant physiological response under water deficit stress. On average, genomic, phenotypic, and the combined strategies had g_{sw} values that were 0.007 mol H_2O m^{-2} s^{-1} higher and VPD values that were 0.047 kPa lower relative to randomly advanced lines (Figure 5). These results suggested that progress toward tolerance to water deficit stress measured by turgor, canopy temperature, g_{sw}, and VPD was achieved with phenotyping workflows and selection indices.

To gain additional insight into the genetics of low canopy temperature and high turgor under water deficit stress, we evaluated our selections for recovery of putative QTL derived from LA1141. Four selections based on phenotype

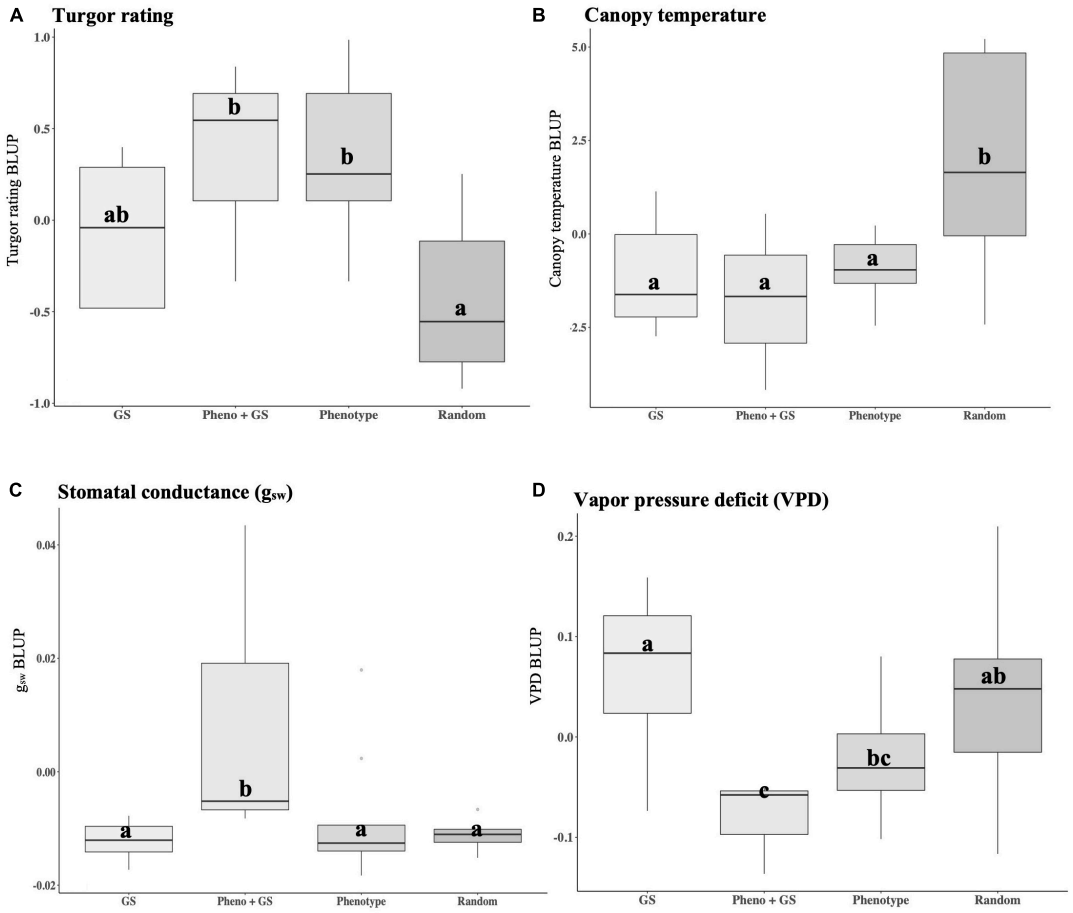


FIGURE 5
Boxplots comparing group performance of groups corresponding to selection strategies under water deficit. Traits were measured as best linear unbiased predictors (BLUPs) and include **(A)** plant turgor ratings, **(B)** canopy temperature, **(C)** stomatal conductance (g_{sw}), and **(D)** vapor pressure deficit (VPD). Selection strategy groups were based on GEBVs (GS, N = 8), LA1141 × OH8245 BC_2S_3 canopy temperature, and turgor BLUPs (Pheno, N = 9), a combination of the two (GS + Pheno, N = 3), and randomly advanced lines (Random, N = 10). Each IBL was replicated three times within the selection strategy. Turgor ratings ranged from 1 to 5 (5 = turgid, 4 = soft to the touch, 3 = beginning to wilt, 2 = wilted with complete loss of turgor, and 1 = dead) as described previously (Waterland et al., 2010). Canopy temperature was estimated using the FLIRONE GEN3 iOS thermal camera (FLIR Systems Wilsonville, OR, United States). Physiological measurements were measured with the LI-600 Porometer/Fluorometer (LI-COR Biosciences, Lincoln, NE, United States) and included stomatal conductance (g_{sw} mol m^{-2} s^{-1}) and vapor pressure deficit (VPD kPa). Values are reported at the time point where they reach their maximums. Different letters indicate statistically significant differences among groups (Tukey's honestly significant difference, $p < 0.05$).

(20% of total selections) possess the LA1141 introgression on chromosome 1b associated with canopy temperature (Table 2). First selection based on GEBV and second selection based on phenotype (15% of total selections) have the introgression from LA1141 on chromosome 6 associated with canopy temperature (Table 2). Three selections based on GEBV, four based on phenotype, and one selected with combined strategies (40% of total selections) have the introgressions from LA1141 on chromosome 5 associated with high turgor (Table 2). Three selections based on phenotype, one selected by combined strategies, and one selected based on genomic selection (20% of total selections) have the introgression from LA1141 on chromosome 7 associated with high turgor (Table 2). The expected frequency of a LA1141 allele in the BC_2S_3 population is 12.5%. The observed frequency of QTLs is, therefore, higher than expected on chromosome 1b (χ^2 = 5.14, p = 0.022), chromosome 5 (χ^2 = 69.14, p = 0.001), and chromosome 7 (χ^2 = 5.14, p = 0.022). However, recovery of the QTL on chromosome 6 approaches the expected frequency (χ^2 = 0.57, p = 0.450).

Evaluation of BC_2S_5 selections for horticultural traits

A field trial was conducted to evaluate the performance of advanced BC_2S_5 selections that were chosen using genomic

TABLE 4 Evaluation of the LA1141 × OH8245 BC_2S_5 selected progenies for horticultural traits compared to the OH8245 recurrent parent.

	Field traits[y]					Color traits[x]		
Genotype[z]	Average fruit weight (g)	Total yield (t ha^{-1})	Marketable yield (t ha^{-1})	Canopy width (cm)	Brix°	% YSD	L*	Hue
OH8245	**68.9 a**	**111.5 a**	**53.5 abcd**	**86 abc**	**5.05 bcde**	**20.30 bc**	**42.81 ab**	**50.55 bc**
SG18-257	55.3 b	93.4 abcd	49.0 abcd	88.5 abc	5.45 abcde	18.12 bc	39.11 abc	50.19 bc
SG18-129	44.5 c	112.7 a	78.2 abc	85 bc	5.50 abcde	11.91 c	36.99 c	48.50 c
SG18-121	36.3 cd	102.7 abc	60.9 abcd	81.5 bc	4.80 bcde	25.66 bc	40.67 abc	52.39 bc
SG18-251	36.3 cd	92.4 abcd	53.2 abcd	67.5 c	4.80 bcde	14.91 c	40.10 abc	50.36 bc
SG18-195	34.5 de	91.1 abcd	61.5 abcd	73.5 bc	4.45 de	25.69 bc	42.27 abc	52.94 bc
SG18-197	33.6 def	69.8 abcd	25.5 bcd	127 ab	5.65 abcde	17.57 bc	40.78 abc	50.56 bc
SG18-223	29.9 defg	90.9 abcd	48.5 abcd	82.5 bc	4.85 bcde	39.39 abc	39.94 abc	54.35 abc
SG18-295	29.9 defg	56.5 bcd	18.8 d	102.5 abc	4.85 bcde	31.97 abc	38.76 abc	54.00 abc
SG18-165	29.0 defg	84.0 abcd	60.7 abcd	58.5 c	4.60 cde	40.22 abc	40.17 abc	54.83 abc
SG18-167	29.0 defg	52.7 cd	15.8 d	98 abc	6.15 abc	61.48 a	43.21 a	62.16 a
SG18-143	28.1 defg	83.1 abcd	43.8 abcd	77.5 bc	5.20 bcde	36.78 abc	39.16 abc	54.49 abc
SG18-177	25.4 efg	108.5 abc	89.6 a	108.5 abc	4.85 bcde	49.28 ab	41.81 abc	57.94 ab
SG18-292	24.5 fgh	102.2 abc	41.6 abcd	141 a	6.45 ab	38.73 abc	40.32 abc	55.74 abc
SG18-145	22.7 ghi	96.4 abcd	40.4 abcd	95 abc	7.00 a	14.71 c	37.24 bc	48.52 c
SG18-182	20.9 ghi	106.0 abc	79.4 ab	75 bc	4.75 cde	38.00 abc	41.26 abc	53.29 bc
SG18-188	15.4 hij	41.4 d	13.6 d	92 abc	5.80 abcde	31.30 abc	41.26 abc	56.17 abc
SG18-248	14.5 ij	43.2 d	19.3 bcd	79.5 bc	6.05 abcd	36.11 abc	39.15 abc	54.14 abc
SG18-255	10.0 j	65.2 abcd	38.9 abcd	71 c	4.20 e	42.07 abc	38.25 abc	54.83 abc
Mean	**31.0**	**84.4**	**46.9**	**88.94**	**5.28**	**31.27**	**40.17**	**53.47**
HSD ($p < 0.05$)[W]	**9.9**	**56.1**	**60.2**	**55.92**	**1.65**	**33.38**	**5.79**	**8.21**

[z]OH8245 is the *S. lycpersicum* elite parent. The remaining genotypes are selected LA1141 × OH8245 progenies that were backcrossed twice to OH8245 and self-pollinated five times (BC_2S_5). Phenotypic selections include SG18-129, SG18-143, SG18-177, SG18-195, SG18-197, SG18-248, SG18-251, and SG18-295. Genomic selections include SG18-121, SG18-145, SG18-167, SG18-182, SG18-188, SG18-255, SG18-257, and SG18-292. Combined selection strategies include SG18-165 and SG18-223.

[y]Average fruit weight (g) of 25 randomly sampled fruit divided by 25 to estimate average fruit weight. Harvested fruit was sorted into marketable, green, and cull groups and each group are weighed separately. Total yield was measured as harvested weight of the three groups.

[x]Color data measured with Tomato Analyzer (Darrigues et al., 2008) on a sub sample of 25 fruit from each plot. Yellow shoulder disorder (%YSD), represents yellow, green-yellow color. L* coordinate indicates darkness (0) to lightness (100) of color. Hue is an angular measurement representing the visible property of color.

[w]The same letters in each treatment indicate non-significant differences among genotypes at $p < 0.05$ based on Tukey's honest significant difference (HSD). Minimum significant difference is reported. Bold values are self-evident.

and phenotypic strategies to establish whether important agronomic and quality traits from the OH8245 recurrent parent were recovered. The fruit size of all BC_2S_5 selections was smaller (p = 8.07e-13) than the OH8245 (Table 4). However, some BC_2S_5 selections have acceptable yield, canopy width, fruit color, and fruit quality (Table 4). Acceptable values of traits were based on ranges observed in processing tomato germplasm (Merk et al., 2012). The acceptable range for canopy width in processing tomatoes is between 75 to 110 cm. In our trial, 65% of selections had an acceptable canopy size, while 25% had canopy size that was too small, and 10% had canopy sizes that were too large (Table 4). Lower values of hue, an angular measurement, represent the visible property of color. Values of L*, a coordinate that indicates the darkness (0) to lightness (100) of color, are associated with increased redness of tomatoes. Ten percent of our selections had improved L* values relative to the recurrent parent, and 100% of selections had an acceptable range of L* (Table 4). Additionally, 10% had improved hue measurements relative to the recurrent parent, and all selections had an acceptable range of hue (Table 4). Low estimates of the physiological color disorder, yellow shoulder (%YSD), are also associated with improved tomato color (Table 4). Forty-five percent of selections exceeded the acceptable cutoff for %YSD (Table 4).

All selections had acceptable Brix° values relative to the OH8245 recurrent parent (Table 4). One selection had a higher Brix° than OH8245. However, this selection ranked 14th out of 20 in fruit size (Table 4). A BC_2S_5 selection chosen based on phenotype, SG18-129, and a BC_2S_5 selection chosen with genomic selection, SG18-145, had improved tomato color relative to OH8245. SG18-129 and SG18-145 also had acceptable canopy sizes (Table 4). Additionally, BC_2S_5 selection SG18-129 had numerically higher yield and numerically higher Brix°

than OH8245. However, those differences in yield were not significant (**Table 4**).

Discussion

Response of LA1141 to water deficit stress

Solanum galapagense accession LA1141 and *S. lycopersium* OH8245 have different physiological responses to water deficit stress. Both morphological and physiological responses can contribute to water deficit tolerance (Koch et al., 2019; Pardo and VanBuren, 2021). Accession LA1141 demonstrates the ability to withstand deficit irrigation for as much as 144 h before wilting is observed. The physiological basis of this tolerance requires further investigation, but a mechanism mediated by stomatal conductance (g_{sw}) is plausible. Similarities of g_{sw} at saturation and genetic variation in advanced selections are promising indicators that it is possible to select for g_{sw} while maintaining plant growth and yield. Previous studies have shown that VPD and g_{sw} are suitable proxies for osmotic adjustment and yield maintenance in plants grown under water deficit stress (Bazzer and Purcell, 2020). Accession LA1141 and advanced selections have lower VPD, higher g_{sw}, and lower canopy temperature after 72 h of water deficit compared to OH8245 and randomly advanced lines, suggesting genetic variation for osmotic adjustments under water deficit stress is present in the population. The putative osmotic adjustment mechanism appears to be independent of leaf anatomy as advanced lines are 87.5% recurrent parent genome and most of the tolerant selections possess the recurrent parent leaf morphological phenotype. For example, the phenotypic selection SG18-197 is shown (**Figure 1**) and leaf morphology is characteristic of the OH8245 recurrent parent.

Tomato plants often exhibit isohydric (saver) and anisohydic (spender) responses to water deficit stress (Sade et al., 2012). A typical isohydric response involves a decline in g_{sw} before any adverse effects of water shortage arise in the canopy (Sade et al., 2012). In contrast, ansiohydric response involves a decline in leaf water potential and stomatal conductance proportional to soil moisture (Sade et al., 2012). Physiological data suggest that LA1141 behaves more like an isohydric plant because of its ability to maintain low canopy temperature, higher g_{sw}, and lower VPD after 72 h of water deficit. In contrast, OH8245 appears to behave like an ansiohydric plant, a response that is associated with higher yields and biomass under moderate stress. This response of OH8245 is consistent with its previously described physiology compared to water deficit tolerant Mediterranean tomato germplasm (Galmés et al., 2011). In situations of water deficit that result in a plant reaching a permanent wilting point, anisohydric behavior may endanger plant survival (Sade et al., 2012). However, plant recovery after deficit irrigation was not evaluated in these studies.

High-throughput thermal image analysis

Image-based estimations of canopy temperature in plants subjected to water deficit can serve as a proxy to plant physiological response. Interest in using canopy temperature as an indicator of plant stress has an extensive history (Jackson et al., 1981; Jones et al., 2002; Gerhards et al., 2016; Perera et al., 2020). Canopy temperature is directly proportional to stomatal conductance in many crops that are subjected to water deficit conditions (Jackson et al., 1988; Ramírez et al., 2016; Fukuda et al., 2018). Canopy temperature can be measured radiometrically using infrared thermometry and by extracting temperature from digital images (Blum et al., 1982; Babar et al., 2006; Gräf et al., 2021). Increased efficiencies achieved using infrared thermometry (Blum et al., 1982) or thermal images (thermography) (Ru et al., 2020; Vieira and Ferrarezi, 2021) enable high throughput data collection to characterize the physiological and genetic consequences of water-deficit stress in large populations or germplasm screens. Genetic variation for canopy temperature has been identified and used for plant improvement in both grain (Lopes et al., 2012) and vegetable crops (Prashar et al., 2013). For example, genetic variation for canopy temperature in runner beans and wheat is linked to maintaining stomatal conductance in water-limited environments (Fischer et al., 1998; Jones, 1999).

Genetic studies suggest tolerance to water deficit stress is a polygenic trait controlled by several small-effect QTLs (Sacco et al., 2013). Hundreds of genes can be activated or repressed in response to water deficit stress (Bray, 2004), making it hard to pinpoint which gene contributes to tolerance. An improved understanding of both the physiological traits and the genetic basis of plant response to water deficit may result in genetic gains in breeding programs by determining the proportion of genetic variation present in populations, estimating the heritability of traits associated with response to water deficit, and modeling environmental effects on traits of interest (Mir et al., 2012).

Generally, when plants experience water deficit stress, their stomata close as a strategy to conserve water. As stomatal conductance declines in response to this closure, leaf temperature can increase rapidly due to loss of evaporative cooling, and an increase in convective heat exchange (Jackson et al., 1988; Lambers et al., 2008). The physical phenomenon of stomatal closure in response to water deficit is inversely proportional to VPD, and the closure also leads to decreased transpiration and carbon assimilation (Lambers et al., 2008). Measuring plant canopy temperature under water deficit may, therefore, provide an efficient phenotyping alternative

to physiological measurements of both stomatal conductance (Zhao et al., 2021) and VPD (Grossiord et al., 2020).

Assessment of canopy temperature provided an efficient measure of response to water deficit stress. Plants under water deficit stress generally have reduced g_{sw}, limited transpiration, and increased canopy temperature. Therefore, canopy temperature was used to proxy these traits. Additionally, canopy temperature was used to phenotype and select tolerant BC_2S_3 families and recovered progenies that maintain higher g_{sw} and lower VPD under deficit irrigation. This suggested that indirect selection was possible with canopy temperature and turgor evaluations. The image analysis workflow was scalable to large populations and efficient for accurate phenotyping in the greenhouse. Water deficit stress evaluations of plants with images partitioned a higher proportion of the total variance into genetics, improved the objectivity of evaluations, and saved time.

Thermal image and IRT estimated canopy temperature are both effective phenotyping strategies, but IRT measures canopy temperature using single-point leaf measurements. These single-point measurements may not capture the entire gradient of temperature across the canopy (**Figure 3**). Temperature gradients across the canopy are likely why point measurements partition more variance into environmental factors and error relative to genetic factors (**Table 1**). Thermal image analysis also detected a greater range of differences in susceptible genotypes than the IRT. Additionally, the image-based analysis identified a canopy temperature's QTL on the distal arm of chromosome 1 in our composite interval mapping study that was detected across seasonal environments but not detected with IRT (**Table 2** and **Figure 4**). Canopy temperature measured by image-based methods may bring greater discrimination and sensitivity for selection and genetic analysis. Importantly, the thermal imaging workflow detected stress before the appearance of wilt symptoms (**Supplementary Figure 4**). Although thermal imaging was an effective method for trait evaluation in our studies, plant turgor ratings had the highest estimates of variance partitioned into genetic effects (**Table 1**). High-throughput phenotyping is promoted to reduce time, costs, and resources to screen populations in breeding programs (Cabrera-Bosquet et al., 2012; Fahlgren et al., 2015; Araus et al., 2018). However, visual evaluation of plant response to water deficit using a visual turgor index (Waterland et al., 2010) appears to be an effective method for evaluating germplasm.

Genetics of water deficit tolerance derived from LA1141

Composite interval mapping helped discover chromosomal regions associated with low canopy temperature and higher turgor under water deficit stress (**Figure 4**). We found evidence for four QTLs, three of which were reproducible across environments and within the combined analysis. As expected, tolerance derived from LA1141 appears to be mediated by many loci and no large-effect QTLs were found. One interpretation of these results is that water deficit tolerance as measured by turgor and canopy temperature is genetically complex, with many QTL falling below the detection threshold. Genomic selection models offered a solution to the genetic complexity of water deficit stress tolerance because of their capacity to handle traits with many small-effect QTL (Crossa et al., 2017). The number of selections that possessed introgression from LA1141 that were associated with putative QTL and were discovered in the mapping study ranged from 15 to 40%, suggesting that we can select for putative QTL and make progress using an introgression strategy, even if the effects of allele substitutions and proportion of phenotypic variance explained are relatively low.

Selection strategies

Phenotypic selection based on high throughput thermal image analysis *via* proximal sensing and genomic selection provides methods to improve response to deficit irrigation in progenies derived from the LA1141 × OH8245 families. Selection strategies based on canopy temperature and turgor also indirectly improved plant physiological response under water deficit measured by g_{sw} and VPD. Overall, prediction accuracy suggested that genomic selection alone may be an effective strategy for evaluating germplasm for tolerance to deficit irrigation as measured by low canopy temperature and high turgor. Additionally, using genomic selection in the future may save time spent phenotyping in additional generations during germplasm screens. Gain under selection was achieved for canopy temperature and turgor ratings. Selection models that incorporate low canopy temperature and visual plant ratings associated with high turgor have also indirectly selected lines with higher g_{sw} and lower VPD under deficit irrigation. Again, this suggests that our phenotyping methodology can be used to proxy plant physiological response and will potentially save time during future germplasm screens. In summary, phenotypic, genomic, and combined selection strategies have identified advanced lines with improved performance when grown under water deficit stress relative to randomly advanced lines. In our evaluations, improved performance was indicated by lower canopy temperature, higher turgor ratings, higher g_{sw}, and lower VPD. These results provided a measure and confirmation of direct and indirect genetic gain toward water-deficit stress tolerance in our parent material, inbred backcross families, and selected progenies.

Water-deficit stress tolerance may also be selected for in-breeding populations using knowledge of marker-trait associations. Quantifying canopy temperature responses in large populations rapidly with objective and repeatable methods may improve QTL discovery. The genetic complexity of water

deficit tolerance and the number of loci that may potentially be involved suggests that alternative strategies that estimate genome-wide effects to predict progeny performance have merit. Genomic selection (GS) (Heffner et al., 2010; Lorenz et al., 2011) changes the focus of analysis from identifying significant associations for QTL to estimating the effect of each marker across the genome. The sum of individual effects provides a genomic estimated breeding value for each family in the population. Our results confirm that GS strategies can be coupled with quantitative phenotyping to develop appropriate selection methods for traits associated with improved physiological status under water deficit. If adequate selection accuracies for complex traits can be achieved, GS has the potential to expedite genetic gain (Heffner et al., 2010; Cabrera-Bosquet et al., 2012). Selection indices developed in this study provide a framework to improve water deficit tolerance using both GS and phenotypic strategies.

Agronomic and quality traits of selections

Acceptable yield, canopy width, and quality parameters were recovered in selections that were chosen based on phenotypic, genomic, and combined strategies. These results demonstrate that we can create selection indices to improve water-deficit tolerance in a recurrent parent background. Additionally, future crossing and evaluation are warranted. The fruit size of selections was small compared to the recurrent parent. Still, some selections yield well and have a canopy size that falls within the acceptable range for commercial processing tomatoes. The failure to recover acceptable fruit size is not surprising in an advanced BC_2 population that used a small-fruited wild accession as the tolerant donor parent. Further crossing and selection will be needed to combine water-deficit stress tolerance and recurrent parent fruit size. However, traits associated with water deficit stress tolerance were successfully introgressed, and at the same time, important agronomic and quality traits associated with commercial processing tomatoes were also recovered.

Conclusion

This work was initiated for the simultaneous introgression and discovery of tolerance to water deficit exhibited by the crop wild relative *S. galapagense* accession LA1141. A thermal image analysis workflow was developed for the population screens, provided an efficient measure of canopy temperature, and was a suitable proxy for physiological traits like g_{sw} and VPD. Analysis of canopy temperature using thermal images at 48 h of water deficit was able to predict turgor ratings, stomatal conductance, and vapor pressure deficit at 72 h water deficit stress. These results suggested that additional efficiencies based on time can be achieved for population evaluations. We were able to identify putative QTL derived from LA1141 associated with low canopy temperature and the ability to maintain high turgor in plants under water deficit. However, no large-effect QTLs were identified. Genomic and phenotypic selection indices offered a feasible strategy to recover tolerance in advanced lines despite the complexity of the trait. Additionally, applying both phenotype-based and genomic selection resulted in the recovery of the putative QTL at a higher-than-expected frequency. Although we successfully selected tomato lines tolerant to water deficit stress, we were unable to recover the fruit size for the direct commercial use of these selections. Finally, the germplasm created in these studies provides a resource for studying traits from LA1141, and we can use the advanced BC_2S_5 selections for future tomato improvement.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Author contributions

SF and DF: conceptualization and experimental design. SF, JC, and JM: phenotyping. SF: high throughput pipeline development and analysis and writing. SF, KM, and DF: data analysis. SF, JC, and DF: population development. JM, KM, and DF: contribution to writing. All authors contributed to the article and approved the submitted version.

Funding

Salaries and research support were provided by state and federal funds appropriated to The Ohio State University, Ohio Agricultural Research and Development Center (OARDC), Hatch project OHO01405, and grant funds from USDA Specialty Crops Research Initiative Award number 2016-51181-25404.

Acknowledgments

We thank the farm crews from the Ohio Agricultural Research and Development Center (OARDC) Wooster for their assistance with the management of the research. We thank Marcela Carvalho Andrade, Regis de Castro Carvalho,

and Wilson Roberto Maluf from The Federal University of Lavras, 37200-000 Lavras, Brazil for assistance with developing the LA1141 IBC population. Additionally, we thank Jessica Cooperstone, Christine Sprunger, and Chieri Kubota, for their constructive comments and insight, which greatly helped to improve the manuscript. Additionally, we also acknowledge Glenn Tattersall, from the department of Biological Sciences at Brock University, for the development of open-source thermal image tools that were used to create our analysis pipeline.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.947538/full#supplementary-material

References

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Babar, M. A., Reynolds, M. P., Van Ginkel, M., Klatt, A. R., Raun, W. R., and Stone, M. L. (2006). Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46, 1046–1057. doi: 10.2135/cropsci2005.0211

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bazzer, S. K., and Purcell, L. C. (2020). Identification of quantitative trait loci associated with canopy temperature in *soybean. Sci. Rep*. 10, 17604. doi: 10.1038/s41598-020-74614-8.

Berry, S. Z., Gould, W. A., and Wiese, K. L. (1991). 'Ohio 8245' processing tomato. *HortScience* 26:1093. doi: 10.21273/HORTSCI.26.8.1093

Blum, A., Mayer, J., and Gozlan, G. (1982). Infrared thermal sensing of plant canopies as a. *Field Crops Res.* 5, 137–146. doi: 10.1016/0378-4290(82)90014-4

Bojacá, C. R., Gil, R., and Cooman, A. (2009). Use of geostatistical and crop growth modelling to assess the variability of greenhouse tomato yield caused by spatial temperature variations. *Comput. Electronics Agric.* 65, 219–227. doi: 10.1016/j.compag.2008.10.001

Bray, E. A. (2004). Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana. J. Exp. Bot.* 55, 2331–2341. doi: 10.1093/jxb/erh270

Broman, K. W., Wu, H., Sen, Ś, and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., and Luis Araus, J. (2012). High-throughput Phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.* 54, 312–320. doi: 10.1111/j.1744-7909.2012.01116.x

Castro, M. (1968). *Tomato Genetics Resource Center. Tomato Genetics Resource Center*. Available online at: https://tgrc.ucdavis.edu/Data/Acc/dataframe.aspx?start=AccSearch.aspx&navstart=nav.html (accessed October, 2021).

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., and de los Campos, G. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011

Dariva, F. D., Copati, M. G. F., Pessoa, H. P., Alves, F. M., Dias, F. D. O., and Picoli, E. (2020). Evaluation of anatomical and physiological traits of Solanum pennellii Cor. associated with plant yield in tomato plants under water-limited conditions. *Sci. Rep.* 10:16052. doi: 10.1038/s41598-020-73004-4

Darrigues, A., Hall, J., Van Der Knaap, E., Francis, D. M., Dujmovic, N., and Gray, S. (2008). Tomato analyzer-color test: a new tool for efficient digital phenotyping. *J. Am. Soc. Hortic. Sci.* 133, 579–586. doi: 10.21273/JASHS.133.4.579

Darwin, S. C., Knapp, S., and Peralta, I. E. (2003). Taxonomy of tomatoes in the galápagos islands: native and introduced species of solarium section lycopersicon (solanaceae). *Syst. Biodiversity* 1, 29–53. doi: 10.1017/S1477200003001026

De Mendiburu, F. (2017). *Package 'Agricolae': Statistical Procedures for Agricultural Research, R Package Version 1.2–4.*

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R Package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Fahlgren, N., Gehan, M. A., and Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opinion Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics (4th)*. Hoboken, NJ: Prentice Hall.

Federer, W. T., and Raghavarao, D. (1975). On augmented designs. *Biometrics* 31, 29–35. doi: 10.2307/2529707

Fenstemaker, S., Sim, L., Cooperstone, J., and Francis, D. (2022). Solanum galapagense-derived purple tomato fruit color is conferred by novel alleles of the Anthocyanin fruit and atroviolacium loci. *Plant Direct* 6:e394 doi: 10.1002/pld3.394

Fischer, R. A., Rees, D., Sayre, K. D., Lu, Z. M., Condon, A. G., and Larque Saavedra, A. (1998). Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. *Crop Sci.* 38, 1467–1475. doi: 10.2135/cropsci1998.0011183X003800060011x

Food and Agriculture Organization [FAO] (2021). *FAOSTAT Statistical Database*. Rome: FAO.

Foolad, M. R. (2007). "Current status of breeding tomatoes for salt and drought tolerance," in *Advances in Molecular Breeding Toward Drought and Salt Tolerant Crops*, Eds M. A. Jenks, P. M. Hasegawa and S. M. Jain. Berlin: Springer

Fukuda, A., Kondo, K., Ikka, T., Takai, T., Tanabata, T., and Yamamoto, T. (2018). A novel QTL associated with rice canopy temperature difference affects stomatal conductance and leaf photosynthesis. *Breed. Sci.* 68, 305–315. doi: 10.1270/jsbbs.17129

Galmés, J., Conesa, M. A., Ochogavía, J. M., Perdomo, J. A., Francis, D. M., Ribas-Carbó, M., et al. (2011). Physiological and morphological adaptations in relation to water use efficiency in mediterranean accessions of *Solanum lycopersicum*. *Plant Cell Environ.* 34, 245–260. doi: 10.1111/j.1365-3040.2010.02239.x

Gerhards, M., Rock, G., Schlerf, M., and Udelhoven, T. (2016). Water stress detection in potato plants using leaf temperature, emissivity, and reflectance. *Int. J. Appl. Earth Obs. Geoinf.* 53, 27–39. doi: 10.1016/j.jag.2016.08.004

Glasbey, C. A. (1993). An analysis of histogram-based thresholding algorithms. *Graph. Models Image Process.* 55, 532–537. doi: 10.1006/cgip.1993.1040

Gräf, M., Immitzer, M., Hietz, P., and Stangl, R. (2021). Water-stressed plants do not cool: leaf surface temperature of living wall plants under drought stress. *Sustainability* 13:3910. doi: 10.3390/su13073910

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., et al. (2020). Plant responses to rising vapor pressure deficit. *New Phytol.* 226, 1550–1566. doi: 10.1111/nph.16485

Haley, C. S., and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315–324. doi: 10.1038/hdy.1992.131

Hamlin, J. A. P., Hibbins, M. S., and Moyle, L. C. (2020). Assessing biological factors affecting postspeciation introgression. *Evol. Lett.* 4, 137–154. doi: 10.1002/evl3.159

Harvey, P. (2020). *ExifTool by Phil Harvey: Read, Write and Edit Meta Information*. Available online at: https://exiftool.org/ (accessed June 2020).

Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662

Horwitz, J. W. (1999). Water at the ice point: a useful quasi-blackbody infrared calibration source. *Appl. Optics* 38:4053. doi: 10.1364/AO.38.004053

Hosmani, P. S., Flores-Gonzalez, M., Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., et al. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *BioRxiv [Preprint]* doi: 10.1101/767764

Inoue, Y., Kimball, B. A., Jackson, R. D., Pinter, P. J., and Reginato, R. J. (1990). Remote estimation of leaf transpiration rate and stomatal resistance based on infrared thermometry. *Agric. For. Meteorol.* 51, 21–33. doi: 10.1016/0168-1923(90)90039-9

Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J. (1981). Canopy temperature as a crop water stress indicator. *Water Resour. Res.* 17, 1133–1138. doi: 10.1029/WR017i004p01133

Jackson, R. D., Kustas, W. P., and Choudhury, B. J. (1988). A reexamination of the crop water stress index. *Irrig. Sci.* 9, 309–317. doi: 10.1007/BF00296705

Jones, H. G. (1999). Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surfaces. *Plant Cell Environ.* 22, 1043–1055. doi: 10.1046/j.1365-3040.1999.00468.x

Jones, H. G., Stoll, M., Santos, T., Sousa, C. D., Chaves, M. M., and Grant, O. M. (2002). Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine. *J. Exp. Bot.* 53, 2249–2260. doi: 10.1093/jxb/erf083

Kabelka, E., Franchino, B., and Francis, D. M. (2002). Two loci from Lycopersicon hirsutum LA407 confer resistance to strains of *Clavibacter michiganensis* subsp. michiganensis. *Phytopathology* 92, 504–510. doi: 10.1094/PHYTO.2002.92.5.504

Kitaya, Y., Niu, G., Kozai, T., and Ohashi, M. (1998). Photosynthetic photon flux, photoperiod, and CO2 concentration affect growth and morphology of lettuce plug transplants. *HortScience* 33, 988–991. doi: 10.21273/HORTSCI.33.6.988

Koch, G., Rolland, G., Dauzat, M., Bédiée, A., Baldazzi, V., Bertin, N., et al. (2019). Leaf production and expansion: a generalized response to drought stresses from cells to whole leaf biomass—a case study in the tomato compound leaf. *Plants* 8:409. doi: 10.3390/plants8100409

Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugenics* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Kustas, W. P., Alfieri, J. G., Anderson, M. C., Colaizzi, P. D., Prueger, J. H., Evett, S. R., et al. (2012). Evaluating the two-source energy balance model using local thermal and surface flux observations in a strongly advective irrigated agricultural area. *Adv. Water Resour.* 50, 120–133. doi: 10.1016/j.advwatres.2012.07.005

Lambers, H., Chapin, F. S., and Pons, T. L. (2008). *Plant Physiological Ecology*. Berlin: Springer. doi: 10.1007/978-0-387-78341-3

Liabeuf, D., Sim, S. C., and Francis, D. M. (2018). Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. *Phytopathology* 108, 392–401. doi: 10.1094/PHYTO-12-16-0431-R

Lopes, M. S., Reynolds, M. P., Jalal-Kamali, M. R., Moussa, M., Feltaous, Y., Tahir, I. S. A., et al. (2012). The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. *Field Crops Res.* 128, 129–136. doi: 10.1016/j.fcr.2011.12.017

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). "Genomic selection in plant breeding. knowledge and prospects," in *Advances in Agronomy*, Sparks, D. L (Ed.) 1st Edn, (Amsterdam: Elsevier). doi: 10.1016/B978-0-12-385531-2.00002-5

Merk, H. L., Yarnes, S. C., Van Deynze, A., Tong, N., Menda, N., Mueller, L. A., et al. (2012). Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato Germplasm. *J. Am. Soc. Hortic. Sci.* 137, 427–437. doi: 10.21273/JASHS.137.6.427

Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., and Varshney, R. K. (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theoretical Appl. Genet.* 125, 625–645. doi: 10.1007/s00122-012-1904-9

Pailles, Y., Awlia, M., Julkowska, M., Passone, L., Zemmouri, K., Negrão, S., et al. (2020). Diverse traits contribute to salinity tolerance of wild tomato seedlings from the galapagos islands. *Plant Physiol.* 182, 534–546. doi: 10.1104/pp.19.00700

Pardo, J., and VanBuren, R. (2021). Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *Plant Cell* 33, 3391–3401. doi: 10.1093/plcell/koab205

Perera, R. S., Cullen, B. R., and Eckard, R. J. (2020). Using leaf temperature to improve simulation of heat and drought stresses in a biophysical model. *Plants* 9:8. doi: 10.3390/plants9010008

Philips, B., Egel, D., Maynard, E., Ingwell, L., and Meyers, S. (2021). *Midwest Vegetable Production Guide*. Available online at: https://mdc.itap.purdue.edu/item.asp?Item_Number=ID-56 (accessed October 2021).

Prashar, A., Yildiz, J., McNicol, J. W., Bryan, G. J., and Jones, H. G. (2013). Infrared Thermography for high throughput field phenotyping in *Solanum tuberosum*. *PLoS One* 8:e65816. doi: 10.1371/journal.pone.0065816

Prodanov, D., and Verstreken, K. (2012). "Automated segmentation and morphometry of cell and tissue structures. selected algorithms in image]," in *Molecular Imaging*, Schaller, B. (Ed.) (London: IntechOpen). doi: 10.5772/36729

R Core Team (2020). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramírez, D. A., Yactayo, W., Rens, L. R., Rolando, J. L., Palacios, S., De Mendiburu, F., et al. (2016). Defining biological thresholds associated to plant water status for monitoring water restriction effects: Stomatal conductance and photosynthesis recovery as key indicators in potato. *Agric. Water Manag.* 177, 369–378. doi: 10.1016/j.agwat.2016.08.028

Rick, C. M. (1956). Genetic and systematic studies on accessions of lycospersicon from the galapagos Islands. *Am. J. Bot.* 43, 687–696. doi: 10.1002/j.1537-2197.1956.tb14433.x

Rick, C. M. (1973). "Potential genetic resources in tomato species: clues from observations in native habitats," in *Genes, Enzymes, and Populations*, ed. S. Adrian (Berlin: Springer), 255–269. doi: 10.1007/978-1-4684-2880-3_17

Rick, C. M. (1978). "Potential improvement of tomatoes by controlled introgression of genes from wild species," in *Proceedings of the Conference on Broadening the Genetic Base of Crops*, (Nertherlands).

Robbins, M. D., Darrigues, A., Sim, S. C., Masud, M. A. T., and Francis, D. M. (2009). Characterization of hypersensitive resistance to bacterial spot race T3 (xanthomonas perforans) from tomato accession pi 128216. *Phytopathology* 99, 1037–1044. doi: 10.1094/PHYTO-99-9-1037

Ru, C., Hu, X., Wang, W., Ran, H., Song, T., and Guo, Y. (2020). Evaluation of the crop water stress index as an indicator for the diagnosis of grapevine water deficiency in greenhouses. *Horticulturae* 6:86. doi: 10.3390/horticulturae6040086

Rush, D. W., and Epstein, E. (1976). Genotypic responses to salinity. *Plant Physiol.* 57, 162–166. doi: 10.1104/pp.57.2.162

Sacco, A., Greco, B., Matteo, A., Di Stefano, R., and De Barone, A. (2013). Evaluation of tomato genetic resources for response to water deficit. *Am. J. Plant Sci.* 4, 131–145. doi: 10.4236/ajps.2013.412A3016

Sade, N., Gebremedhin, A., and Moshelion, M. (2012). Risk-taking plants: anisohydric behavior as a stress-resistance trait. *Plant Signal. Behav.* 7, 767–770. doi: 10.4161/psb.20505

Saeed, A., and Fatima, N. (2021). "Wild germplasm: shaping future tomato breeding," in *Wild Germplasm for Genetic Improvement in Crop Plants*, eds M. T.

Azhar and S. H. Wani (Amsterdam: Elsevier). doi: 10.1016/B978-0-12-822137-2.00011-4

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., and Pietzsch, T. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Shannon, L. M., Yandell, B. S., and Broman, K. (2013). *Users Guide for New BCsFt Tools for R / qtl*. Available online at: http://cran.fhcrc.org/web/packages/qtl/vignettes/bcsft.pdf (accessed December 2020).

Sharma, H., Shukla, M. K., Bosland, P. W., and Steiner, R. (2017). Soil moisture sensor calibration, actual evapotranspiration, and crop coefficients for drip irrigated greenhouse chile peppers. *Agric. Water Manag.* 179, 81–91. doi: 10.1016/j.agwat.2016.07.001

Smith, D. T., Potgieter, A. B., and Chapman, S. C. (2021). Scaling up high-throughput phenotyping for abiotic stress selection in the field. *Theoretical Appl. Genet.* 134, 1845–1866. doi: 10.1007/s00122-021-03864-5

Snijdgers, T. A. B., and Bosker, R. J. (2012). *Multilevel Analysis : an Introduction to Basic and Advanced Multilevel Modeling (2nd)*. Thousand Oaks, CA: Sage Publications.

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., and Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci.* 61, 839–852. doi: 10.1002/csc2.20377

Tattersall, G. J. (2019). *ThermImageJ: Thermal Image Functions and Macros for ImageJ*. Available online at: https://doi.org/10.5281/zenodo.2652896 (accessed October 2021).

Tattersall, G. J. (2021). *Thermimage: Thermal Image Analysis*. Available online at: https://doi.org/10.5281/zenodo.1069704 (accessed October 2021).

Vieira, G. H. S., and Ferrarezi, R. S. (2021). Use of thermal imaging to assess water status in citrus plants in greenhouses. *Horticulturae* 7:249. doi: 10.3390/horticulturae7080249

Waterland, N. L., Campbell, C. A., Finer, J. J., and Jones, M. L. (2010). Abscisic acid application enhances drought stress tolerance in bedding plants. *HortScience* 45, 409–413. doi: 10.21273/HORTSCI.45.3.409

World Bank Group (2021). *World Bank Climate Change Knowledge Portal*. Available online at: https://climateknowledgeportal.worldbank.org/ (accessed October 2021).

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhao, L., Wang, L., Li, J., Bai, G., Shi, Y., and Ge, Y. (2021). "Toward accurate estimating of crop leaf stomatal conductance combining thermal IR imaging, weather variables, and machine learning," in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping VI*, eds J. A. Thomasson and A. F. Torres-Rua (Bellingham: SPIE). doi: 10.1117/12.2587577

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fpls.2022.951903

OPEN ACCESS

EDITED BY
Mohd. Kamran Khan,
Seluck University, Turkey

REVIEWED BY
Xiaoming Zheng,
Institute of Animal Sciences (CAAS), China
Anjana Rustagi,
University of Delhi,
India
Tofazzal Islam,
Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

*CORRESPONDENCE
Zhiping Song
songzp@fudan.edu.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 24 May 2022
ACCEPTED 15 August 2022
PUBLISHED 06 September 2022

CITATION
Chen H, Dong S, He Z, Chen Y, Tian D, Liu Y, Wang Y, Zhang W, Li L, Yang J and Song Z (2022) Effects of land use change on population survival of three wild rice species in China since 2001. *Front. Plant Sci.* 13:951903. doi: 10.3389/fpls.2022.951903

Effects of land use change on population survival of three wild rice species in China since 2001

Hao Chen[1], Shanshan Dong[2], Zhizhou He[1], Yuhong Chen[1], Defeng Tian[1], Yan Liu[2], Yuguo Wang[1], Wenju Zhang[1], Linfeng Li[1], Ji Yang[1] and Zhiping Song[1]*

[1]The Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, Institute of Botany, Fudan University, Shanghai, China, [2]Nanjing Institute of Environmental Sciences of the Ministry of Ecology and Environment, Nanjing, China

Land use change stemming from human activities, particularly cropland expansion, heavily threatens the survival of crop wild relatives that usually occur nearby or scatter in farming systems. Understanding the impacts of land use change on wild populations is critical in forming the conservation decision-making of wild relatives. Based on the investigations on the population survival of three wild rice species (*Oryza rufipogon*, *O. officinalis*, and *O. granulata*) in China over the past 40years (1978–2019), the effect of land use change during the past 20years (2001–2019) on the natural populations of the three species was examined using the land use type data of satellite-based Earth observations (data from GlobCover). From 1978 to 2019, the number of populations (distribution sites) of the three wild rice species had decreased by 65–87%, mainly because of the habitat destruction or disappearance caused by human-induced land use change. The three wild rice species display different habitat preferences, resulting in specific land use types surrounding their populations. In the recent 20years, although the surrounding community composition of the wild rice population has been relatively stable, the surrounding vegetation cover area of the survived populations was significantly more extensive than that of the extinct ones ($p<0.05$). These findings suggest that habitat vegetation plays a "biological barrier" role in the survival of wild populations through resisting or mitigating the disturbing impact of land use change on wild populations. This study provides not only direct guidelines for the conservation of wild rice but also new insights into the mechanisms underlying the influence of land use change on wild populations.

KEYWORDS

conservation, habitat, land use, community, wild rice, wild population

Introduction

Crop wild relatives (CWRs) are the most important genetic resources for germplasm innovation and crop improvement (Xiao et al., 1996; Hajjar and Hodgkin, 2007; Feuillet et al., 2008). They play an important role in raising the yield and quality of grains and other economic crops (Vitousek et al., 1997). However, CWRs have been seriously threatened by

human activities around the world. On the one hand, people need to increase land use to supply food for the rapidly growing population; this increased land use involves land exploitation and agricultural expansion that inevitably leads to the destruction or even loss of natural habitats, threatening the survival of wild populations. On the other hand, CWRs have habitat requirements similar to those of domesticated crops. Thus, they usually scatter in agricultural systems, and their habitats tend to be directly exploited or disturbed by agricultural activities, such as the use of agrochemicals (pesticides and fertilizers), grazing, and the introduction of nonnative competitors, resulting in population decline or even extinction of CWRs (Newbold et al., 2015; Pimm et al., 2018; Nicholson et al., 2019). Therefore, land use change is the main threat to CWRs survival (Davies et al., 2006). Evaluating the effect of land use change on wild populations is critical for CWRs conservation.

As an important genetic resources, CWRs have attracted the attention of conservationists and plant breeders. However, little is known about how land use change impacts the survival of their wild populations. Several studies that focused on the population genetics of CWRs suggested that the genetic population decline is associated with habitat fragmentation due to human activities (Ceballos et al., 2017). Consequently, *in situ* conservation strategies have been established to maintain population genetic variations, thereby protecting the evolutionary potential of the wild population. The survival of wild populations mainly depends on the habitat quality and the stability of the natural community surrounding wild populations (Blowes et al., 2019; Newbold et al., 2020). Community stability is highly related to landscape structure and heterogeneity. However, land cover changed stemming from land use change directly alter landscape properties, thereby affecting the CWRs population through cascading effects (da Silva and Hernández, 2014). For example, the change of land cover causes landscape composition changes. Thus, the change or disappearance of a biological community and the loss of the natural or seminatural habitat of wild populations eventually lead to population decline or even extinction of CWRs (Niedrist et al., 2009; Malavasi et al., 2018). The community and its shift, which is usually accompanied by land use change, have a great effect on wild populations (Newbold et al., 2015). Thus, we need a deep understanding of how land use change affects the communities surrounding CWRs populations.

The practices of biodiversity conservation, including those for CWRs, have been continuously moving forward. For example, the European Union (EU) has implemented sustainable enhancement, organic agriculture, and agricultural environment programs to reduce the negative impact of agricultural activities on biodiversity (Andersson et al., 2012; Batáry et al., 2015). Meanwhile, China has carried out the most extensive programs for ecosystem services worldwide, such as the Natural Forest Conservation Program (NFCP) and Grain-to-Green Program (GTGP; also known as the Farm-to-Forest Program), as well as important agricultural biological resource protection actions, to restore natural vegetation and protect biodiversity (Xu et al., 2006; Liu et al., 2008). Since the implementation of these programs, many national *in situ* and *ex situ* protection zones of CWRs have been established. Thus far, China has built 205 *in situ* conservation zones for 39 CWRs, including 65 *in situ* conservation zones for wild rice species. In particular, China established two national wild rice germplasm resource nurseries, where the plants of more than 17,000 wild rice germplasm resources are growing (Xu et al., 2020). Although these conservation practices greatly benefit the survival of wild populations of CWRs, the attention to the effects of these practices experiencing the impacts of land use change on the community surrounding the CWRs population and the comparison of *in situ* conserved populations with not conserved ones remain lacking.

Wild rice species are the wild relatives of cultivated rice (*Oryza sativa*; Zheng et al., 2019). They are also important germplasm resources for rice breeding because they contain abundant insect-resistant, stress-tolerant, high-yield, or/and male sterility genes. For instance, the utilization of male sterility genes from the common wild rice (*O. rufipogon*) generated the world-famous hybrid rice (*O. sativa*) (Liu et al., 2014). Three wild rice species are distributed in China: *O. rufipogon*, *O. officinalis*, and *O. granulata* (Dai et al., 2004). The three wild rice species mainly occur on the edges of farmlands and forest zones. Their habitats are constantly disturbed by land use changes mainly caused by agricultural activities, resulting in a serious decline in their population. The Chinese government has listed the three wild rice species as national class II protected species and launched two national surveys for wild rice resources (1978–1982 and 2001–2004; Xu et al., 2020). Consequently, dozens of natural populations of wild rice were conserved *in situ*. Moreover, the results of the two national surveys show that Chinese wild rice populations are still declining, even though great effort has been paid to wild rice population conservation and the surviving wild rice populations generally hold relatively rich genetic diversity (Yang et al., 2013). This population decline may be due to habitat destruction or loss caused by land use change. In addition, the three wild rice species display remarkable niche differentiation (Zhang and Yang, 2003). The three plants have specific habitat preferences and occur in different ecological systems. Moreover, their populations may face different land use changes due to agricultural activities. Therefore, the three wild rice species provide us an ideal system to compare the population survival statuses of CWRs with different typic or intensive disturbances and obtain general insights into the effects of land use change on CWRs.

The present study used the three wild rice species as a model, aims to examine the effects of land use change on the survival of wild populations. We performed another national wild rice population survey in 2019 based on the historical survey records (1978–2004). We investigated the influence of land use change on wild rice populations by using the data of population survival status combined with the Landsat Earth observation information of land use change in the recent 20 years (2001–2019), covering the range of the three wild rice species. This study aims to answer the following questions: (1) What is the habitat preference of wild rice species? (2) What kind of land use change have the three wild

rice species experienced? (3) Do land use change and population decline in wild rice species have significant correlations? The obtained data provide not only direct guidelines for the conservation of wild rice but also new insights into the mechanisms underlying the influence of land use change on CWRs.

Materials and methods

Study area

The basic unit of the research object is the population in the ecological concept. We recorded the three wild rice species (*O. rufipogon*, *O. officinalis*, and *O. granulata*) with a geographical space interval of less than 300 m as the same population. The first national wild rice survey (1978–1982) comprehensively recorded the geographical distribution range of all wild rice populations in China (excluding Taiwan Province), showing that *O. rufipogon* is distributed in Jiangxi, Hunan, Yunnan, Guangxi, Guangdong, Fujian and Hainan Province; *O. officinalis* in Yunnan, Guangxi, Guangdong and Hainan Province; and *O. granulata* in Yunnan and Hainan Province (Supplementary Figure S1; The Cooperative Team of Wild Rice Resources Survey and Exploration of China, 1984; Xu et al., 2020). Subsequent studies are all based on these data of geographical distribution range and population status. The second national survey (2001–2004; Xu et al., 2020) covered the geographical range entirely as the first one. In the present study, the sampled populations were randomly selected from those survived in the second national survey and also cover the whole distribution range of the three wild rice species in China. GPS data were recorded at the center of the geographical range of each population.

Population information

The data of the first national survey were from the survey reports (The Cooperative Team of Wild Rice Resources Survey and Exploration of China, 1984). The population data of the second national survey were obtained from the literature reports, survey reports, books, and other records of the Provincial Academy of Agricultural Sciences (Pang et al., 2000; Pang and Chen, 2002; Zhang and Yang, 2003; Gai et al., 2005; Xu et al., 2020). Some of the original recorded "populations" were combined and recorded as the same population when the distance between individual "populations" is less than 300 m, then we obtained the basic data of population survival status of the three wild rice species in 2001. To further monitor the population survival status of the three wild rice species in China recently, we conducted a filed survey across the entire range of the three wild rice species from 2016 to 2019. A total of 195 populations survived were investigated in our survey, including 118 *O. rufipogon*, 35 *O. officinalis*, and 42 *O. granulata* (Supplementary Tables S1–S3), in which wild rice plants still exist was recorded as present population; otherwise, it was noted as lost one. We used the data of population status in 2001 and 2019 to do land use change analysis.

Land use type

Land use type is usually characterized as land cover type, such as farmland, water body, natural vegetation, and so on, that all are easily recorded by satellite (Zalles et al., 2021). Therefore, we can describe the land use change of wild rice habitats during a period based on the land cover data observed by satellite in different years. The satellite-observed land cover data were derived from GlobCover version 2.32009.[1] These data are available as a raster with 23 land cover classes at a resolution of 300 m. The data include natural land cover and artificial land use types, which are uniformly recorded as land use in this study. We used land use types in grids of GPS points that were repositioned in the center of the population to obtain the land use types of the 195 populations. To obtain the data of the land use types of communities around the habitats of wild rice populations, we extracted all land use types of each population within a radius of 5 km centered on the GPS point and calculated the area of each land use type in the circle to represent the composition and area of the community surrounding the population. The information of four populations of *O. rufipogon* could not be extracted. Finally, we extracted the data of 191 populations and their surrounding communities. Among them, for *O. rufipogon,* there are 64 population in presence, and 50 population lost; for *O. officinalis*, there are 23 population in presence, and 12 population lost; for *O. granulata*, there are 31 population in presence, and 11 population lost (Supplementary Tables S1–S3). In this study, 12 land types were obtained. They comprise three types of TC (tree cover): tree cover, broadleaved, evergreen; tree cover, broadleaved, deciduous; tree cover, needle leaved, evergreen. The following are the other nine land types: cropland, rainfed (RC); cropland, irrigated, or postflooding (IC); mosaic cropland (MCN, > 50% natural vegetation, < 50%); mosaic natural vegetation (MNC, > 50% cropland, < 50%); mosaic tree and shrub (MTH, > 50% herbaceous cover, < 50%); shrubland (SL); tree cover, flooded, fresh, or brackish water (TCF); urban areas (UA); water bodies (WB). The extracted data also include NA, indicating no data. The data remained stable from 2000 to 2019, so we leave them out. The extraction of land use types of each population in satellite observation data was performed in R 4.0.1 using a self-written R script.

1 http://due.esrin.esa.int/page_globcover.php

Data analysis

The survey data in 1978–2019 that we used were divided into two periods: 1978–2000 and 2001–2019. We also calculated the loss proportion of wild rice populations in each period. The loss proportion of the population from 1978 to 2000 was obtained by comparing the second national survey data (2001–2004) with the first national survey data (1978–1982). The loss proportion of the population in 2001–2019 was obtained by comparing the populations survey data of 2016–2019 with the data of the second national survey (2001–2004).

We considered 4 years as the statistical unit to reduce the statistical error of the community area around the population. This statistical unit was divided into five periods: 2000–2003, 2004–2007, 2008–2011, 2012–2015, and 2016–2019. The average value of the 4-year data in each period represents this time. The period from 2000 to 2003 was used as the baseline, and the area proportion of each type in the other four periods was compared with the baseline to obtain the change proportion of the current period relative to the baseline period. In this study, in order to compare the difference of land use area between the present and lost populations, the significance test between the present population and the lost population (the area of each type within the 5 km range in 2001) was conducted in R 4.0.1 using Wilcoxon rank sum test.

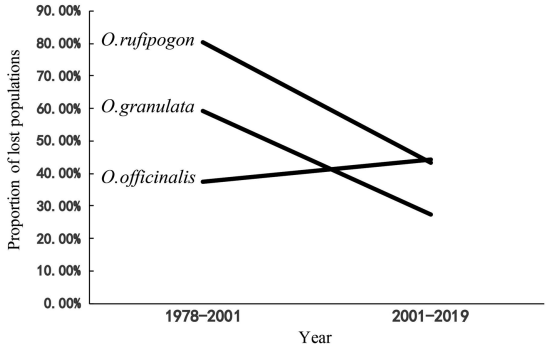


FIGURE 1
The proportion of lost populations in two period (1978–2001,2001–2019). The data of surveys in 1978–1982 as baseline, the proportion of loss population in 1978–2001 was compared with data of two national surveys 1978–1982 and 2001–2004; The proportion of lost populations in 2001–2019 was compared with data of national surveys (2001–2004) and our data in 2019.

Results

Survival dynamics of wild rice populations

After 20 years since the first national survey [1978–2000 (S1)], the survival rates of wild populations of *O. rufipogon*, *O. officinalis*, and *O. granulata* decreased to 20, 63, and 40%, respectively, when the land change was first clearly observed by satellites in 2001. Population decline for the three wild rice species has continued in the recent 20 years [from 2001 to 2019 (S2)], with population loss rates of 43.2, 44.3, and 26.2% (Figure 1). These findings indicate that the three wild rice populations have experienced a high-speed extinction trend in the past 40 years. The comparison of the loss proportions of the populations in the two periods indicated that the population loss proportion of *O. rufipogon* and *O. granulata* decreased considerably in the S2 period. This decrease was approximately 1/2 of that in the S1 period (80.3–43.2% and 59.3–26.2%). However, the population loss proportion of the *O. officinalis* population has accelerated (37.5–44.3%; Figure 1).

Land use types of habitats of the three wild rice populations in 2001

The land use types of the three wild rice population habitats are basically the same in 2001 (Figure 2). They include the land use types of RC, IC, MCN, MNC, TC, and MTH, but their proportions vary greatly. The proportion of farmland (RC and IC) and mosaic farmland (MCN and MNC) in the habitat of *O. rufipogon* populations is 94.7% (Figure 2A; Supplementary Table S4). The proportions of either type in the habitats of *O. officinalis* (31.4%) and *O. granulata* (47.6%) are low (Figures 2B,C; Supplementary Table S4). The proportion of TC in the habitats of *O. officinalis* and *O. granulata* is the highest, accounting for 51.4 and 31.0%, respectively; however, it is less than 1% in the habitat of *O. rufipogon*. This finding shows that obvious habitat preference differences exist among the three wild rice species.

Further analyses of the habitat types of the lost populations from 2001 to 2019 showed that the populations whose habitats were in the farmland system (RC, IC, MCN, and MNC) reached a high loss proportion: 96% of the lost population habitats of *O. rufipogon* (Figure 2A) and 66.7% of the lost population habitats of *O. officinalis* (Figure 2B) and *O. granulata* (Figure 2C) are in the farmland system. This finding suggests that the populations in agricultural systems are under a great threat from agricultural activities.

Composition and dynamics of communities around the habitat of the three wild rice populations

The land use types of communities around the habitats of the three wild rice populations include 10 types: RC, IC, MCN, MNC, TC, MTH, SL, TCF, UA, and WB. However, the area proportion varies greatly. The communities around the habitats of the three wild rice populations mainly comprise farmland systems (RC, IC, MCN, and MNC; Figure 3). The results show that the present population of *O. rufipogon* has significantly low RC (Wilcoxon rank sum test, $p = 0.0047$) area proportion and significantly high

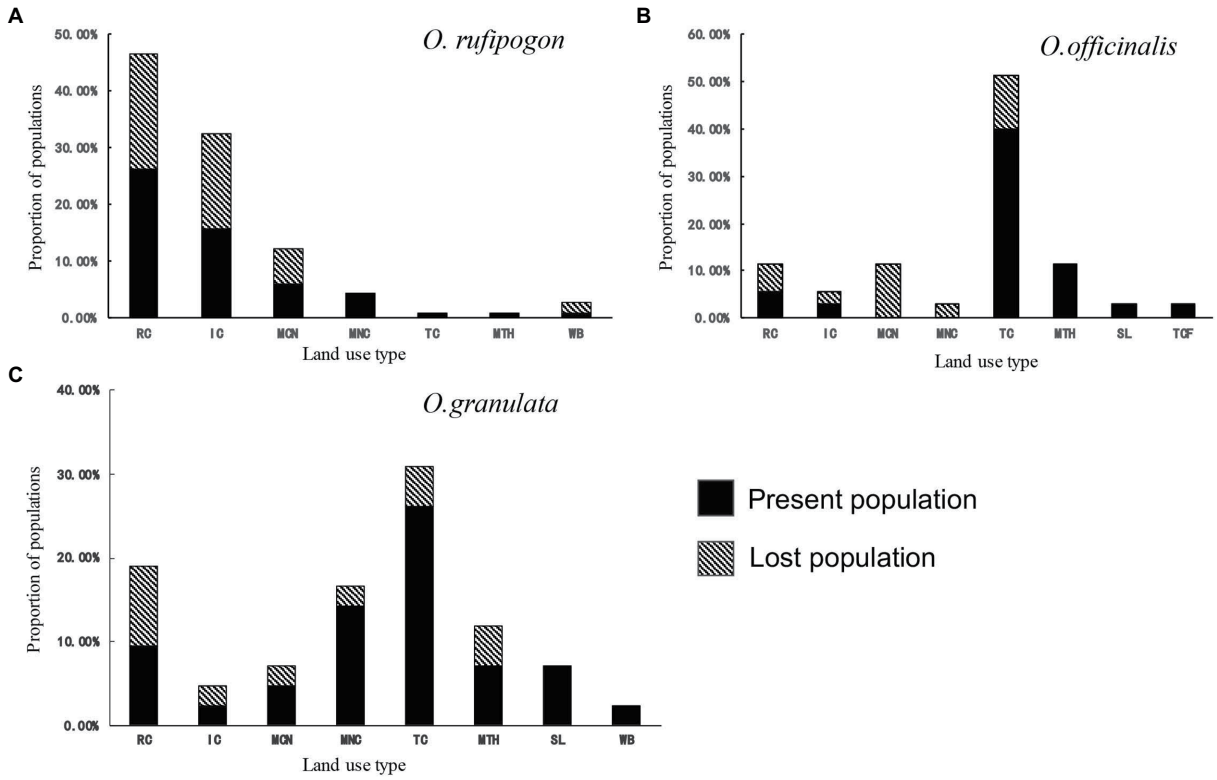


FIGURE 2
Population proportion per land use type of (A) *O. rufipogon* (B) *O. officinalis* (C) *O. granulata* during 2001~2019. Present population is the population survived during 2001~2019; Lost population presented in 2001 but disappeared in 2019. RC, rainfed cropland; IC, irrigated or postflooding cropland; MCN, mosaic cropland, >50% / natural vegetation, <50%; MNC, mosaic natural vegetation, >50% / cropland, <50%; TC, tree cover; MTH, mosaic tree and shrub, >50% / herbaceous cover, <50%; SL, shrubland; TCF, tree cover, flooded, fresh, or brackish water; WB, water bodies.

MCN ($p = 0.0338$) and MNC ($p = 0.0036$) area proportion (Figure 3A; Table 1). The present population of *O. officinalis* has a significantly low area proportion of MCN ($p = 0.0106$) and SL ($p = 0.0207$; Figure 3B; Table 1). The present population of *O. granulata* has a significantly low area proportion of RC ($p = 0.0402$), and significantly high area proportion of MCN ($p = 0.0149$) and SL ($p = 0.0014$; Figure 3C; Table 1). Compared with the lost populations, the communities around the habitat have a significantly high proportion of vegetation area, indicating the potential protective role of vegetation in the survival of the three wild rice populations.

The dynamics of the communities around the populations from 2001 to 2019 were analyzed further (Supplementary Figures S2–S4). The area proportion of farmland systems (RC and IC) around almost all populations of the three wild rice species showed a gradually decreasing trend in four periods (2004–2007, 2008–2011, 2012–2015, and 2016–2019), whereas the area proportion of the natural communities (TC, MTH, SL and TCF) showed an increasing trend. This observation shows that the area of farmland around the population has been decreasing and the area of vegetation has been increasing in the recent 20 years. The proportion of mosaic farmland (MCN and MNC; Supplementary Figures S3C,D) in some populations for *O. officinalis* has shown a considerable decline in lost populations relative to the present populations, whereas the proportions of TC and SL for *O. officinalis* have shown a considerable increase (Supplementary Figures S3E,G) in the recent period (2016–2019). This finding suggests that the transformation of mosaic farmland (MNC and MCN) to natural vegetation (TC and SL) has occurred in the recent period.

Discussion

Monitoring habitat change and evaluating its effect on population survival help in understanding the impact of land use change on biodiversity (Fahrig et al., 2011; Kennedy et al., 2013; Zabel et al., 2019; Zalles et al., 2021). The present study evaluated the influence of land use change on the population survival of the three wild rice species in China for the first time through integrative analysis of the data on population survival in the recent 40 years and the remote sensing data of land use change across the range of wild rice distribution in the recent 20 years. The populations of the three wild rice species are affected by different land use change depending on their specific habitat preference. *O. rufipogon*, which constantly exists in farmland systems, suffers

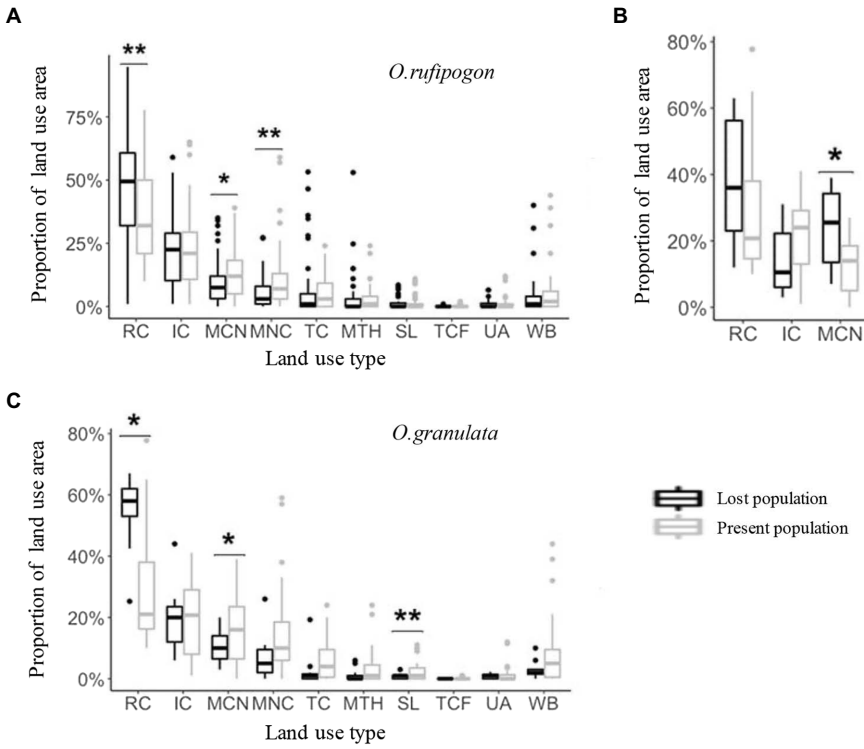


FIGURE 3
The land use area proportion in (A) *O. rufipogon* (B) *O. officinalis* (C) *O. granulata* within a radius of 5 km during 2001~2019. Present population is the population survived during 2001~2019; Lost population presented in 2001 but disappeared in 2019. *p<0.05, and **p<0.01 (Wilcoxon rank sum test). RC, rainfed cropland; IC, irrigated or postflooding cropland; MCN, mosaic cropland, >50% / natural vegetation, <50%; MNC, mosaic natural vegetation, >50% / cropland, <50%; TC, tree cover; MTH, mosaic tree and shrub, >50% / herbaceous cover, <50%; SL, shrubland; TCF, tree cover, flooded, fresh, or brackish water; UA, urban areas; WB, water bodies.

TABLE 1 *p*-Values generated from Wilcoxon Rank Sum test testing the difference of land use types area around populations.

Species		RC	IC	MCN	MNC	TC	MTH	SL	TCF
O. rufipogon	W	1121.5	1,627	1994	2,132	1725.5	1783.5	1,654	1725.5
	P	0.0047	0.9706	0.0338	0.0036	0.5432	0.3301	0.8339	0.1235
O. officinalis	W	90.5	183.5	64	134.5	143	166.5	75	162
	P	0.1022	0.1171	0.0106	0.9167	0.8725	0.3054	0.0207	0.1385
O. granulata	W	78.5	164.5	67.5	132	143	166.5	50	168
	P	0.0402	0.3654	0.0149	0.8479	0.8736	0.3055	0.0014	0.0909

RC, rainfed cropland; IC, irrigated or postflooding cropland; MCN, mosaic cropland, > 50% /natural vegetation, < 50%; MNC, mosaic natural vegetation, > 50% /cropland, < 50%; TC, tree cover; MTH, mosaic tree and shrub, > 50% /herbaceous cover, < 50%; SL, shrubland; TCF, tree cover, flooded, fresh, or brackish water. W, Wilcoxon Rank sum test statistics; *p*, *p*-values.

the most consequential impacts of land use change. The natural vegetation surrounding wild populations can considerably mitigate the destructive effect of land use change on the habitat of wild rice species, allowing the refuge of wild populations.

The combined analyses of historical records of population distribution and our field investigation data show that the three wild rice species are distributed inconsistently, resulting in populations isolated from one another. Although their habitat land cover types are basically the same, the proportion of each cover type considerably varies (Dai et al., 2004; Yang et al., 2013). *O. rufipogon* is mainly distributed in the shallow water or wet regions of inland swamps, such as ponds, swamps, streams, and riversides in hills and plains. The land cover types of its habitat are mainly farmland (RC and IC) and mosaic farmland (MCN and MNC; Figures 2, 3), resulting in its populations being often nearby rice fields (Song et al., 2003), and suggesting the similar habitat requirements between *O. rufipogon* and cultivated rice. *O. officinalis* is a hydrophilic plant and commonly distributed in mountainous streams, puddles, and other humid areas. The land cover type of its population is mainly forest (Figure 2). *O. granulata* is a xerophyte and understory plant. It is often found in the gaps of shrub and arbor forests on the hillside. The land cover type of its habitat is also mainly forest (Figure 2). These results

demonstrate that the land cover types of population habitats are different between the three rice species and the habitat preferences of the three rice species are specific. This pattern is related to niche differentiation during speciation.

The analysis of land use change reveals that the three wild rice species have experienced changes in habitats (land cover types) with different extents. The populations of *O. rufipogon* are usually scattered in farmland systems with little space for reclamation (wasteland or barren land). Thus, *O. rufipogon* shows a relatively stable state (Supplementary Figure S5A). Conversely, the populations of *O. officinalis* and *O. granulata*, commonly occurring in seminatural or natural systems nearby farmland systems, face a serious threat of habitat change (Supplementary Figures S5B,C). This pattern hints that land use change is much intensive at the edge of the farmland system or forest edge (Bryant et al., 2020). Further analysis of the community (land cover area) around the population shows that the populations of the three wild rice species have a similar trend of community shift, with a slightly decreased area of farmland and increased area of natural vegetation (Supplementary Figures S2–S4). This pattern may be attributed to the implementation of NFCP and GTGP (Liu et al., 2008; Deng et al., 2014).

We further analyzed the relationship between land use change and the population decline of wild rice species. The population numbers of the three wild rice species have considerably varied since the first survey in the 1980s (Figure 1). This population decline is obviously caused by habitat development or/and destruction (Dai et al., 2004). We found that the survival rates vary between the wild rice populations occurring in different habitats (different land cover types and type proportions) and almost all the lost populations of the three wild rice species come from the farmland system, and the lost rate of populations in other habitats is very low (Figure 2). These results can be explained by the fact that the habitat of wild rice is directly occupied because of land use change or/and habitat destruction due to agricultural activities, such as the use of chemical fertilizers, pesticides, and herbicides (Tilman et al., 2001). It supports the view that the change in agricultural practices strongly affects the biodiversity in the farmland ecosystem (Sala et al., 2000; Titeux et al., 2016; Ellis, 2019). The populations of *O. officinalis* and *O. granulata* are influenced by other types of land use (Figure 2), such as abandonment of farmland or orchard that leads to intensive interspecific competition due to weed spring up (Matus et al., 2003; Deák et al., 2011).

The status of wild rice populations (including population stability and survival) in different habitats is correlated with the area of their surrounding vegetation. Our field investigation showed that dozens of *in situ* conserved populations, including the three wild rice species, grow well; however, the community type and the proportion of different land cover types of these conserved populations seem to be similar to those of populations that disappeared (Supplementary Figure S6). The reason is that human disturbance is effectively avoided by physical barriers, such as fences or wire entanglements, or mitigated by the buffer zone around the conserved populations (Gonçalves-Souza et al., 2021). For the un-conserved populations, those with a small area of natural vegetation exhibit high extinction rates (Figure 3). The lower the area of vegetation/farmland surrounding wild rice populations is, the lower the landscape heterogeneity is, and the weaker the ecological flexibility of wild populations is (Niedrist et al., 2009; Malavasi et al., 2018; Newbold, 2018). This finding indicates that natural vegetation can effectively mitigate the impacts of human activities on wild populations. Playing the role of a "biological barrier" (Santoro et al., 2012), natural vegetation can buffer the pressure of herbivore grazing and trampling and the spillover effect of pesticide spraying and chemical fertilizer on wild rice. This "biological barrier" effect coincides with the good population survival of *O. rufipogon* and *O. granulata*, following the increase of natural vegetation since the implementation of natural ecological programs (Liu et al., 2008). Thus, the present study agrees with the view that, like physical barriers, natural vegetation plays the role of a "biological barrier" to alleviate the impacts of land use change on wild populations (Reyers et al., 2010).

With the Landsat Earth observation information of land use change and the field population investigation, the present study reveals that the habitats of three wild rice populations are constantly affected by land use change due to human activities, particularly agricultural activities. Our findings suggest that remote sensing technology can be used effectively to monitor population status, land use changes, and community around wild population in real time and quantify the role of vegetation type and area in buffering the impact of human activities. This study also aims to consider the composition of communities around wild populations as a critical component of the assessment system in monitoring the population dynamics of wild rice species. In addition, our results call for the management and limitation of human agricultural activities, such as the use of herbicides, the expansion of cultivated lands, and grazing, to reduce the impact of agriculture on community integrity and improve the resilience and sustainability of the ecosystem of sympatric wild relative species. Generally, the present study highlights the biological barrier role of natural vegetation surrounding wild populations in suffering from the destructive effect of land use change on wild rice species. To further understanding the effects of land use change on CWRs, we need more studies monitoring the dynamics of population and its surrounding vegetations by using remote sensing technology.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ZS, YW, WZ, LL, and JY conceived the study. HC, SD, ZH, YC, DT, and YL collected the data. HC performed the statistical analyses and wrote the first draft of manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was supported by the National Natural Science Foundation of China (31570383) and the Biodiversity Survey, Observation and Assessment Program of Ministry of Ecology and Environment of China (2019-5-12).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.951903/full#supplementary-material

References

Andersson, G. K., Rundlöf, M., and Smith, H. G. (2012). Organic farming improves pollination success in strawberries. *PLoS One* 7:e31599. doi: 10.1371/journal.pone.0031599

Batáry, P., Dicks, L. V., Kleijn, D., and Sutherland, W. J. (2015). The role of Agri-environment schemes in conservation and environmental management. *Conserv. Biol.* 29, 1006–1016. doi: 10.1111/cobi.12536

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., et al. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science* 366, 339–345. doi: 10.1126/science.aaw1620

Bryant, B. P., Kelsey, T. R., Vogl, A. L., Wolny, S. A., MacEwan, D., Selmants, P. C., et al. (2020). Shaping land use change and ecosystem restoration in a water-stressed agricultural landscape to achieve multiple benefits. *Front. Sustain. Food Syst.* 4:138. doi: 10.3389/fsufs.2020.00138

Ceballos, G., Ehrlich, P. R., and Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. U. S. A.* 114, E6089–E6096. doi: 10.1073/pnas.1704949114

da Silva, P. G., and Hernández, M. I. (2014). Local and regional effects on community structure of dung beetles in a mainland-island scenario. *PLoS One* 9:e111883. doi: 10.1371/journal.pone.0111883

Dai, L. Y., Wu, L. H., Wang, L., Yang, Q. W., Tang, C. F., and Yu, T. J. (2004). Analysis on the current status of wild rice resource distributed in Yunnan province based on the investigation. *Chinese. J. Rice. Sci.* 24, 107–128. doi: 10.1300/J064v24n01_09

Davies, R. G., Orme, C. D., Olson, V., Thomas, G. H., Ross, S. G., Ding, T. S., et al. (2006). Human impacts and the global distribution of extinction risk. *Proc. Biol. Sci.* 273, 2127–2133. doi: 10.1098/rspb.2006.3551

Deák, B., Valkó, O., Kelemen, A., Török, P., Miglécz, T., Ölvedi, T., et al. (2011). Litter and graminoid biomass accumulation suppresses weedy forbs in grass-land restoration. *Plant. Biosyst. Int. J. Aspects. Plant. Biol.* 145, 730–737. doi: 10.1080/11263504.2011.601336

Deng, L., Liu, G. B., and Shangguan, Z. P. (2014). Land-use conversion and changing soil carbon stocks in China's 'Grain-for-Green' program: a synthesis. *Glob. Chang. Biol.* 20, 3544–3556. doi: 10.1111/gcb.12508

Ellis, E. C. (2019). Evolution: biodiversity in the Anthropocene. *Curr. Biol.* 29, R831–R833. doi: 10.1016/j.cub.2019.07.073

Fahrig, L., Baudry, J., Brotons, L., Burel, F. G., Crist, T. O., Fuller, R. J., et al. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. *Ecol. Lett.* 14, 101–112. doi: 10.1111/j.1461-0248.2010.01559.x

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends Genet.* 24, 24–32. doi: 10.1016/j.tig.2007.11.001

Gai, H. M.， Chen, C. B., Shen, F. F., Zhang, W. X., Ren, M., Wang, Y. W., et al. (2005). Genetic diversity and conservation strategy of *Oryza rufipogon* along the Haojiang River in Guangxi Zhuang autonomous region. *J. Plant. Genetic. Resour.* 6,156–162. doi: 10.3969/j.issn.1672-1810.2005.02.007

Gonçalves-Souza, D., Vilela, B., Phalan, B., and Dobrovolski, R. (2021). The role of protected areas in maintaining natural vegetation in Brazil. *Sci. Adv.* 7, eabh2932. doi: 10.1126/sciadv.abh2932

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1–13. doi: 10.1007/s10681-007-9363-0

Kennedy, C. M., Lonsdorf, E., Neel, M. C., Williams, N. M., Ricketts, T. H., Winfree, R., et al. (2013). A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecol. Lett.* 16, 584–599. doi: 10.1111/ele.12082

Liu, J., Li, S., Ouyang, Z., Tam, C., and Chen, X. (2008). Ecological and socioeconomic effects of China's policies for ecosystem services. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9477–9482. doi: 10.1073/pnas.0706436105

Liu, W., Liu, J., Triplett, L., Leach, J. E., and Wang, G. L. (2014). Novel insights into rice innate immunity against bacterial and fungal pathogens. *Annu. Rev. Phytopathol.* 52, 213–241. doi: 10.1146/annurev-phyto-102313-045926

Malavasi, M., Bartak, V., Carranza, M. L., Simova, P., and Acosta, A. T. R. (2018). Landscape pattern and plant biodiversity in Mediterranean coastal dune ecosystems: do habitat loss and fragmentation really matter? *J. Biogeogr.* 45, 1367–1377. doi: 10.1111/jbi.13215

Matus, G., Tóthmérész, B., and Papp, M. (2003). Restoration prospects of abandoned species-rich sandy grassland in Hungary. *Appl. Veg. Sci.* 6, 169–178. doi: 10.1111/j.1654-109x.2003.tb00577.x

Newbold, T. (2018). Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. *Proc. Biol. Sci.* 285, 20180792. doi: 10.1098/rspb.2018.0792

Newbold, T., Hudson, L. N., Hill, S. L., Contu, S., Lysenko, I., Senior, R. A., et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature* 520, 45–50. doi: 10.1038/nature14324

Newbold, T., Oppenheimer, P., Etard, A., and Williams, J. J. (2020). Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nat. Ecol. Evol.* 4, 1630–1638. doi: 10.1038/s41559-020-01303-0

Nicholson, E., Fulton, E. A., Brooks, T. M., Blanchard, R., Leadley, P., Metzger, J. P., et al. (2019). Scenarios and models to support global conservation targets. *Trends Ecol. Evol.* 34, 57–68. doi: 10.1016/j.tree.2018.10.006

Niedrist, G., Tasser, E., Lüth, C., Via, J. D., and Tappeiner, U. (2009). Plant diversity declines with recent land use changes in European Alps. *Plant Ecol.* 202, 195–210. doi: 10.1007/s11258-008-9487-x

Pang, H. H., and Chen, C. B. (2002). *The Wild rice Resources in China*. Nanning: Guangxi Science and Technology Press.

Pang, H. H., Yang, Q. W., and Zhao, J. (2000). Exploration, collection, evaluation and conservation of wild rice resources in China. *J. Plant. Genetic. Resour.* 1, 52–56.

Pimm, S. L., Jenkins, C. N., and Li, B. V. (2018). How to protect half of earth to ensure it protects sufficient biodiversity. *Sci. Adv.* 4, eaat2616. doi: 10.1126/sciadv.aat2616

Reyers, B., Roux, D. J., Cowling, R. M., Ginsburg, A. E., Nel, J. L., and O'Farrell, P. (2010). Conservation planning as a transdisciplinary process. *Conserv. Biol.* 24, 957–965. doi: 10.1111/j.1523-1739.2010.01497.x

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., et al. (2000). Global biodiversity scenarios for the year 2100. *Science* 287, 1770–1774. doi: 10.1126/science.287.5459.1770

Santoro, R., Jucker, T., Prisco, I., Carboni, M., Battisti, C., and Acosta, A. T. (2012). Effects of trampling limitation on coastal dune plant communities. *Environ. Manag.* 49, 534–542. doi: 10.1007/s00267-012-9809-6

Song, Z. P., Xu, X., Wang, B., Chen, J. K., and Lu, B. R. (2003). Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. *Theor. Appl. Genet.* 107, 1492–1499. doi: 10.1007/s00122-003-1380-3

The Cooperative Team of Wild Rice Resources Survey and Exploration of China (1984). A general survey and exploration of wild rice germplasm in China. *Sci. Agric. Sinica.* 17, 27–34.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., et al. (2001). Forecasting agriculturally driven global environmental change. *Science* 292, 281–284. doi: 10.1126/science.1057544

Titeux, N., Henle, K., Mihoub, J. B., Regos, A., Geijzendorffer, I. R., Cramer, W., et al. (2016). Biodiversity scenarios neglect future land-use changes. *Glob. Chang. Biol.* 22, 2505–2515. doi: 10.1111/gcb.13272

Vitousek, P. M., Mooney, H. A., Lubchenco, J., and Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science* 277, 494–499. doi: 10.1126/science.277.5325.494

Xiao, J., Grandillo, S., Ahn, S., McCouch, S. R., and Tanksley, S. D. (1996). Genes from wild rice improve yield. *Nature* 384, 223–224. doi: 10.1038/384223a0

Xu, Z. J., Wang, J. L., Zheng, X. Y., Fan, Z. L., Tang, C. F., Wang, X. H., et al. (2020). Collection and conservation of wild Rice resources in China. *J. Plant. Genetic. Resour.* 21, 1337–1343. doi: 10.13430/j.cnki.jpgr.20200420003

Xu, J. T., Yin, R. S., Li, Z., and Liu, C. (2006). China's ecological rehabilitation. *Ecol. Econ.* 57, 595–607. doi: 10.1016/j.ecolecon.2005.05.008

Yang, Q. W.，Qin, W. B.，Zhang, W. X.，Qiao, W. H.，Yu, S. N., and Guo, Q. (2013). In-situ conservation practices and future development of wild relatives of crops in China. *J. Plant. Genetic. Resour.* 14, 1–7. doi: 10.3969/j.issn.1672-1810.2013.01.001

Zabel, F., Delzeit, R., Schneider, J. M., Seppelt, R., Mauser, W., and Václavík, T. (2019). Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nat. Commum.* 10, 2844. doi: 10.1038/s41467-019-10775-z

Zalles, V., Hansen, M. C., Potapov, P. V., Parker, D., Stehman, S. V., Pickens, A. H., et al. (2021). Rapid expansion of human impact on natural land in South America since 1985. *Sci. Adv.* 7, eabg1620. doi: 10.1126/sciadv.abg1620

Zhang, W. X., and Yang, Q. W. (2003). Collecting, valuation and conservation of wild rice resources in China. *J. Plant. Genetic. Resour.* 2003, 369–373. doi: 10.3969/j.issn.1672-1810.2003.04.020

Zheng, X. M., Chen, B. X., Song, Y., Li, F., Wang, J. R., Qiao, W. H., et al. (2019). In-situ conservation of wild relatives of crops. *J. Plant. Genetic. Resour.* 20, 1103–1110. doi: 10.13430/j.cnki.jpgr.20190513002

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpls.2022.1010799

OPEN ACCESS

EDITED BY
Sait Gezgin,
Selcuk University, Turkey

REVIEWED BY
Satinder Kaur,
Punjab Agricultural University, India
Rajesh Singhal,
Indian Grassland and Fodder Research Institute (ICAR), India

*CORRESPONDENCE
Javier García-Algarra
javier.algarra@u-tad.com

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 03 August 2022
ACCEPTED 14 October 2022
PUBLISHED 03 November 2022

CITATION
Civantos-Gómez I, Rubio Teso ML, Galeano J, Rubiales D, Iriondo JM and García-Algarra J (2022) Climate change conditions the selection of rust-resistant candidate wild lentil populations for *in situ* conservation. *Front. Plant Sci.* 13:1010799. doi: 10.3389/fpls.2022.1010799

Climate change conditions the selection of rust-resistant candidate wild lentil populations for *in situ* conservation

Iciar Civantos-Gómez[1,2], María Luisa Rubio Teso[3], Javier Galeano[1], Diego Rubiales[4], José María Iriondo[3] and Javier García-Algarra[5]*

[1]Complex System Group, Universidad Politécnica de Madrid, Madrid, Spain, [2]Faculty of Economics and Business Administration, Universidad Pontificia Comillas, Madrid, Spain, [3]ECOEVO Research Group, Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Madrid, Spain, [4]Instituto de Agricultura Sostenible (CSIC) Avenida Menéndez Pidal s/n Campus Alameda del Obispo, Córdoba, Spain, [5]DRACO Research Group, Centro Universitario de Tecnología y Arte Digital, Las Rozas, Spain

Crop Wild Relatives (CWR) are a valuable source of genetic diversity that can be transferred to commercial crops, so their conservation will become a priority in the face of climate change. Bizarrely, *in situ* conserved CWR populations and the traits one might wish to preserve in them are themselves vulnerable to climate change. In this study, we used a quantitative machine learning predictive approach to project the resistance of CWR populations of lentils to a common disease, lentil rust, caused by fungus *Uromyces viciae-fabae*. Resistance is measured through a proxy quantitative value, DSr (Disease Severity relative), quite complex and expensive to get. Therefore, machine learning is a convenient tool to predict this magnitude using a well-curated georeferenced calibration set. Previous works have provided a binary outcome (resistant *vs.* non-resistant), but that approach is not fine enough to answer three practical questions: which variables are key to predict rust resistance, which CWR populations are resistant to rust under current environmental conditions, and which of them are likely to keep this trait under different climate change scenarios. We first predict rust resistance in present time for crop wild relatives that grow up inside protected areas. Then, we use the same models under future climate IPCC (Intergovernmental Panel on Climate Change) scenarios to predict future DSr values. Populations that are rust-resistant by now and under future conditions are optimal candidates for further evaluation and *in situ* conservation of this valuable trait. We have found that rust-resistance variation as a result of climate change is not uniform across the geographic scope of the study (the Mediterranean basin), and that candidate populations share some interesting common environmental conditions.

KEYWORDS

crop wild relatives, climate change, machine learning, rust resistance, lentils, *in situ* conservation, predictive characterization

1 Introduction

In the coming decades, food security will be seriously compromised by the lack of adaptive resilience to climate change of cultivars currently used in crops (Smith et al., 2017; Wiebe et al., 2019; Anderson et al., 2020). This lack of adaptive resilience is caused by the low genetic diversity that is inherent to most modern cultivars (Rauf et al., 2010; Van de Wouw et al., 2010; Rufo et al., 2019). Indeed, it has been estimated that major crops are likely to experience sensible yield reductions in the coming decades. In their review of 2015 for Eastern Africa, Adhikari et al. depict a gloomy scenario for main staples by the end of this century. They predict a 72% drop for wheat, around 40% for other cereals, and 10% for potatoes (Adhikari et al., 2015). A recent study predicts a drop that ranges from 3% to 12% by 2050 and from 11% to 25% by 2090 for rice and soybeans (Wing et al., 2021). This will be the result of losses caused by the arrival of new pests and pathogens, the intensification of the effects of those active right now, and potential mismatches to the new climate regimes, including increasing temperature and drought and higher incidence of extreme events (e.g., hail, strong winds, floods, etc). Anyway, those statistical projections could hide the fact that climate change may result beneficial for some crops and regions (Ray et al., 2019).

Crop wild relatives (CWR) are one of the most important sources of genetic diversity to transfer key adaptations to crops, a relevant fact in the context of climate change (Heywood et al., 2007; Maxted, 2008; Zhang et al., 2017). For example, they have been effectively used in plant breeding in crops such as sunflower (*Helianthus annuus* L.) (Seiler et al., 2017), narrow–leafed lupin (*Lupinus angustifolius* L.) (Mousavi-Derazmahalleh et al., 2018), durum wheat (*Triticum turgidum* subsp. *durum* (Desf.) Husn.) (El Haddad et al., 2021) or pea (*Pisum sativum* L.) (Rubiales et al., 2020).

Interestingly, the genetic diversity of crop wild relatives is, at the same time, being severely eroded mainly due to habitat modification and destruction by human activities (Iriondo et al., 2008; Khoury et al., 2022). A recent survey of 600 species of crop wild relatives in the United States, estimated that more than one-half are endangered and 7% in a critical condition (Khoury et al., 2020). As a result, a global effort is being made to promote the establishment of genetic reserves or the *in situ* conservation of crop wild relatives (Maxted et al., 2008; van Treuren et al., 2017; Labokas et al., 2018).

In the process of deciding which populations of a given crop wild relative should be selected for *in situ* conservation, there are a number of considerations that must be taken into account. On one hand, genetic reserves should be representative of the range of genetic diversity present in the species (Dempewolf et al., 2017). The best way of characterizing genetic diversity is through sequencing as it provides complete information about the genome and its cost is drastically decreasing. Nevertheless, the characterization of hundreds or thousands of populations that a given crop wild relative might have is a highly time-consuming task that would require huge human and economic resources as well. Ecogeographic information arises as a useful tool that provides a proxy to estimate among-population genetic diversity and helps in the selection of representative populations (Vincent et al., 2019). Plant breeders are often interested in identifying and conserving candidates that display a targeted phenotypic trait (e.g., resistance to a pathogen). Once again, ecogeographic information associated with the CWR populations is critical to identifying these candidate populations given the impossibility of conducting evaluation experiments with plant material from so many sites. In this sense, predictive characterization techniques based on FIGS (Focused Identification Germplasm Strategy) have been applied to identify wheat resistance to stem rust, caused by the fungus *Puccinia graminis* (Endresen et al., 2012; Bari et al., 2014), and barley resistance to leaf rust, caused by *Puccinia hordei* (Amouzoune et al., 2022). Different methodological approaches predicted phenotypic traits through a calibration approach, using as a starting point a set of training data where the targeted trait of a set of populations is known and the corresponding ecogeographic information may be retrieved from public repositories (Endresen, 2010; Sánchez et al., 2019). In the search for resistance to pathogens, several researchers have followed qualitative approaches in which the material is previously evaluated as resistant or non-resistant (Bari et al., 2012; Rubio Teso et al., 2022). However, resistance to pathogens could be evaluated using numerical variables as well (Arojju et al., 2018; Ren et al., 2021). Predictive characterization of quantitative resistance traits such as DSr (Disease Severity relative) to lentil rust (*Uromyces vicia-fabae* (Pers.) Schröt), is likely to benefit from machine-learning models with quantitative dependent variables (Rubiales et al., 2013).

The second aspect of great relevance when identifying the most appropriate CWR populations relates to the adequacy of the site for the long term *in situ* conservation. The land use of the site has to be compatible with the long-term viability and persistence of the candidate population (Hunter, 2012; Hunter et al., 2012). Those within protected areas are less vulnerable to human disturbance and are, therefore, preferred in this context (Maxted et al., 2012). In any case, the *in situ* conservation of CWR in genetic reserves may also be feasible in other instances (e.g., farms), whenever there is a long-term commitment by the landowners (Maxted and Kell, 2009).

Ex situ conservation is the best and most adequate approach for conserving and utilizing plant genetic reserves. One of the main benefits of complementing it with *in situ* conservatio of CWR is that *in situ* conserved wild populations are constantly evolving as a result of changing biotic and abiotic environmental factors (Fu, 2017). Adapting to such changes favours genotypes that maximize their fitness under current and potentially future environmental conditions (Meilleur and Hodgkin, 2004). On the

contrary, the germplasm conserved ex situ in genebanks allows rapid access to genetic variation but represent a static genetic diversity of the population in the moment of sampling (Castañeda-Álvarez et al., 2016) that, in any case can be useful in recovering populations. In the identification of the most appropriate CWR populations for *in situ* conservation, one must take into account not only the environmental conditions currently present in the target population but also those expected to occur in the future as a result of climate change, and whether those future conditions are compatible with the preservation of the population or the targeted traits. The vulnerability of protected areas to the effects of climate change is starting to be assessed in terms of global biodiversity or emblematic species but has not been studied in the context of CWR *in situ* conservation (Hannah, 2010; Triviño et al., 2018).

In this study we used a predictive characterization approach, based on machine learning, to quantitatively project the rust resistance of crop wild relative populations of lentils (*Lens culinaris* subsp. *culinaris*) in the Mediterranean basin. Rust is a severe foliar disease in lentils (Rubiales et al., 2011). Breeding with CWRs to increase rust resistance of cultivars is a convenient method for this disease control in legume crops (Barilli et al., 2009; Rubiales et al., 2011; Negussie and Pretorius, 2012; Sillero et al., 2017).

Climate change is expected to unevenly affect agriculture in different parts of the world (Howden et al., 2007). There is large variation in climatic conditions, soils, land use, infrastructure, and political and economic conditions across the European continent (Olesen and Bindi, 2002; Cramer et al., 2018; Fellmann et al., 2018). These differences are expected to influence the responsiveness of CWR populations to climate change. Here we apply climate change projections and Shared Socio-Economic Pathways (SSPs) to predictive characterization in order to compare the potential changes in rust-resistance of lentil CWR populations in Europe and Turkey (Dufresne et al., 2013; Voldoire et al., 2013; O'Neill et al., 2014; Wu et al., 2014). We, then, aimed to identify a set of rust-resistant candidate populations that could be designated for *in situ* conservation in genetic reserves, searching for those that occur in protected areas and selecting those with the lowest vulnerability to changes in the environment that might result in the loss of this trait. The results of the calibration method of the predictive characterization techniques should answer the following questions: (i) which variables are the most important to predict rust resistance? (ii) which CWR populations are likely to show strong resistance to rust under present time environmental conditions? and (iii) which of them are likely to keep this trait under forthcoming climate change scenarios? We expect selected populations to have evolved to develop rust resistance over time and consider them to be a valuable asset under present and future conditions.

2 Data description

Lentil was cultivated for the first time in the Fertile Crescent around 5000 BP (Zohary et al., 2012). It probably spread out through the Mediterranean basin, the Indian subcontinent, and the Horn of Africa at a relatively fast pace driven by its high yield (Liber et al., 2021). Thus, it didn't face the selective pressure of its wild relatives, the subspecies *orientalis* and *odemensis* and the three species *Lens nigricans* (M. Bieb.) Godr., *Lens ervoides* (Brign.) Grande and *Lens lammotei*. Czefr. We considered in our study all *Lens* taxa naturally occurring in Europe and Turkey. These are *L. ervoides*, *L. nigricans* and *L. lammotei*, as well as *L. culinaris* subsp. *orientalis* (Boiss.) Ponert. and *L. culinaris* subsp. *odemensis*. (Ladiz.). *Lens* taxa distribution data were extracted from a database of crop wild relative populations in Europe and Turkey generated for the Farmer's Pride project (www.farmerspride.eu) (Rubio Teso et al., 2020). Due to the imbalance in the number of samples ofthe original database (443 populations of *L. nigricans* found in 12 countries, 145 of *L. ervoides* in 9 countries, 29 of *L. lamottei* and 9 of *L. culinaris subsp. orientalis*), we decided to build a unique model and exclude the species as an input variable. We made this choice to avoid overfitted models, being aware that there is a loss of input information, but taking into account that the four taxa belong to the same genus.

Raw data downloaded were further cleaned and filtered as indicated in (Rubio Teso et al., 2022). The calibration dataset holding rust evaluation data has 351 samples of five *Lens* taxa, both wild and cultivated (*L. culinaris*, *L. culinaris* subsp. *culinaris*, *L. culinaris* subsp. *orientalis*, *L. ervoides* and *L. nigricans*). Each sample is georeferenced and its DSr value is the mean of four years' field trials. Each field trial followed a complete block design with 3 replications, artificially inoculating the samples and including frequent rows of susceptible checks to act as spreaders to ensure a high and uniform disease pressure. Disease Severity on mature plants in the field is assessed as a visual estimation of the leaf area covered by rust pustules, which is influenced by environmental factors. This value (DS) is standardized each year by expressing each DS value as a percentage of the highest one in each location that is set at 100% (DSr) (Sillero et al., 2017).

Each record of the filtered lentil CWR distribution database and of the calibration dataset was associated to the values of 65 bioclimatic, 35 edaphic and 18 geophysic variables at 2.5 arc-min resolution, corresponding to the sites of the populations. This information was obtained from the ecogeographic database of CAPFITOGEN3 (Parra-Quijano et al. 2020). Latitude and longitude were also added to the variable selection procedure. *SelecVar* function of CAPFITOGEN3 was used to estimate variable importance according to the random forest classification (RFC) and detected redundant variables through

bivariate correlation analysis (Garcia et al., 2017). The first 15 variables of each bioclimatic, edaphic and geophysic component with the highest Mean Decrease Accuracy (MDA) values (Cutler et al., 2007; Rubio Teso et al., 2022) were checked for colinearity. Pairs of variables with Pearson correlation coefficient > |0.50| and p-value< 0.05 in the same ecogeographical component were identified and the variable with the lowest MDA removed. Hence, in the bioclimatic component, only annual mean temperature (°C) and annual precipitation (mm) were kept. In the edaphic component, four non-correlated variables were selected: bulk density (fine earth) of topsoil, topsoil available soil water capacity until wilting point, topsoil total exchangeable bases and topsoil sand fraction. Finally, in the geophysic component, three non-correlated variables were selected: annual solar radiation ($kJ/m^2perday$), December solar radiation, and longitude. Further information and details about the generation and characteristics of this environmental database can be found in (Rubio Teso et al., 2022).

To identify the populations of lentil, crop wild relatives of Europe and Turkey that occur within protected areas, we considered the protected areas registered at the World Database of Protected Areas (WDPA) for this territory and those included in the Natura 2000 network. The file with the polygons of the WDPA was downloaded in April 2021 from the website 'Explore the World's Protected Areas' (protectedplanet.net). The polygons of the Natura 2000 network were downloaded from the European Environment Agency website (https://www.eea.europa.eu/data-and-maps/data/natura-11/natura-2000-spatial-data/natura-2000-shapefile-1, last accessed 2021/07/20). The shapefile polygons from N2000 and WDPA obtained were merged into a single shapefile that contained all available protected areas in Europe and Turkey, using the function 'join vector layers' in QGIS v.3.18.2-Zürich (QGIS Org, 2021). All protected areas in the resulting shapefile were considered for the selection of candidate populations.

The dataset of lentil CWRs comprises 583 populations of four *Lens* taxa; 236 out of them grow inside a protected area (Supplementary Table S1).

3 Climate change models

Considering that climate change is expected to influence the evolutionary dynamics of CWR populations, one of the aims of this study was also to apply predictive characterization under future climate conditions. To do so, we incorporated climate change projections and Shared Socio-Economic Pathways (SSPs) to assess whether the environmental conditions that are likely to promote at present the presence of rust resistance in wild populations still remain in the forthcoming future. According to this, we combined the ecogeographic variables with the projected temperature and precipitation to quantitatively project the rust resistance of crop wild relative populations in future climate scenarios. A potential role of the results extracted here is their applicability in the selection of candidate populations (Stockwell and Peterson, 2002; Guisan and Thuiller, 2005; Araújo and Guisan, 2006).

The future climate Geographical Information System (GIS) layers were downloaded from the Worldclim database (http://www.worldclim.org/) at 2.5 arc-min resolution (around 5x5 km). From the available periods, we selected the 2021-2040 as the future climate scenario. We chose three global circulation models (GCMs) for climate change projections, produced by the Coupled Model Intercomparison Project Phase 6 (CMIP6) (O'Neill et al., 2016). For every GCM, we analysed three different shared socioeconomic pathways(SSPs), the most "pessimistic" or "conservative" scenario, a "balanced" scenario and an "optimistic" scenario, so we can cover the range of expectations to do a sensitivity analysis. Only the two bioclimatic variables that had been previously selected (Annual mean temperature and Annual mean precipitation) were considered.

4 Methods

4.1 Predictive characterization for evaluation accessions

In this study, we rely on the evaluated accessions that constituted the calibration dataset to carry out a quantitative predictive characterization using a machine learning regression approach. This dataset includes, on the one hand, the continuous range of the DSr as the dependent variable and the selected ecogeographical variables, at present time.

Rubio Teso et al. addressed the issue of predicting resistance to rust by means of the calibration method and a qualitative strategy with the same dataset (Rubio Teso et al., 2022). This method consisted of the binarization of DSr numerical values into qualitative values (resistant, susceptible), prior to prediction. According to this, accessions with the lowest DSr values were classified as rust-resistant, i.e., those located in the first decile of the distribution. The binarized levels of expression (0 = susceptible; 1 = resistant) were used as the dependent variable. Ecogeographical variables were the inputs to predict the binarized DSr resistance through nine classification algorithms. Then, the best predictor model was projected on the non-evaluated populations.

We followed a different path in this research, predicting the numerical value of DSr. We built and evaluated three different families of predictive models with the present time environmental values and DSr values for evaluated accessions: Ridge Regression (Hoerl, 1962, Random Forest (Breiman et al., 1984) and XGBoost (eXtreme Gradient Bosting) (Friedman, 2001; Chen and Guestrin, 2016).

4.1.1 Regression models

DSr quantitative prediction is a regression problem with tabular data. We have implemented in Python the three regression models to tackle the task to identify populations potentially resistant to lentil rust.

Initially, we used the calibration dataset at present to train the models. From this dataset, we carry out a cleaning process in which we discard duplicate values and samples with incomplete variables, leaving us with a total of 255 samples. Considering the small sample size situation, in order to avoid cross-validation overfitting Ng et al. (1997) and achieving robust predictions for the different scenarios proposed, we address the following approach. We build 500 models that only differ in the random split of training and testing sets, including all variables. According to this, the dataset was split into different random train and test subsets, set to a 70/30 ratio. We train the model using the training subset and then we perform the predictions with the test subsets. Once the 500 models had been trained we collected all the predictions to better understand their distribution.

For visualization purposes we used the R programming language. A full list of the packages used is provided at the end of the methods section.

Ridge Regression is the simplest choice to achieve a balance between interpretability and precision. Since linear regression establishes a relationship between dependent variable and one or more independent variables that might be correlated, Ridge Regression imposes a penalty term on the size of the coefficients to overcome this issue, which is called multicollinearity (Gruber, 1998).

The aim of both Ordinary Least Squares (OLS) and Ridge Regression coefficients is to minimize the residual sum of squares (Saleh et al., 2019) and thus, the MSE. According to this, the penalty hyperparameter must be tuned so that model coefficients change in order to optimize the model error, by decreasing the residual sum of squares. As well as linear regression, the Ridge regressor explains the outcome as a function of multiple input variables. Thus, as a result, each input variable has an associated weight that will be positive or negative depending on its contribution to the model.

As well as linear regression, the Ridge regressor explains the outcome as a function of the multiple input variables. Thus, as a result, each input variable has an associated weight that will be positive or negative depending on its contribution to the model.

Random Forest Regression (RFR) is a tree-based ensemble method and belongs to the family of Classification and Regression Trees (CART). An ensemble method is a technique that combines the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model (Breiman, 2001).

RFR operates by constructing a multitude of decision trees, which are trained with a random subset of samples that have been drawn with replacement from the training sample. The number of variables included in each tree is limited to a percentage of the total variables that must be initially set. This ensures that the ensemble model does not rely too heavily on any individual variable, and makes fair use of all potentially predictive variables. As a result, the output estimation is the mean prediction of the individual trees.

Regarding interpretability, it is known that decision trees can be easily converted into rules which increase human interpretability of the results and explain why a decision was made. However, in the case of Random Forest it is not straightforward to find out the contribution of each of the variables (Rogers and Gunn, 2005). In Random Forests each Decision Tree is a set of internal nodes and leaves. In the internal node, the selected variable is used to make decision how to divide the data set into two separate sets with similar responses within. The variables for internal nodes are selected with some criterion, which for regression tasks is variance reduction. We can measure how each variable decrease the impurity of the split (the variable with highest decrease is selected for internal node). For each variable we can collect how on average it decreases the impurity. The average over all trees in the forest is the measure of the variable importance.

XGBoost (eXtreme Gradient Boosting) is another ensemble method that relies on the concept of gradient tree boosting. Boosting is a sequential algorithm that makes predictions for several rounds on the entire training sample and iteratively improves the performance of the algorithm with the information from the prior round's prediction accuracy. However, XGBoost produces black box models, hard to visualize and tune compared to RFR (Agarwal and Das, 2020). Note that our aim is not to compare performance across a wide range of modelling techniques, but to show how different modelling approaches ranging from simple Ridge regression to more complex XGBoost can be explored within our framework.

4.1.2 Variable engineering

To optimize the regressors, we followed a three-step process. In particular, we proceeded as follows:

I. Variable selection. The environmental variables that might be the most relevant for explaining lentil taxa distribution were identified using a modified R script developed for the *SelecVar* tool from CAPFITOGEN3 (Parra-Quijano et al. 2020).
II. Data normalization. Machine learning regressors typically require variables to have a close scale (Kotsiantis, 2011). The scale difference between variables can influence the performance of a ML regressor. Hence, we performed a data normalization scenario, namely standardization.
III. Variable Importance. For the three regression models (Ridge Regression, Random Forests, and XGBoost), we ran a procedure after the training process to estimate

variable importance. The importance of variables explaining rust resistance was carried out by analysing the weights assigned to the predictors for each model during the training step.

4.1.3 Model evaluation

To assess the performance of regression models we computed the Root Mean Square Error (RMSE). Although Pearson correlation coefficient is widely used in quantitative genetics (González-Recio et al., 2014), Mean Squared Error (MSE) and RMSE yield better performance in model selection when the sample size is small and there is a high variance in the outcome variable (Oliveira et al., 2018; Waldmann, 2019). RMSE is a distance between the vectors of recorded values (y_i) and predicted values ($\hat{y}_i$) (Chai and Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

4.2 Non-evaluated projections under current conditions

As it was previously mentioned, regression models were initially trained using present time data for rust-resistance evaluated accessions. After that, we applied the best performing model to non-evaluated populations. This dataset had the same present-time ecogeographic variables than the calibration dataset.

We used the trained regression model to perform DSr projections on crop wild relative populations. Thus, we can identify those wild populations that are most likely to be resistant according to their predicted DSr value. Those populations for which DSr projection falls within the first quartile for the continuous range of rust-evaluated populations ($DSr \leq 30.48$) and which are located in a protected area were selected as candidates to the long term *in situ* conservation.

4.3 DSr variation under climate change

Climate change models provide the future estimations for average temperature and yearly precipitation. According to this, we replaced both bioclimatic variables with future climate projections and applied these datasets as inputs to the predictive model. Those candidate populations to *in situ* conservation whose projected DSr under future climate conditions still falls within the first quartile of DSr distribution at present time were considered to be the populations most likely to retain the rust resistance trait, and, therefore, were selected as the most valuable populations for the establishment of genetic reserves.

List of statistical packages used

Python: python 3.8.8 (Van Rossum and Drake, 1995), matplotlib 3.3.4 (Hunter, 2007), numpy 1.20.1 (Harris et al., 2020), pandas 1.2.4 (Wes McKinney, 2010), seaborn 1.11.1 (Waskom, 2021), scikit-learn 0.24.1 (Pedregosa et al., 2011), verde 1.6.1 (Uieda, 2018), xgboost 1.4.2 (Chen and Guestrin, 2016). R: r-base 4.1.0 (R Core Team, 2020), countrycode 1.2.0 (Arel-Bundock et al., 2018), dplyr 3.4.0 (Wickham et al., 2022), forcats 0.5.1 (Wickham, 2021), gensysr 1.0.0 (Obreza, 2019), ggplot2 3.3.3 (Wickham, 2016), maps 3.3.0 (Brownrigg, 2018), readxl 1.3.1 (Wickham and Bryan, 2022), rgbif 0.9.8 (Chamberlain et al., 2017), scico 1.3.0 (Pedersen and Crameri, 2021), seewave 2.2.0 (Sueur et al., 2008).

5 Results

5.1 Predictive characterization

The training models generated with the three regression models rendered similar global results. Figure 1A shows the RMSE distributions for each of the three predictive models at present time, evaluated for the samples whose DSr values were known in advance. The Ridge model yielded the lowest median value, however this measure of centrality is not the only criterion to decide which is the most appropriate predictive model. Thus, the Ridge model showed a larger error spread than Random Forest, and, in addition, it was less accurate for small DSr values, precisely those corresponding to the most resistant samples and, therefore, the most valuable for conservation purposes (Figure 1B, Supplementary Table S2). Finally, a visual comparison of the distribution of the actual values and those predicted by the three methods reveals that the Ridge model tends to overpredict in the intermediate value range. Figures 1C–E include the values of the Kolmogorov-Smirnov distance which measures how the predicted and the evaluated values differ. Although the value of this difference is slightly lower for XGB than for Random Forest (0.114 vs. 0.116), the RMSE distribution of XGB showed a greater median value and error spread. For all these considerations, we selected Random Forest as the most suitable model to predict the DSr value of non-evaluated populations both at present and under different climate change scenarios.

We run the Variable Importance method for the Random Forest predictor, where results show that annual precipitation (mm) is the second most relevant, after Longitude. This fact becomes relevant specially when introducing the Climate Change models (Table 1), since annual precipitation was a variable susceptible to be modified, as well as annual average temperature that was at the fifth position in the variable importance ranking.

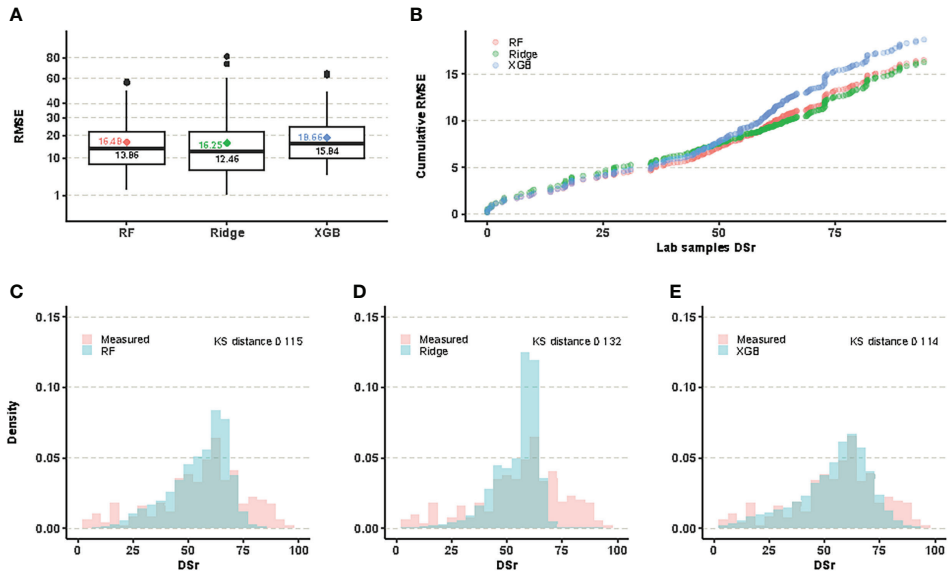


FIGURE 1
(A) RMSE of predicted DSr for evaluated samples for each of the three methods, average values are marked as color-filled diamonds. **(B)** The cumulative RMSE per sample is the sum of RMSE over the number of evaluated samples, and its value is equal to the average value for the last one. Random Forest is the best performer for low values of DSr. **(C–E)** plots compare the evaluated data distribution of DSr to the predicted one.

5.2 DSr variation under climate change

One of the main questions of this research is to assess how climate change may impact the ability of populations to maintain resistance to rust. An interesting result when comparing projections under current conditions and under climate change scenarios is that the global distribution of DSr values doesn't drift in a clear direction. Figure 2A shows that the distribution of the DSr value is very similar for the present and future predictions under the BCC370 model. The median remains almost unchanged, going from 33.70 to 33.85.

We have built the Random Forest predictor for all the climate change models described in methods. In what follows we always refer to the Random Forest regression under the conditions of change of the BCC370 model.

TABLE 1 Variable importance for the Random Forest predictor.

Variable	Importance
Longitude	0.2916
Annual precipitation	0.1623
Bulk density topsoil	0.0917
Available soil water capacity until wilting point	0.0866
Average annual temperature	0.0708
Topsoil sand fraction	0.0697
Topsoil total exchangeable bases	0.0651
Annual solar radiation	0.0615
Latitude	0.0548
Solar radiation December	0.0455

The variation in annual precipitation is weakly positive, with a median increase of 6 mm (Figure 2B). This change is highly concentrated around the median with -9 mm variation at the limit of the first quartile and 22 mm at the limit of the third quartile. In other words, locations with the lowest precipitation tended to lower their precipitation even further, whereas the other locations increased their precipitation, especially those which initially had the highest precipitation. Average annual temperature increases, on the other hand, are remarkable with values 1.28 °C, 1.56 °C and 1.71 °C for the first quartile, the median and the third quartile (Figure 2C).

The highest values of sensitivity to rust are found in some locations in the interior of the Iberian Peninsula, in regions with very hot and dry summers where rust cannot thrive (Figure 3A). Something similar occurs in the interior of France, but in this same country there is a cluster of potentially resistant populations in the final course of the Rhone, an area with much higher humidity conditions. Proximity to sea seems to foster populations with low DSr, as it happens in Greece, the Anatolian shoreline and Southern Crimea. Figure 3B shows the map of DSr variation with respect to the present time for the same wild accessions using Model BCC_370 instead. Spatial patterns are easy to spot. According to this, DSr values will decrease in the mountains of the Iberian Peninsula, but will be higher in the Rhone region and Southern Greece (See Supplementary Figures S2 and S3 for details). Annual precipitation is the main driver of

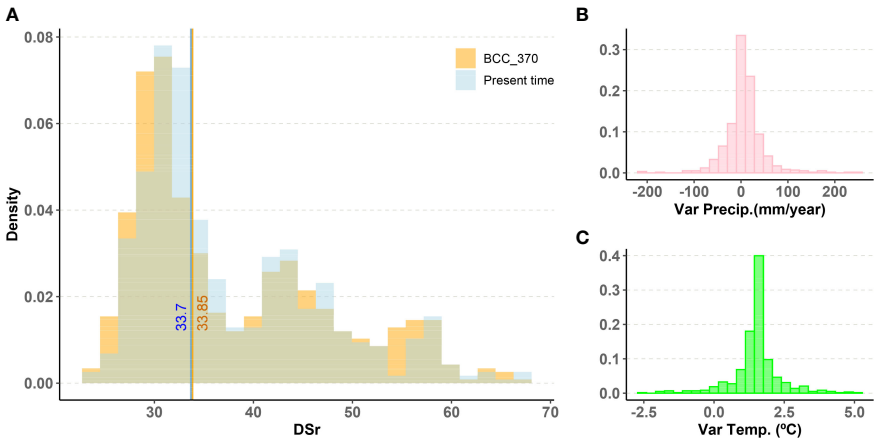


FIGURE 2
(A) Histogram of DSr prediction using Random Forest for non-evaluated wild populations in present time and in the future under BCC370 climate change model. Vertical lines mark the median value of each distribution. **(B)**, **(C)** Histograms of variation of average annual precipitation and average annual temperature between present time and future conditions under BCC370 climate model.

these changes, with a general trend to drought. Figure 3C shows the variation of precipitation according to the BCC_370 model.

Supplementary Table S1 shows the number of populations whose present DSr value falls under the first quartile for the present time (*DSr* ≤ 30.48) and are inside a protected area: 14 belong to *Lens ervoides* and 36 to *Lens nigricans*. Out of these 51 populations there are 16 that have a DSr value under 30.48 under the BCC370 hypothesis of Climate Change, 7 of *Lens ervoides* and 9 of *Lens nigricans*. They are encircled in magenta in Figure 3B. They are the most valuable populations for *in situ* conservation and are listed in Table 2.

6 Discussion

In this study, we addressed the pressing problem of the lack of adaptive resilience in modern cultivars in the context of climate change and the need to identify crop wild relative populations that might provide the genetic diversity needed to obtain specific traits. Rust is a severe disease for lentil production and breeding with resistant CWRs is a convenient way to reduce this problem but systematic straight identification of wild samples that have evolved to resist the fungus is not possible because of the expensive and time-consuming method to estimate Disease Severity. We tackled this problem by building a Machine Learning model using a rich dataset of calibrated samples grown under controlled environmental conditions. We used the model to predict the sensitivity to rust of a set of samples of wild relatives of *Lens culinaris* subsp. *culinaris*, in the present time and under 9 scenarios of Climate Change. As DSr is a quantitave measurement, we have gone straight to a regressive model instead of a qualitative binarized approach as in previous works. Our goal was to identify natural CWR populations in protected areas, that are likely to be resistant to rust at present time and to maintain this trait in the future due to the maintenance of selective pressures (i.e., have a projected low DSr at present time and in the future scenarios of climate change models). A quantitative estimation is better suited for the purpose of finding extremely resistant accessions.

With this methodological approach we aimed to answer three questions. The first one is which variables have more impact the model. Variable importance analysis revealed that Longitude is the most relevant one. This fact comes as no surprise given the East-West distribution of lentil wild species populations across the Mediterranean basin (Ladizinsky et al., 1983; Ladizinsky et al., 1984), and the evidences of westward migration from Near East of other wild *Fabaceae* like wild peas (Smýkal et al., 2017; Hellwig et al., 2022) or wild lupins (Mousavi-Derazmahalleh et al., 2018). Rust sensitivity is higher in the West Mediterranean basin, with extreme values in inner areas of the Iberian Peninsula and France, where the conditions are less prone to rust development. These results are in line with (Singh et al., 2014) who experimentally evaluated 405 wild lentil accessions and identified 27 promising rust-resistant populations which were mostly located in the Eastern Mediterranean (Syria and Turkey). Besides longitude, annual precipitation, bulk density topsoil, available soil water capacity and annual average temperature are the most relevant for the RFR predictor.

In response to the second question we identify 51 populations whose DSr values are under the first quartile in present time and grow inside a protected area. These are good rust-resistant candidate populations which should be evaluated for this trait and could be easily conserved *in situ*.

The third question constituting a relevant landmark of this work was to identify which of those 51 populations are likely to

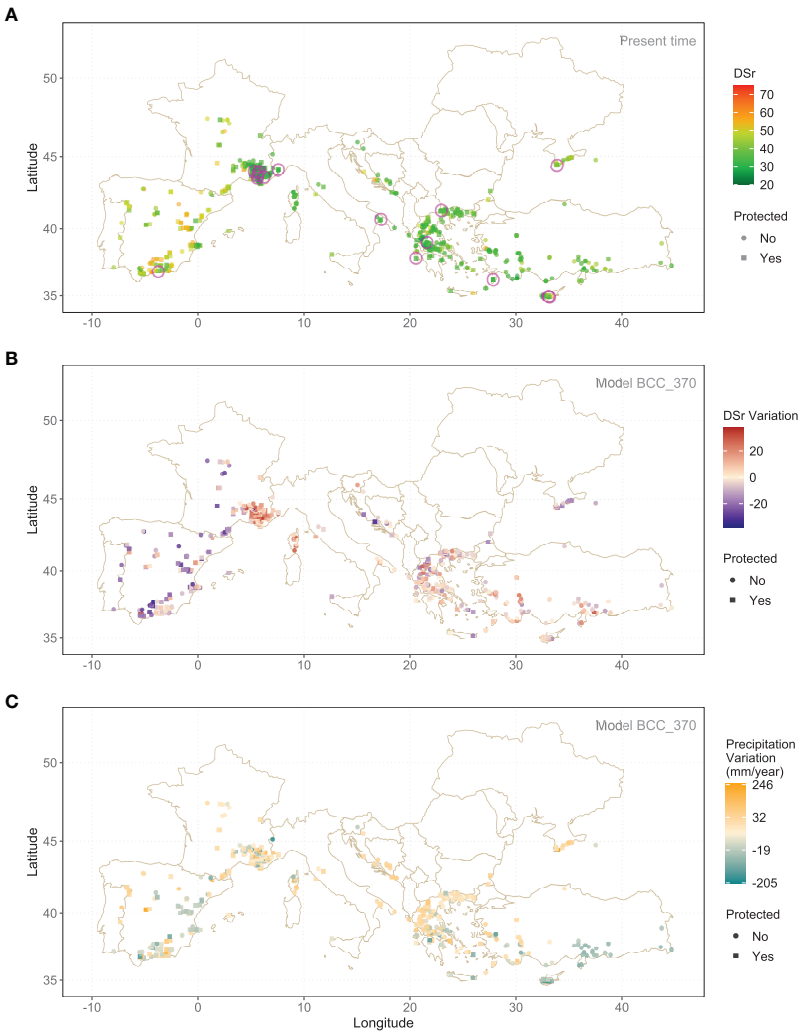


FIGURE 3
(A) DSr predicted median value, using Random Forest, for wild accessions under present climate conditions. **(B)** Variation of the same parameter according to BCC370 model. Encircled samples in **(A)** show a low DSr (first quartile) both in present time and in the future and are located inside a protected area. **(C)** Variation of annual precipitation.

maintain the environmental conditions that are favourable for the occurrence of rust-resistant genotypes under the projected Climate Change scenarios. Although longitude is the variable with the greatest weight in the prediction of our model, this variable is constant through time and not affected by climate change. Annual precipitation and average annual temperature were the variables that had an impact in the different scenarios on the future sensitivity to rust. The effect on the variation of DSr was similar for them all (Supplementary Figure S1; Supplementary Table S3), possibly because of the short span over which the change is projected (year 2040). Locations with the lowest precipitation tended to lower their precipitation even further, whereas the other locations increased their precipitation, especially those which initially had the highest precipitation. Average annual temperature increases, on the other hand, are remarkable with values 1.28 °C, 1.56 °C and 1.71 °C for the first quartile, the median and the third quartile. This could suggest that rust resistance is not affected by the changes in annual precipitation and mean temperature, but quite the opposite, we found clear patterns of DSr variation by geographic area. Annual precipitation was found to be the main driver of change and those regions with increased precipitation were associated to conditions more favourable for rust-resistant populations (lower DSr values). In particular, the accessions of South Crimea, the Iberian plateaus and Western France would be the ones that experiment a greater increase in precipitation and consequently a decrease in projected DSr. Just the opposite would happen on the Southern shore of Anatolia, where precipitation is going to decrease sharply reducing the chances of finding rust resistant genotypes in those locations.

TABLE 2 Selected subset of wild lentil populations for *in situ* conservation, occurring in protected areas and most likely to be rust-resistant at present and in the future according to the tested climate change projection models.

DSr Present	DSr Future	Long.	Lati.	Species	Site
29.57	28.84	33.16	34.88	*Lens ervoides*	Nicosia District, Ora, Greece
29.22	30.35	33.09	34.91	*Lens ervoides*	Nicosia District, Palaichori, Greece
28.96	29.89	27.87	36.20	*Lens ervoides*	Rodos, Mt. Attaviros, Greece
29.28	26.95	-3.72	36.82	*Lens ervoides*	PN de las Sierras de Tejeda, Almijara y Alhama, Spain
27.80	28.78	20.64	37.82	*Lens ervoides*	Zakynthos, Porto Vromi, Greece
29.07	24.21	21.66	38.98	*Lens ervoides*	Evrytania, NW Nea Viniani, Greece
29.24	29.37	33.83	44.40	*Lens ervoides*	Near Black Sea and Sanatome, Crimea
27.95	29.16	17.25	40.67	*Lens nigricans*	Puglia, 14 km from Massafra, Italy
30.04	24.68	23.02	41.29	*Lens nigricans*	Serrai, S Kato Poroia, Greece
26.49	29.37	5.66	43.54	*Lens nigricans*	Brians, France
28.55	28.95	6.28	43.61	*Lens nigricans*	Ampus, France
28.84	28.08	5.67	43.72	*Lens nigricans*	Mirabeau, France
28.52	26.61	5.68	43.90	*Lens nigricans*	Reillane, France
30.28	27.10	5.81	43.99	*Lens nigricans*	Forcalquier, France
27.54	30.11	5.31	44.00	*Lens nigricans*	Lioux, France
30.22	26.94	7.56	44.11	*Lens nigricans*	Tende, France

We identified 16 populations in protected areas with DSr values below 30.48 (first quartile of the distribution of the DSr projection) at the present time and under the future environmental conditions (Table 2). Eight populations are located in the Eastern Mediterranean basin (6 belonging to *Lens ervoides* and 2 to *Lens nigricans*). In the Western basin, 7 populations are in the South of France (all of them belonging to *Lens nigricans*) and only 1 is in the Iberian Peninsula (*Lens ervoides*). There is a remarkable difference among the CWR species. While 7 out of 14 populations of *Lens ervoides* selected at present time will remain of high interest in the future from the rust-resistance point of view, only 9 of 37 populations of *Lens nigricans* fall within this category. They all share a common geographic feature, they are very close to the coast, and 5 of them are on islands (Rhodes, Zakhyntos, Crete and Cyprus). There are not populations of *Lens lamottei* and *Lens culinaris* subsp. *orientalis* in the selected subset of high interest populations for *in situ* conservation, but the initial number of populations for these species was small compared to the other species. This fact suggests an interest for focusing future field work on these underrepresented taxa.

7 Conclusion

Crop Wild Relatives of *Lens culinaris* subsp. *culinaris* are a source of genetic resistance to a commonrust disease caused by the fungus *Uromyces viciae-fabae*. The quantitative field evaluation of rust resistance is a hard, expensive, and time-consuming task. The use of Machine Learning approaches provided a way to mitigate this obstacle, using a carefully built calibration set. Our results identified 16 populations that are likely to be resistant to rust, occur in protected areas, and are expected to be resilient under predicted Climate Change conditions. Thus, they are sound candidates for the establishment of genetic reserves for *in situ* conservation. Further characterization by field evaluation of these populations is needed to check the validity of Machine Learning predictions and improve the genetic value of the calibration set. The same method may be extended to predict pest and pathogen resistance traits of CWRs of other crops.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://zenodo.org/record/6883274.

Author contributions

Conceptualization: IC-G, MLRT, JI, JG, JG-A. Data curation: DR, MLRT, JI. Funding acquisition: JI, JG, DR. Methodology: IC-G, JG-A, JG, JI. Machine learning models: IC-G, JG-A, MLRT. Visualization: JG-A. Writing, original draft: IC-G, JG-A, JI, JG. Writing, review & editing: IC-G, MLRT, JG-A, JG, DR, JI. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by the Farmer's Pride project: Networking, Partnerships, and Tools to Enhance *in*

situ Conservation of European Plant Genetic Resources, a 3-year EU-funded project funded by the Horizon 2020 Framework Programme of the European Union, grant agreement no. 774271. J.G. acknowledges financial support provided by the Ministerio de Ciencia, Innovación y Universidades (*PGC*2018-093854-*B*-100). This research was partially funded too by *I* + *D* + *i* Plan Andaluz Investigación LEGAND project (P20_0986).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1010799/full#supplementary-material

References

Adhikari, U., Nejadhashemi, A. P., and Woznicki, S. A. (2015). Climate change and Eastern Africa: A review of impact on major crops. *Food Energy Secur.* 4, 110–132. doi: 10.1002/fes3.61

Agarwal, N., and Das, S. (2020). "Interpretable Machine Learning Tools: A Survey. In *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)* Canberra, Australia, 1–4, 1528–1534

Amouzoune, M., Amri, A., Benkirane, R., Kehel, Z., Al-Jaboobi, M., Moulakat, A., et al. (2022). Mining and predictive characterization of resistance to leaf rust (*Puccinia hordei otth*) using two subsets of barley genetic resources. *Genet. Resour. Crop Evol.* 69, 839–853. doi: 10.1007/s10722-021-01268-4

Anderson, R., Bayer, P. E., and Edwards, D. (2020). Climate change and the need for agricultural adaptation. *Curr. Opin. Plant Biol.* 56, 197–202. doi: 10.1016/j.pbi.2019.12.006

Araújo, M., and Guisan, A. (2006). Araujo mb, guisan a. five (or so) challenges for species distribution modelling. journal of biogeography. *J. Biogeogr.* 33, 1677–1688. doi: 10.1111/j.1365-2699.2006.01584.x

Arel-Bundock, V., Enevoldsen, N., and Yetman, C. (2018). Countrycode: An r package to convert country names and country codes. *J. Open Source Software.* 3, 848. doi: 10.21105/joss.00848

Arojju, S. K., Conaghan, P., Barth, S., Milbourne, D., Casler, M. D., Hodkinson, T. R., et al. (2018). Genomic prediction of crown rust resistance in lolium perenne. *BMC Genet.* 19, 1–10. doi: 10.1186/s12863-018-0613-z

Bari, A., Amri, A., Street, K., Mackay, M., De Pauw, E., Sanders, R., et al. (2014). Predicting resistance to stripe (yellow) rust (*Puccinia striiformis*) in wheat genetic resources using focused identification of germplasm strategy. *J. Agric. Sci.* 152, 906–916. doi: 10.1017/S0021859613000543

Barilli, E., Sillero, J. C., Fernández-Aparicio, M., and Rubiales, D. (2009). Identification of resistance to *Uromyces pisi* (pers.) wint. in *Pisum* spp. germplasm. *Field Crops Res.* 114, 198–203. doi: 10.1016/j.fcr.2009.07.017

Bari, A., Street, K., Mackay, M., Endresen, D. T. F., De Pauw, E., and Amri, A. (2012). Focused identification of germplasm strategy (figs) detects wheat stem rust resistance linked to environmental variables. *Genet. Resour. Crop Evol.* 59, 1465–1481. doi: 10.1007/s10722-011-9775-5

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. (*1st ed.*). New York: Routledge. doi: 10.1201/9781315139470

Brownrigg, R. (2018) *Package 'maps'*. Available at: https://cran.r-project.org/package=maps.

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2, 1–6. doi: 10.1038/nplants.2016.22

Chai, T., and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae). *Geosci. Model. Dev. Discussions.* 7, 1525–1534. doi: 10.5194/gmdd-7-1525-2014

Chamberlain, S., Ram, K., Barve, V., Mcglinn, D., and Chamberlain, M. S. (2017) *Package 'rgbif'*. Available at: https://cran.r-project.org/package=rgbif.

Chen, T., and Guestrin, C. (2016). "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. (ACM), 785–794.

Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., et al. (2018). Climate change and interconnected risks to sustainable development in the mediterranean. *Nat. Climate Change* 8, 972–980. doi: 10.1038/s41558-018-0299-2

Cutler, D. R., Edwards, T. C.Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792. doi: 10.1890/07-0539.1

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the ipsl-cm5 earth system model: from cmip3 to cmip5. *Climate Dynamics.* 40, 2123–2165. doi: 10.1007/s00382-012-1636-1

El Haddad, N., Kabbaj, H., Zaïm, M., El Hassouni, K., Tidiane Sall, A., Azouz, M., et al. (2021). Crop wild relatives in durum wheat breeding: Drift or thrift? *Crop Sci.* 61, 37–54. doi: 10.1002/csc2.20223

Endresen, D. T. F. (2010). Predictive association between trait data and ecogeographic data for nordic barley landraces. *Crop Sci.* 50, 2418–2430. doi: 10.2135/cropsci2010.03.0174

Endresen, D. T. F., Street, K., Mackay, M., Bari, A., Amri, A., De Pauw, E., et al. (2012). Sources of resistance to stem rust (ug99) in bread wheat and durum wheat identified using focused identification of germplasm strategy. *Crop Sci.* 52, 764–773. doi: 10.2135/cropsci2011.08.0427

Fellmann, T., Witzke, P., Weiss, F., Van Doorslaer, B., Drabik, D., Huck, I., et al. (2018). Major challenges of integrating agriculture into climate change mitigation policy frameworks. *Mitigation. Adapt. Strat. Global Change* 23, 451–468. doi: 10.1007/s11027-017-9743-2

Food and of the United Nations A. O (2014). *Herramientas CAPFITOGEN: Programa para El fortalecimiento de las capacidades en programas nacionales de recursos* (FAO).

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat*, 29, 1189–1232. doi: 10.1214/aos/1013203451

Fu, Y.-B. (2017). The vulnerability of plant genetic resources conserved ex situ. *Crop Sci.* 57, 2314–2328. doi: 10.2135/cropsci2017.01.0014

Garcia, R. M., Parra-Quijano, M., and Iriondo, J. M. (2017). Identification of ecogeographical gaps in the spanish aegilops collections with potential tolerance to drought and salinity. *PeerJ* 5, e3494. doi: 10.7717/peerj.3494

González-Recio, O., Rosa, G. J., and Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock. Sci.* 166, 217–231. doi: 10.1016/j.livsci.2014.05.036

Gruber, M. (1998). *Improving efficiency by shrinkage: The James-stein and ridge regression estimators* Vol. 156 (New York, CRC Press).

Guisan, A., and Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x

Hannah, L. (2010). A global conservation system for climate-change adaptation. *Conserv. Biol.* 24, 70–77. doi: 10.1111/j.1523-1739.2009.01405.x

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2

Hellwig, T., Abbo, S., and Ophir, R. (2022). Drivers of genetic differentiation and recent evolutionary history of an eurasian wild pea. *J. Biogeogr.* 49, 794–808. doi: 10.1111/jbi.14274

Heywood, V., Casas, A., Ford-Lloyd, B., Kell, S., and Maxted, N. (2007). Conservation and sustainable use of crop wild relatives. *Agricult. Ecosyst. Environ.* 121, 245–255. doi: 10.1016/j.agee.2006.12.014

Hoerl, A. (1962). Application of ridge analysis to regression problems. *Chem. Eng. Prog.* 58, 54–59.

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M., and Meinke, H. (2007). Adapting agriculture to climate change. *Proc. Natl. Acad. Sci.* 104, 19691–19696. doi: 10.1073/pnas.0701890104

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Hunter, D. (2012). *Crop wild relatives: a manual of in situ conservation* (Routledge, London), 452.

Hunter, D., Maxted, N., Heywood, V., Kell, S., and Borelli, T. (2012). Protected areas and the challenge of conserving crop wild relatives. *Parks* 18, 87.

Iriondo, J. M., Maxted, N., and Dulloo, M. E. (2008). *Conserving plant genetic diversity in protected areas: population management of crop wild relatives* (Wallingford (UK), CAB International).

Khoury, C. K., Brush, S., Costich, D. E., Curry, H. A., de Haan, S., Engels, J. M., et al. (2022). Crop genetic erosion: Understanding and responding to loss of crop diversity. *New Phytol.* 233, 84–118. doi: 10.1111/nph.17733

Khoury, C. K., Carver, D., Greene, S. L., Williams, K. A., Achicanoy, H. A., Schori, M., et al. (2020). Crop wild relatives of the united states require urgent conservation action. *Proc. Natl. Acad. Sci.* 117, 33351–33357. doi: 10.1073/pnas.2007029117

Kotsiantis, S. (2011). Feature selection for machine learning classification problems: a recent overview. *Artif. Intell. Rev.* 42, 157–176. doi: 10.1007/s10462-011-9230-1

Labokas, J., Maxted, N., Kell, S., Brehm, J. M., and Iriondo, J. M. (2018). Development of national crop wild relative conservation strategies in european countries. *Genet. Resour. Crop Evol.* 65, 1385–1403. doi: 10.1007/s10722-018-0621-x

Ladizinsky, G., Braun, D., Goshen, D., and Muehlbauer, F. (1984). The biological species of the genus. *Lens. L. Bot. Gazette.* 145, 253–261. doi: 10.1086/337454

Ladizinsky, G., Braun, D., and Muehlbauer, F. (1983). Evidence for domestication of *Lens nigricans* (M. bieb.) godron in s Europe. *Bot. J. Linn. Soc.* 87, 169–176. doi: 10.1111/j.1095-8339.1983.tb00988.x

Liber, M., Duarte, I., Maia, A. T., and Oliveira, H. R. (2021). The history of lentil *(Lens culinaris subsp. culinaris)* domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. *Front. Plant Sci.* 12, 628439. doi: 10.3389/fpls.2021.628439

Maxted, N., Iriondo, J. M., Dulloo, M. E., and Lane, A. (2008). The integration of PGR Conservation with Protected Area Management. In: Conserving plant genetic diversity in protected areas. *Population management of crop wild relatives* J. M. Iriondo, M. E. Dulloo and N. Maxted (Wallingford. (UK): CAB International) 1–22.

Maxted, N., Ford-Lloyd, B.V., Kell, S.P., Iriondo, J.M., and Dulloo Turok, M.E. J. (2008). *Population management of crop wild relatives* (Wallingford. (UK): CAB International) 1–22.

Maxted, N., Ford-Lloyd, B.V., Kell, S.P., Iriondo, J.M., and Dulloo Turok, M.E. J. (eds.) (2008). *P Crop Wild Relative Conservation and Use* (Wallingford. (UK): CAB International).

Maxted, N., Iriondo, J. M., Dulloo, M. E., and Lane, A. (2008). Introduction: The integration of PGR Conservation with Protected Area Management. In: Conserving plant genetic diversity in protected areas. *Population management of crop wild relatives* J. M. Iriondo, M. E. Dulloo and N. Maxted (Wallingford. (UK): CAB International) pp. 1–22.

Maxted, N., and Kell, S. (2009). *Establishment of a global network for the situ conservation of crop wild relatives: Status and needs* (Rome, Italy: FAO Commission on Genetic Resources for Food and Agriculture).

Maxted, N., Kell, S., Ford-Lloyd, B., Dulloo, E., and Toledo, Á. (2012). Toward the systematic conservation of global crop wild relative diversity. *Crop Sci.* 52, 774–785. doi: 10.2135/cropsci2011.08.0415

Meilleur, B. A., and Hodgkin, T. (2004). *In situ* conservation of crop wild relatives: status and trends. *Biodivers. Conserv.* 13, 663–684. doi: 10.1023/B:BIOC.0000011719.03230.17

Mousavi-Derazmahalleh, M., Bayer, P. E., Nevado, B., Hurgobin, B., Filatov, D., Kilian, A., et al. (2018). Exploring the genetic and adaptive diversity of a pan-mediterranean crop wild relative: narrow-leafed lupin. *Theor. Appl. Genet.* 131, 887–901. doi: 10.1007/s00122-017-3045-7

Negussie, T., and Pretorius, Z. (2012). Lentil rust: Present status and future prospects. *Crop Prot.* 32, 119–128. doi: 10.1016/j.cropro.2011.11.004

Ng, A. Y. (1997). "Preventing" overfitting" of cross-validation data," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML '97)*, vol. 97, 245–253.

Obreza, M. (2019) Preventing "overfitting" of cross-validation data. Machine Learning: Proceedings of the Fourteenth International Conference (ICML '97), 245–253

Olesen, J. E., and Bindi, M. (2002). Consequences of climate change for european agricultural productivity, land use and policy. *Eur. J. Agron.* 16, 239–262. doi: 10.1016/S1161-0301(02)00004-7

Oliveira, M., Pereira, G., Ferreira, E., Santos, J., Knezevic, S., and Werle, R. (2018). Additive design: the concept and data analysis. *Weed. Res.* 58, 338–347. doi: 10.1111/wre.12317

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., et al. (2014). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* 122, 387–400. doi: 10.1007/s10584-013-0905-2

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (scenariomip) for cmip6. *Geosci. Model. Dev.* 9, 3461–3482. doi: 10.5194/gmd-9-3461-2016

QGIS Org (2021). *Qgis geographic information system (QGIS Association)*. Open source geospatial foundation project. Available at: https://qgis.org/es.

Parra-Quijano, M., Iriondo, J. M., Torres, M. E., López, F., Maxted, N., Kell, , , S., et al (2020). CAPFITOGEN3. A Toolbox for the Conservation and Promotion of the use of Agricultural Biodiversity. (Bogotá, Colombia: Universidad Nacional de Colombia. Bogotá campus – Faculty of Agricultural Sciences). Available online at: http://www.capfitogen.net/es/acceso/manuales/

Pedersen, T. L., and Crameri, F. (2021) *Package 'scico'*. Available at: https://cran.r-project.org/package=scico.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. Available at: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Rauf, S., da Silva, J. T., Khan, A. A., and Naveed, A. (2010). Consequences of plant breeding on genetic diversity. *Int. J. Plant Breed.* 4, 1–21.

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., and Chatterjee, S. (2019). Climate change has likely already affected global food production. *PloS One* 14, e0217148. doi: 10.1371/journal.pone.0217148

R Core Team (2020). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Ren, J., Li, Z., Wu, P., Zhang, A., Liu, Y., Hu, G., et al. (2021). Genetic dissection of quantitative resistance to common rust (*Puccinia sorghi*) in tropical maize (*zea mays* l.) by combined genome-wide association study, linkage mapping, and genomic prediction. *Front. Plant Sci.* 12, 1338. doi: 10.3389/fpls.2021.692205

Rogers, J., and Gunn, S. (2005). "Identifying feature relevance using a random forest," in *International statistical and optimization perspectives workshop" subspace, latent structure and feature selection* (Berlin, Springer), 173–184.

Rubiales, D., Castillejo, M., Madrid, E., Barilli, E., and Rispail, N. (2011). Legume breeding for rust resistance: lessons to learn from the model *Medicago truncatula*. *Euphytica* 180, 89–98. doi: 10.1007/s10681-011-0367-4

Rubiales, D., Fondevilla, S., and Fernández-Aparicio, M. (2020). Development of pea breeding lines with resistance to *Orobanche crenata* derived from pea landraces and wild *Pisum* spp. *Agronomy* 11, 36. doi: 10.3390/agronomy11010036

Rubiales, D., Rojas-Molina, M. M., and Sillero, J. C. (2013). Identification of pre- and posthaustorial resistance to rust (*Uromyces viciae-fabae*) in lentil (*Lens culinaris*) germplasm. *Plant Breed.* 132, 676–680. doi: 10.1111/pbr.12096

Rubio Teso, M. L., Álvarez Muñiz, C., Gaisberger, H., Kell, S., Lara-Romero, C., Magos Brehm, J., et al. (2020). *In situ plant genetic resources in Europe: crop wild relatives*. Farmer's Pride Project, https://more.bham.ac.uk/farmerspride/wp-content/uploads/sites/19/2020/10/D1.2_In_situ_PGR_in_Europe_crop_wild_relatives.pdf

Rubio Teso, M. L., Lara-Romero, C., Rubiales, D., Parra-Quijano, M., and Iriondo, J. M. (2022). Searching for abiotic tolerant and biotic stress resistant wild lentils for introgression breeding through predictive characterization. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.817849

Rufo, R., Alvaro, F., Royo, C., and Soriano, J. M. (2019). From landraces to improved cultivars: Assessment of genetic diversity and population structure of mediterranean wheat using snp markers. *PloS One* 14, e0219867. doi: 10.1371/journal.pone.0219867

Saleh, A. M. E., Arashi, M., and Kibria, B. G. (2019). *Theory of ridge regression estimation with applications*. 285 (Oxford (UK): John Wiley & Sons).

Sánchez, R. M. G., Parra-Quijano, M., Greene, S., and Iriondo, J. M. (2019). Predictive characterisation identifies global sources of acyanogenic germplasm of a key forage species. *Crop Pasture Sci.* 70, 546–554. doi: 10.1071/CP18346

Seiler, G. J., Qi, L. L., and Marek, L. F. (2017). Utilization of sunflower crop wild relatives for cultivated sunflower improvement. *Crop Sci.* 57, 1083–1101. doi: 10.2135/cropsci2016.10.0856

Sillero, J. C., Rojas-Molina, M. M., Emeran, A. A., Kharrat, M., Winkler, J., Khan, H. R., et al. (2017). Identification and multi-environment validation of resistance to rust (*Uromyces viciae-fabae*) in *vicia faba*. *Crop Pasture Sci.* 68, 1013–1023. doi: 10.1071/CP17099

Singh, M., Bisht, I. S., Kumar, S., Dutta, M., Bansal, K. C., Karale, M., et al. (2014). Global wild annual lens collection: a potential resource for lentil genetic base broadening and yield enhancement. *PloS One* 9, e107781. doi: 10.1371/journal.pone.0107781

Smith, P., Howden, M., Krug, M-D, Mbow, C., Pörtner, H.O., Reisinger, A., et al (2017). Special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (sr2). Geneva: IPCC, 650

Smýkal, P., Hradilová, I., Trněný, O., Brus, J., Rathore, A., Bariotakis, M., et al. (2017). Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-17623-4

Stockwell, D. R., and Peterson, A. (2002). Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13. doi: 10.1016/S0304-3800(01)00388-X

Sueur, J., Aubin, T., and Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* 18, 213–226. doi: 10.1080/09524622.2008.9753600

Triviño, M., Kujala, H., Araújo, M. B., and Cabeza, M. (2018). Planning for the future: identifying conservation priority areas for iberian birds under climate change. *Landscape Ecol.* 33, 659–673. doi: 10.1007/s10980-018-0626-z

Uieda, L. (2018). Verde: Processing and gridding spatial data using green's functions. *J. Open Source Software.* 3, 957. doi: 10.21105/joss.00957

Van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., and Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* 120, 1241–1252. doi: 10.1007/s00122-009-1252-6

Van Rossum, G., and Drake, F. L.Jr (1995). *Python Reference manual* (Amsterdam, Centrum voor Wiskunde en Informatica Amsterdam).

van Treuren, R., Hoekstra, R., and van Hintum, T. J. (2017). Inventory and prioritization for the conservation of crop wild relatives in the netherlands under climate change. *Biol. Conserv.* 216, 123–139. doi: 10.1016/j.biocon.2017.10.003

Vincent, H., Amri, A., Castañeda-Álvarez, N. P., Dempewolf, H., Dulloo, E., Guarino, L., et al. (2019). Modeling of crop wild relative species identifies areas globally for *in situ* conservation. *Commun. Biol.* 2, 1–8. doi: 10.1038/s42003-019-0372-z

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., et al. (2013). The cnrm-cm5. 1 global climate model: description and basic evaluation. *Climate Dynamics.* 40, 2091–2121. doi: 10.1007/s00382-011-1259-y

Waldmann, P. (2019). On the use of the pearson correlation coefficient for model evaluation in genome-wide prediction. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00899

Waskom, M. L. (2021). Seaborn: statistical data visualization. *J. Open Source Software.* 6, 3021. doi: 10.21105/joss.03021

McKinney, W (2010). "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in science conference*. Eds. S. van der Walt and J. Millman, 51–56. doi: 10.25080/Majora-92bf1922-00a

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

Wickham, H. (2021) *Package 'forcats'*. Available at: https://cran.r-project.org/package=forcats.

Wickham, H., and Bryan, J. (2022) *Readxl: Read excel files*. Available at: https://github.com/tidyverse/readxl.

Wickham, H., François, R., Henry, L., and Müller, K. (2022) *Dplyr: A grammar of data manipulation*. Available at: https://github.com/tidyverse/dplyr.

Wiebe, K., Robinson, S., and Cattaneo, A. (2019). Climate change, agriculture and food security: impacts and the potential for adaptation and mitigation. *Sustain. Food Agric.* 2019, 55–74. doi: 10.1016/B978-0-12-812134-4.00004-2

Wing, I. S., De Cian, E., and Mistry, M. N. (2021). Global vulnerability of crop yields to climate change. *J. Environ. Econ. Manage.* 109, 102462. doi: 10.1016/j.jeem.2021.102462

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., et al. (2014). An overview of bcc climate system model development and application for climate change studies. *J. Meteorol. Res.* 28, 34–56. doi: 10.1007/s13351-014-3041-7

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B.-H. (2017). Back into the wild–apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* 10, 5–24. doi: 10.1111/eva.12434

Zohary, D., Hopf, M., and Weiss, E. (2012). *Domestication of plants in the old world: The origin and spread of domesticated plants in southwest Asia, Europe, and the Mediterranean basin* (Oxford (UK): Oxford University Press).

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 December 2022
DOI 10.3389/fpls.2022.1059197

OPEN ACCESS

EDITED BY
Mohd. Kamran Khan,
Selcuk University, Turkey

REVIEWED BY
Deyong Ren,
China National Rice Research Institute (CAAS), China
Md Nashir Uddin,
The University of Tokyo, Japan

*CORRESPONDENCE
Wenfu Chen
wfchen@syau.edu.cn
Jian Sun
sunjian811119@syau.edu.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 01 October 2022
ACCEPTED 11 November 2022
PUBLISHED 05 December 2022

CITATION
Li Z, Gui R, Yu X, Liang C, Cui J, Zhao X, Zhang X, Yu P, Chen W and Sun J (2022) Genetic basis of the early heading of high-latitude weedy rice. *Front. Plant Sci.* 13:1059197. doi: 10.3389/fpls.2022.1059197

Genetic basis of the early heading of high-latitude weedy rice

Zhuan Li, Rui Gui, Xiaoyu Yu, Chengwei Liang, Juan Cui, Xue Zhao, Xuemin Zhang, Pengcheng Yu, Wenfu Chen* and Jian Sun*

Rice Research Institute, Shenyang Agricultural University, Shenyang, China

Japonica rice (*Oryza sativa* L.) is an important staple food in high-latitude regions and is widely distributed in northern China, Japan, Korea, and Europe. However, the genetic diversity of *japonica* rice is relatively narrow and poorly adapted. Weedy rice (*Oryza sativa f. spontanea*) is a semi-domesticated rice. Its headings are earlier than the accompanied *japonica* rice, making it a potential new genetic resource, which can make up for the defects of wild rice that are difficult to be directly applied to *japonica* rice improvement caused by reproductive isolation. In this study, we applied a natural population consisting of weedy rice, *japonica* landrace, and *japonica* cultivar to conduct a genome-wide association study (GWAS) of the heading date and found four loci that could explain the natural variation of the heading date in this population. At the same time, we developed recombinant inbred lines (RILs) crossed by the early-heading weedy rice WR04-6 and its accompanied *japonica* cultivar ShenNong 265 (SN265) to carry out a QTL mapping analysis of the heading date and mapped four quantitative trait locus (QTLs) and three epistatic effect gene pairs. The major locus on chromosome 6 overlapped with the GWAS result. Further analysis found that two genes, *Hd1* and *OsCCT22*, on chromosome 6 (Locus 2 and Locus 3) may be the key points of the early-heading character of weedy rice. As minor effect genes, *Dth7* and *Hd16* also have genetic contributions to the early heading of weedy rice. In the process of developing the RIL population, we introduced fragments of Locus 2 and Locus 3 from the weedy rice into super-high-yielding *japonica* rice, which successfully promoted its heading date by at least 10 days and expanded the rice suitable cultivation area northward by about 400 km. This study successfully revealed the genetic basis of the early heading of weedy rice and provided a new idea for the genetic improvement of cultivated rice by weedy rice.

KEYWORDS

GWAS, genetic resources, heading date, weedy rice, QTL mapping

Introduction

Japonica rice (*Oryza sativa* L.) is a staple food for most people in China, particularly in the northeast. It is crucial for economic development and ensuring global and national food security (Mao et al., 2017). Superior *japonica* varieties have been applied over a wide range for a long time. However, their long-term application will reduce genetic diversity (Wang et al., 2014). Due to genetic bottlenecks, new improvement strategies have encountered challenges.

Although wild rice resources have been widely suggested for rice improvement, the unfavorable characteristics of the hybrid offerings of wild rice and *japonica* rice far exceed the favorable characteristics due to reproductive isolation and geographic and genetic distance. Thus, wild rice is difficult to apply in breeding programs (Brambilla and Fornara, 2013; Brar and Khush, 2018; Xie et al., 2019).

As a semi-domesticated genetic resource, weedy rice (*Oryza sativa f. spontanea*) has no reproductive isolation with accompanied cultivar and adapts to the current paddy field ecosystem. In addition, weedy rice also possesses some adaptive alleles that cultivars lack. Among these alleles, the early-heading genes have great application potential for improving *japonica* cultivars, while the underlying molecular genetic mechanisms of the early heading in weedy rice are still not revealed, which hinders its use in breeding applications (Zhao et al., 2018; Sun et al., 2019).

With the rapid development of next-generation high-throughput sequencing technology and multi-omics, many genes related to the rice heading have been identified. The identification of genes related to photoperiodic pathway and analysis of their molecular regulatory network mechanisms can not only reveal the genetic variation in the rice heading but also guide the genetic improvement of rice breeding, which is of great significance and necessary to the high-yield breeding of rice (Ebana et al., 2011; Brambilla and Fornara, 2013; Shrestha et al., 2014; Hori et al., 2015; Onogi et al., 2016). Rice has two known relatively conserved photoperiodic flowering pathways: the *Hd1*-centered gene and its related genes under short-day (SD) conditions and the *Ehd1*-centered gene and its regulatory genes under long-day (LD) conditions (Li et al., 2015; Chen et al., 2022).

So far, many heading genes have been identified; however, the reason for weedy rice heading earlier than their accompanied cultivars is still largely unknown. We performed a genome-wide associated study (GWAS) and QTL mapping to reveal this genetic basis in this study. Meanwhile, we provide new genetic materials and gene resources for breeding early-heading and widely adapted *japonica* cultivated rice.

Materials and methods

Plant materials and growth conditions

A total of 274 accessions, including 154 modern *japonica* cultivars, 87 *japonica* landrace, and 33 *japonica* weedy rice at Asian high latitudes (WRAH), were planted in the paddy field in three consecutive years from 2019 to 2021 for GWAS of the heading date. Another recombinant inbred line (RIL) population of 165 individuals derived from the weedy rice WR04-6 and *japonica* cultivar ShenNong 265 (SN265 called super rice) was used for QTL mapping of the heading date. All materials were planted at Shenyang Agricultural University (123°25′E, 41°48′N, Liaoning Province, China, temperate semi-humid continental climate, under natural long day-length conditions, average day length >14 h from May to September, average temperature of 23.5°C, fertilizer applied in the field with the following standards: urea 300 kg/hm^2, diammonium 220 kg/hm^2, potassium chloride 220 kg/hm^2). Each variety was planted in three rows with 1.2-m row length, 30 cm apart between rows, and 12-cm space in rows, with one plant per hill. Field management followed normal agricultural practices. The list of accessions used in this study is displayed in Supplementary Tables S1, S2.

Phenotyping and statistical analysis

The first panicle beyond flag leaf 1 cm was recorded as the heading. The heading date was calculated as the date from sowing to the heading of half of the accession. The phenotype data are displayed in Supplementary Tables S1 and S2. Microsoft Excel 2019 and GraphPad Prism 9.0 were used for statistical analysis. The heading date difference of three ecotypes was tested by one-way ANOVA multiple comparisons. The D'Agostino–Pearson test method was used for the normal distribution test of the RIL population.

Genome-wide association study

The genotype of the GWAS panel for the heading date has been effectively used to study the genetic basis of agronomic traits in weedy rice (Sun et al., 2022). All raw reads were screened for high quality with Q20 quality scores >95% and Guanine and cytosine (GC) content <50%. The reads of each accession were aligned and then mapped to the reference genome (IRGSP1.0) using BWA software. We used samtools v0.1.19 and GATK v4.0 for population SNP variant data, and potential PCR duplicates were removed. It is also necessary to control the single nucleotide polymorphism (SNP) quality with a minor allele

frequency >0.05 and no more than 50% missing data. Finally, we got a high-quality SNP haplotype map with an average coverage depth of 23.2 times for each sample to proceed with follow-up analysis (Supplementary Figure S1). We performed the GWAS based on 1,311,445 genetic markers and a heading date of 3 years by using EMMAX software to fit a linear mixed model. To control the population structure during the GWAS, the first two principal components of the principal component analysis (PCA) were used as covariates. The Manhattan plots of GWAS results were drawn by R package CMplot. The threshold for genome-wide significance was determined by 10^{-5} and 10^{-6}.

Linkage disequilibrium block analysis and candidate gene association analysis

Based on the result of the GWAS, we performed linkage disequilibrium (LD) block analysis for the 2MB region surrounding the leading SNP. The LD block was defined by LDblockShow software. The genes located within the high LD blocks were selected for checking the gene annotation (https://rapdb.dna.affrc.go.jp/). Then, a haplotype-based association analysis for the genes within the LD block was conducted by candihap software (Li et al., 2020).

Genotyping for genetic map construction and quantitative trait locus (QTL) mapping

DNA was extracted from fresh leaves of RILs by the Cetyltrimethylammonium Bromide (CTAB) method. The genotypes were obtained using a 50K liquid-phase sequence capture chip (Guo et al., 2021). According to the genotyping results, 13,097 polymorphic sites were used for genetic map construction. Finally, a linkage map containing 2,275 bins was constructed based on the 13,097 polymorphic sites, spanning a total genetic distance of 2,068.6 cM, and the average genetic distance between adjacent markers was 0.91 cM. Marker information for the genetic map constructed is listed in Supplementary Table S3. This linkage map was used for QTL mapping of the heading date (Supplementary Figure S2).

QTL mapping for RILs

The heading date of the RIL population was used as phenotype data combined with a high-quality bin marker map to proceed with QTL mapping by IciMapping software with a threshold Likelihood of Odd (LOD) value of 2.5. Two models were applied in this study: inclusive composite interval mapping with an additive effect (ICIM-ADD) and inclusive composite interval mapping with an epistasis effect (ICIM-EPI).

Molecular marker-assisted selection for breeding improvement

In developing RILs derived from WR04-6 and SN265, individuals with the early heading and other elite characteristics were selected. We extracted these individuals' DNA by the CTAB method and designed primers by https://primer3.ut.ee/to detect the locus responsible for the early heading. Primers used in this study are displayed in Supplementary Table S6.

Haplotype network analysis

In order to further reveal the domestication and evolution relationship of *Hd1* in different rice ecotypes, we used the sequencing data of 1,293 accessions to perform haplotype network analysis including five ecotypes, temperate *japonica*, tropical *japonica*, *japonica* intermediate, weedy rice, and wild rice [*Oryza rufipogon* (*Or-IIIa*)]. The sequence data of wild rice and weedy rice were from Sun et al. (2022), and the sequence of the *japonica* population was obtained from the database http://ricevarmap.ncpgr.cn/. The haplotype network was built based on 14 SNP variations of the *Hd1* Coding sequence (CDS) region with *Or-IIIa* as an outgroup by using the "Median Joining Network" approach implemented in Popart software (Version 1.7).

Results

High-latitude weedy rice heading earlier than that of cultivated rice and landrace

In high-latitude paddy ecosystems, weedy rice tends to heading earlier than cocultivated rice for rapid completion of reproduction. In the present study, we collected 274 accessions from three *japonica* rice ecotypes as a GWAS panel: the high-latitude modern *japonica* cultivar and its accompanied *japonica* weedy rice and *japonica* landrace. The heading date of this panel was investigated for three consecutive years (2019, 2020, and 2021) as the phenotype for GWAS. In three ecotypes, weedy rice (average 85 days) exhibited significantly earlier heading than cultivated rice (average 96 days) and landrace (average 106 days) (Figures 1A–C).

Dissecting the genetic basis of the weedy rice early heading by genome-wide association study

In the present *japonica* rice panel, the heading date exhibits significant variations. To find quantitative trait nucleotides

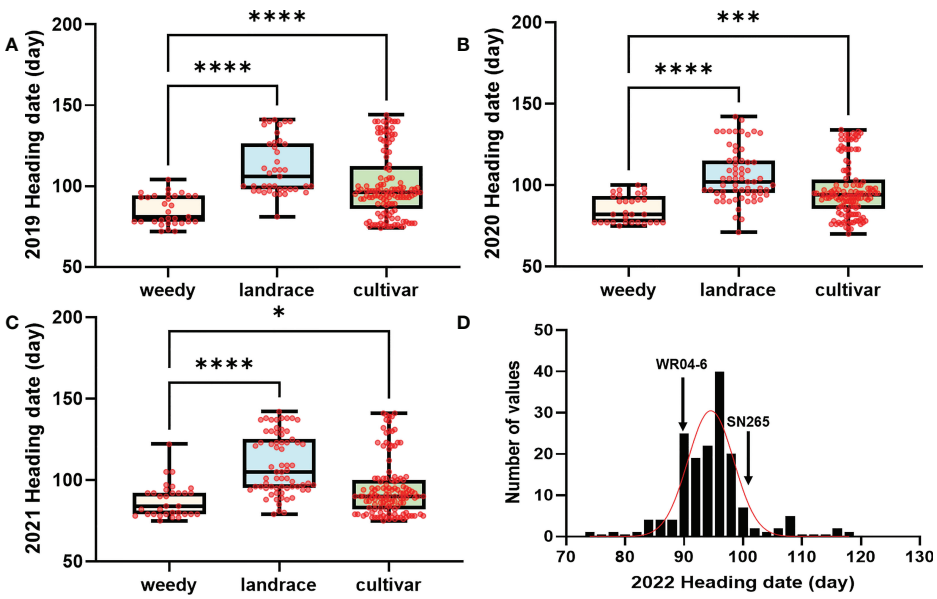


FIGURE 1
The heading date in the GWAS and RIL population. **(A–C)** The heading date of weedy rice, landrace, and *japonica* cultivar for 2019, 2020, and 2021 in the GWAS population. **(D)** Distribution of the RIL progenies of the weedy rice WR04-6 and *japonica* cultivar SN265. Bars are the maximum and minimum values of samples. Data were analyzed by one-way ANOVA multiple comparisons. ****, ***, and * represent significant differences at P = 0.0001, P = 0.001, and P = 0.05, respectively; red dots indicate samples. The D'Agostino–Pearson test method was used for the normal distribution test. weedy, weedy rice; landrace, japonica landrace; cultivar, modern japonica cultivar.

(QTNs) that are significantly associated with this phenotypic variation, the heading date was investigated for three consecutive years (2019, 2020, and 2021). The Manhattan plots of the GWAS (using a linear mixed model and the first two principal components of PCA as covariates) showed that the four strongest associated signals (Locus 1, Locus 2, Locus 3, and Locus 4) were stable and detected for at least 2 years (Figures 2A–C). The specific GWAS results and the locus distribution of ecotype information are displayed in Table 1.

Considering the following QTL analysis in this study and the background knowledge on the known heading date genes in rice, Locus 2 and Locus 3 on chromosome 6 were the main objects of interest for further analysis. We expanded the 2MB physical distance surrounding the lead SNP of Locus 2 and Locus 3 to find highly LD regions. We finally detected two LD blocks in Locus 2 and one LD block in Locus 3 that involved 361 and 85 genes, respectively (Figure 2G). Based on the candidate gene association analysis, we further narrowed down the candidate genes to 24 in Locus 2 and 38 in Locus 3 (Supplementary Table S4). On the other hand, three genes according to the annotation (http://rice.uga.edu/), *OsCCT20*, *OsCCT21* (*Hd1*), and *OsCCT22*, within the three LD blocks were considered as candidate causal genes due to all of them being CCT family members that were reported to have important roles in flowering regulation (Zhang et al., 2021) (Figures 2D–F). However, considering the candidate gene association analysis, different haplotypes of OsCCT20 did not show significant differences in the heading date. Finally, *OsCCT21* (*Hd1*) and *OsCCT22* were speculated to be the causal genes for responding to the natural variation of the heading date in Locus 2 and Locus 3 (Figures 3A–C).

QTL mapping for the heading date differences between WR04-6 and SN265

Japonica rice SN265 with a high yield advantage was recognized as a "super rice" by the Chinese Ministry of Agriculture, and its accompanied weedy rice (named WR04-6) showed 10–15 days earlier heading than SN265. The heading date showed continuous variation in the RIL population from 74 days to 123 days, but not following a normal distribution (k^2 = 30.19) (Figure 1D). We crossed SN265 with WR04-6 to develop an RIL population for detecting the genetic basis of these differences in the heading date by QTL mapping. We applied the ICIM-ADD model for additive-effects QTL mapping and found four QTLs above the threshold, as shown in Figure 4A; they were distributed on chromosomes 3, 6, 7, and 12 (Table 2). Among them, we noticed that three genes within the three QTL genomic regions were reported before: *Hd1*/*OsCCT21* (Chr6), *Dth7* (Chr7), and *Hd16* (Chr3). Among them, *Hd1* can explain the largest phenotypic variation (PVE = 35.9%); the phenotypic variance explained (PVE) of *Hd16* is 8.3% and that of *Dth7* is 6.1%.

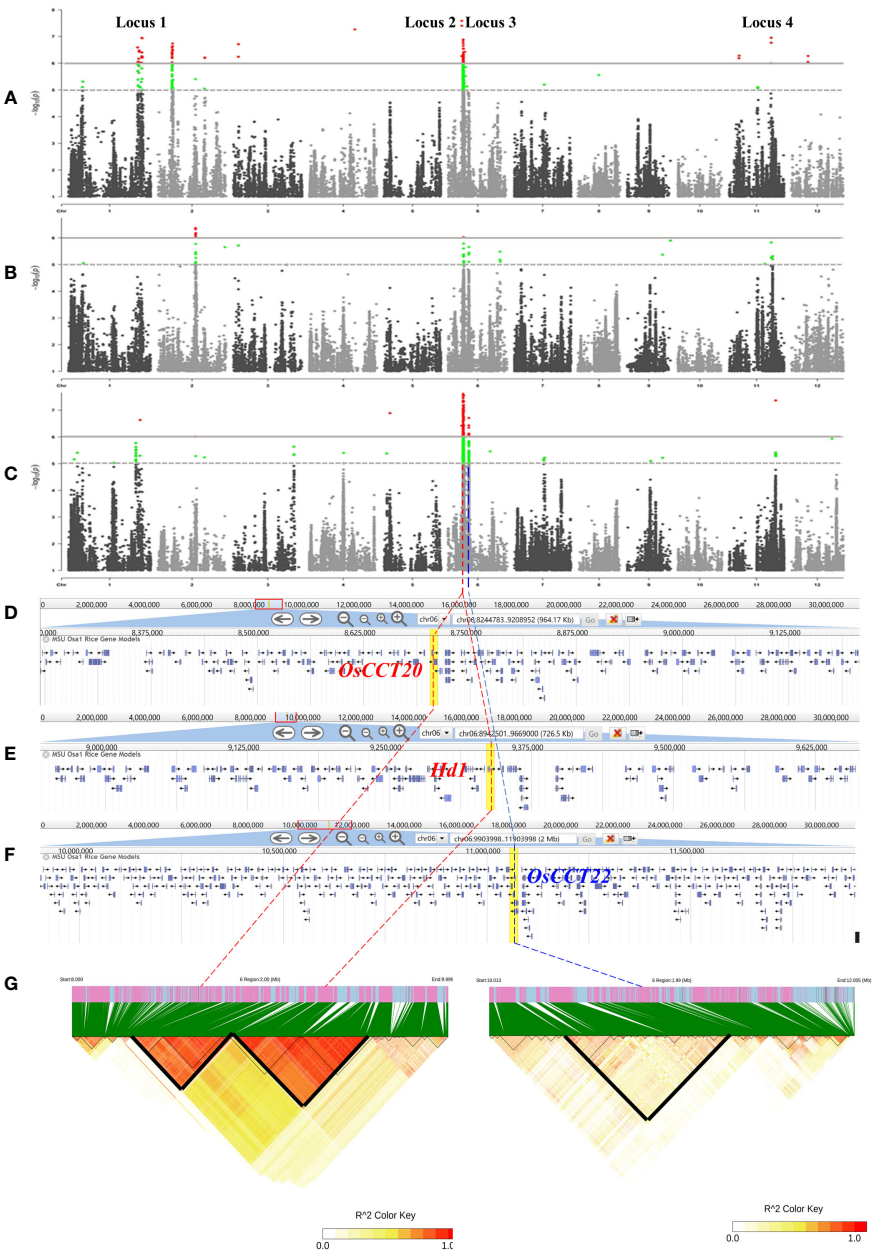


FIGURE 2
Results of the genome-wide association study. **(A–C)** Manhattan plot of 274 accessions in 2019, 2020, and 2021. Gray lines indicate threshold lines by 10^{-5} and 10^{-6}; red and green dots mean SNPs above the threshold line. **(D–F)** Annotation of genes that are in highly linkage disequilibrium block **(G)** linkage disequilibrium block analysis of 2MB region surrounding the lead SNPs, highly linkage disequilibrium blocks are displayed in bold lines.

Moreover, the additive effect of *Hd1* is 6.3111, while the additive effect of *Hd16* and *Dth7* is -2.1566 and -1.83, respectively. We further analyzed the genotype of the three genes between the two parents, weedy rice WR04-6 and *japonica* cultivar SN265, to confirm the specific variants (Figures 5A–C). We found that the genotype of *Hd1* in WR04-6 in the coding region is different from that of SN265 and Nipponbare (reference genome). The changed base located in 134 coding regions in the zinc finger domain of *Hd1* caused a non-synonymous substitution from valine to glycine, which may cause the functional variation of *Hd1* between SN265 and WR04-6 (Figure 5A). Considering the GWAS results of the heading date in the *japonica* panel of the present study, we have reason to believe that *OsCCT21 (Hd1)* is the common cause of the early-heading feature in weedy rice.

How *Hd1* of weedy rice evolves is another important scientific issue. Therefore, we established the genetic

TABLE 1 Results of the genome-wide association analysis.

Locus	Chr	Lead SNP (bp)	-log$_{10}$(*P*)	Year	Variation	Weedy	Landrace	Cultivar
Locus 1	1	35,416,622	5.77	2021	T/C	0.455	0.689	0.903
	1	36,347,597	6.59	2019	G/GA	1.000	0.770	0.903
	2	7,738,314	6.73	2019	T/A	0.969	0.851	0.968
	2	19,840,643	6.32	2020	G/A	0.000	0.172	0.084
Locus 2	6	8,621,625	7.4	2019	A/G	0.879	0.598	0.617
	6	8,718,748	8.1	2021	A/G	0.879	0.621	0.643
	6	8,721,885	6.02	2020	A/C	0.121	0.299	0.425
Locus 3	6	10,973,241	4.87	2019	C/T	0.849	0.736	0.636
	6	10,962,053	5.66	2020	T/A	0.819	0.759	0.630
	6	10,958,625	6.71	2021	A/G	0.030	0.770	0.266
Locus 4	11	22,255,595	6.95	2019	G/A	0.969	0.805	0.864
	11	22,305,044	5.83	2020	A/G	1.000	0.851	0.877
	11	24,600,828	7.40	2021	ATG/A	0.848	0.471	0.623

The right three columns represent the distribution of the lead SNP variant in the weedy, landrace, and modern cultivar.
Chr, Chromosome; weedy, weedy rice; landrace, *japonica* landrace; cultivar, modern *japonica* cultivar.

relationship between weedy rice and other ecotypes of *japonica* through the haplotype network. As shown in Supplementary Figure S4, weedy rice has four haplotypes, *Hd1*:Hap2, Hap3, Hap6, and Hap7. Most of the weedy rice (76.9%) is Hap7, which is exclusive to weedy rice and derived from Hap4, the major haplotype shared by temperate *japonica*, tropical *japonica*, and *japonica* intermediate. This result implies that selection pressure on weedy rice resulted in its *Hd1* diverging from cultivated rice under long-term natural selection, which is further evidence of the genetic contribution of *Hd1* in the early heading of weedy rice.

The genotype of *Dth7* and *Hd16* in minor effect QTL genomic regions also showed functional variations between WR04-6 and SN265. *Dth7* has two non-synonymous substitutions in the WR04-6 coding region nearby the PR domain; although this haplotype of *Dth7* has been reported before, its function is still unknown (Gao et al., 2014; Li et al., 2015; Li et al., 2018). A non-synonymous substitution in the WR04-6 S/K kinase domain caused the loss of function of *Hd16* (Kwon et al., 2014) (Figures 5B, C). The other QTLs located on chromosome 12 were new unknown QTLs for the rice heading date reported by this study. In addition, we found three gene

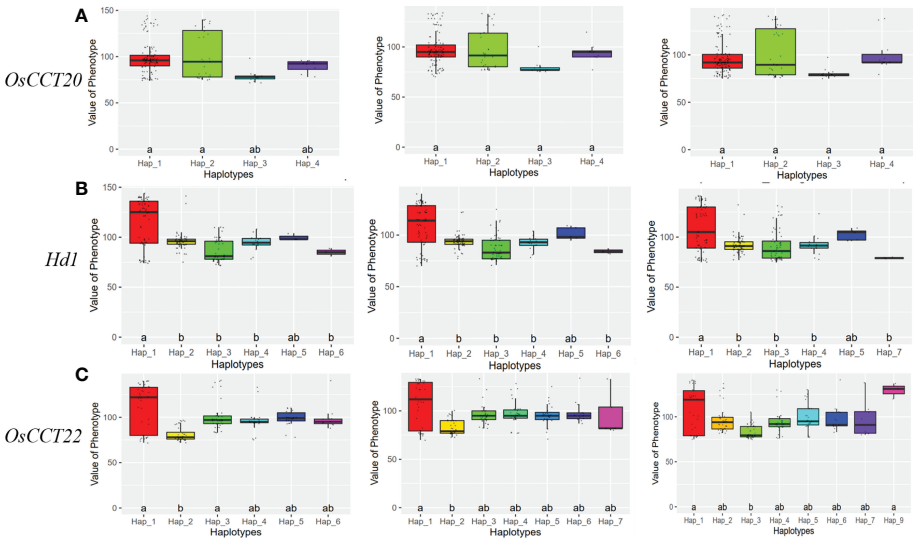


FIGURE 3
Boxplots of the heading date phenotype for different haplotypes of candidate genes. **(A)** A 3-year heading date box for the different haplotypes of *OsCCT20*. **(B)** A 3-year heading date box for the different haplotypes of *Hd1*. **(C)** A 3-year heading date box for the different haplotypes of *OsCCT22*. Duncan's test at p ranks the phenotypic differences. The letter a and b are ranked by Duncan's test at p<0.05.

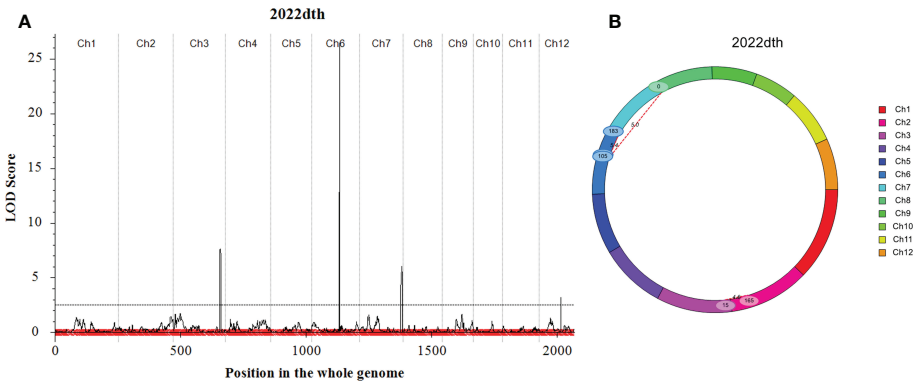


FIGURE 4
QTL mapping results of the heading date. **(A)** Additive effect for the heading date calculated by the ICIM-ADD model. **(B)** The epistatic effect for the heading date calculated by the ICIM-EPI model. The threshold line was set with an LOD score ≥ 2.5.

pairs that may regulate the heading date by epistasis effect interaction based on the ICIM-EPI (Figure 4B). Interestingly, one epistasis gene pair occurred between *Hd1* and another locus; the specific details are displayed in Supplementary Table S5.

Breeding early-heading high-yield *Japonica* rice by applying early-heading genes from weedy rice

In developing the RIL population using the weedy rice WR04-6 and the super-high-yielding *japonica* rice SN265, we also performed genetic improvement practices of early-heading lines. From the F_3 to the F_8 generation, individuals with early heading and good comprehensive traits were selected as the selection targets in each generation. Finally, six genetically stable lines with different genetic backgrounds were successfully bred, and their heading dates ranged from 81 to 92 days, at least 10 days earlier than that of the *japonica* parent SN265. After detecting the genotype, we found that all of the fragments of Locus 2 and Locus 3 on chromosome 6 of these six lines were derived from the weedy rice parent WR04-6 (Supplementary Figure S3). We then cultivated the six lines in higher latitudes, Wuchang, Heilongjiang province (127°16′E, 44° 93′N), and confirmed that they could safely head. Thus, our practice of this genetic improvement using weedy rice as a genetic resource has successfully expanded the planting range of super *japonica* rice northward by about 400 km. We also developed molecular markers for this early-heading interval that combined with the early-heading genetic resources of weedy rice will greatly contribute to breeding the early-heading *japonica* rice.

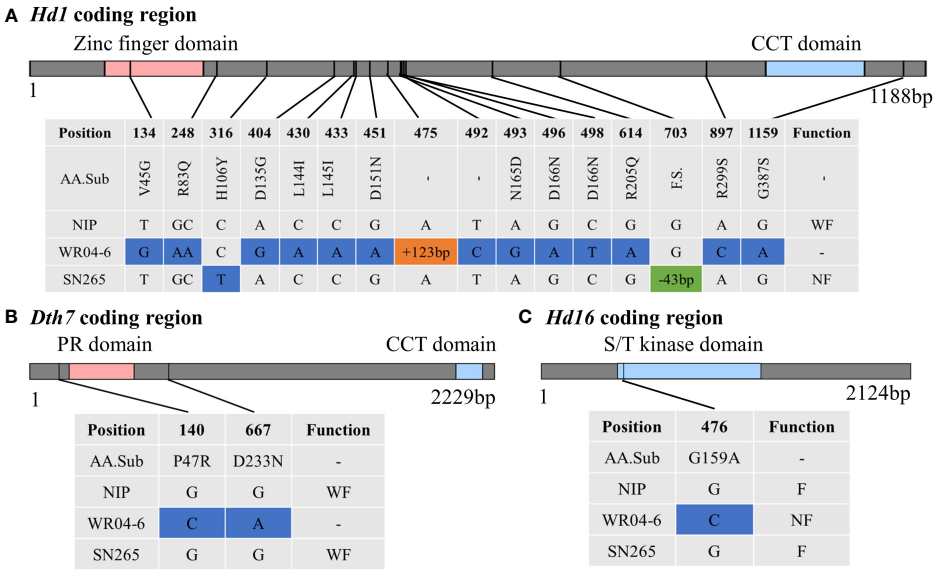
Discussion

Molecular genetic basis of early-heading characteristics in weedy rice

In high-latitude paddy ecosystems, weedy rice tends to head early than cocultivated rice for rapid completion of reproduction. In this study, we performed a GWAS and QTL mapping to reveal the genetic basis of the early-heading feature of weedy rice. Our research has found that *Hd1* and *OsCCT22* may be responsible for the heading date variations in the natural population by GWAS (Takahashi et al., 2009; Zhang et al., 2021). Furthermore, *Hd1*, *Dth7*, and *Hd16* were colocated using the ICIM-ADD model in the RIL population, and the three genes do have variations between parents. We also found the epistatic effect of *Hd1* for another QTL by the ICIM-EPI model. Considering the GWAS results with QTL mapping, we believe that *Hd1* is the common cause of the early-heading feature in weedy rice. Our findings provide a new

TABLE 2 RIL QTL mapping result by ICIM-ADD.

Trait Name	Chr	Position(cM)	L-Position(bp)	R-Position(bp)	LOD	PVE (%)	Add	Known Gene	Reference
2022dth	3	184.7	32,196,269	33,001,571	7.6207	8.316	-2.1566	*Hd16*	Hori et al., 2013
2022dth	6	110.5	9,676,469	9,035,697	26.4224	35.9071	6.3111	*Hd1*	Yano et al., 2000
2022dth	7	169	28,914,563	29,617,569	5.9928	6.1069	-1.8338	*Dth7*	Yan et al., 2013
2022dth	12	85.6	21,255,481	21,603,023	3.1934	3.1561	-1.3223	–	–



Position	134	248	316	404	430	433	451	475	492	493	496	498	614	703	897	1159	Function
AA.Sub	V45G	R83Q	H106Y	D135G	L144I	L145I	D151N	-	-	N165D	D166N	D166N	R205Q	F.S.	R299S	G387S	-
NIP	T	GC	C	A	C	C	G	A	T	A	G	C	G	G	A	G	WF
WR04-6	G	AA	C	G	A	A	A	+123bp	C	G	A	T	A	G	C	A	-
SN265	T	GC	T	A	C	C	G	A	T	A	G	C	G	-43bp	A	G	NF

Position	140	667	Function
AA.Sub	P47R	D233N	-
NIP	G	G	WF
WR04-6	C	A	-
SN265	G	G	WF

Position	476	Function
AA.Sub	G159A	-
NIP	G	F
WR04-6	C	NF
SN265	G	F

FIGURE 5
Genotype analysis between the weedy rice WR04-6 and *japonica* cultivar SN265. **(A)** The variation of *Hd1* between WR04-6 and SN265 with Nipponbare genotype as a reference in the coding region. **(B)** The variation of *Dth7* between WR04-6 and SN265 with Nipponbare genotype as a reference in the coding region. **(C)** The variation of *Hd16* between WR04-6 and SN265 with Nipponbare genotype as a reference in the coding region. The blue box indicates a single base variation; the orange and green boxes represent insertion and deletion, respectively. F, Functional; WF, Weak functional; NF, Non-functional. The gray top bar represents the coding region; the colorful region means the domain of the gene.

perspective for studying the molecular genetic mechanisms and regulatory networks governing the heading date in rice.

Hd1 contains the B-box zinc finger domain and CCT domain, which is the first cloned gene that can regulate the heading in rice from Nipponbare (Yano et al., 1997); it may be a selection target in the process of domestication of flowering diversity of cultivated rice (Takahashi and Shimamoto, 2011). In our study, we found a new haplotype of *Hd1* in WR04-6, which contains a new non-synonymous substitution in the B-box binding domain. Based on available data, we speculate that this new haplotype is functional because the conserved CCT domain has no variation in WR04-6, and the additive effect of *Hd1* is 6.311 in the RIL population, which means that the early-heading characteristic comes from the weedy rice WR04-6 (Takahashi et al., 2009). In addition, this substitution in the B-box binding domain may change the protein interaction and affect the other aspect of rice growth. More importantly, this single base substitution variation only exists in weedy rice and is not detected in other accessions, but we did not validate the activity of the *Hd1* protein of weedy rice.

Variant genes in weedy rice change the complex regulatory network of the heading date

The flowering time of rice is regulated by plenty of genes; different combinations of these allelic genes determine the rice adaptability and regional distribution (Ebana et al., 2011; Guo et al., 2013; Lee and An, 2015; Brambilla et al., 2017). *Hd1* has a basic function of promoting the expression of florigen genes *Hd3a*/*RFT1* and heading regardless of the day length (Zong et al., 2021). *Hd1* interacts with the *Ghd7* (*OsCCT26*) CCT domain to form a complex that represses the rice heading. The *Dth8*/*Hd1* complex binds to the promoter of *Ghd7* (*OsCCT26*) to form a ternary complex (Lin et al., 2003; Nemoto et al., 2016; Du et al., 2017; Wang et al., 2019; Zhang B. et al., 2019; Zhang Z. et al., 2019; Zong et al., 2021). *Dth7* (*OsCCT28*) is a major gene regulating the flowering time in the northeast of China, which also regulates the plant height and grains per panicle. *Dth7* acts downstream of phyB, inhibiting the expression of *Ehd1* (upstream regulator of rice florigen genes *Hd3a* and *RFT1*), thereby delaying flowering under day length (Liu et al., 2013; Gao et al., 2014; Liu et al., 2021). We speculate that the *Dth7* haplotype found in WR04-6 is non-functional because the additive effect of *Dth7* is -1.8338 in the RIL population (Table 2); the non-functional *Dth7* leads to the early heading of weedy rice by reduced suppression of *Ehd1* and promotes the expression of florigen genes *Hd3a*/*RFT1* in the shoot apical meristem. The non-functional *Dth7* reduced the competition for *Hd1*, which means that *Hd1* will play a greater role in promoting the heading under long day-length conditions. *Hd16* can phosphorylate *Ghd7* (*OsCCT26*), which contributes to the later heading in *japonica* rice under LD conditions. *Hd16* can also interact with and phosphorylate *Dth7*, although the underlying mechanism is still unknown (Kwon et al., 2015).

The genetic interactions between *Ghd7* and *Dth7* and between *Dth7* and *Hd16* are also involved in the heading under LD conditions (Shibaya et al., 2011). The non-functional *Hd16* in weedy rice will reduce the expression of *Dth7* and *Ghd7*, thereby enhancing the expression of *Ehd1* and promoting the heading. Variations found in weedy rice change the complex regulatory network of the heading date and other aspects of rice growth. Based on the above research, we can conclude that the early heading of weedy rice results from natural selection and competition with cultivated rice; however, how selections shape the early-heading loci in the weedy rice genome is still unknown. We uncovered this scientific question through the GWAS and QTL mapping in this study.

Weedy rice as a gene resource for improving cultivated rice

As an important ecological characteristic, the heading date plays a decisive role in the rice breeding practice. It affects cultivation environments, yield, rice quality, nutritional value, and other traits; suitable heading dates will maximize the use of light, heat, and other ecological resources in the local ecological environment (Brambilla et al., 2017). The early heading is a necessary feature for weedy rice to survive. Under long-term natural selection, the elite dominant allele of the early heading in weedy rice has been fixed. It is easy to reproduce naturally in the paddy field and breed offspring for weedy rice (Xia et al., 2011; Sun et al., 2013). Therefore, exploring early-heading gene resources of weedy rice is of great value for cultivating the early-heading cultivated rice.

Weedy rice can be an excellent genetic resource for its large number of elite characteristics, the most important is that weedy rice has no reproductive isolation with cultivars. Thus, these elite genes can be applied in the breeding practice and improvement programs for cultivars by Molecular marker-assisted selection (MAS). The genetic diversity of cultivated rice has decreased yearly due to the single breeding objective, so the genetic resource from weedy rice is very precious (Ebana et al., 2011). Applying the early-heading genes from weedy rice makes more elite cultivated varieties that can be planted in more expansive areas (Imaizumi et al., 2021). In the present study, we also developed several molecular markers to assist the selection of the early-heading varieties and expand the rice-suitable cultivation area northward.

Besides the early-heading characteristics, weedy rice still has many other fine biological characteristics that can be used in breeding. All these make weedy rice more competitive than the cultivated rice. Weedy rice is a precious genetic material for cultivated rice that is reported to be a hidden gold mine in the paddy field (Tang et al., 2011; Qiu et al., 2017; Sun et al., 2019; Wu et al., 2022). Especially in high latitudes where wild rice is difficult to apply, weedy rice might be the best choice to expand the genetic diversity of *japonica* rice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

JS and WC designed the study. ZL, RG, JC, and XZ investigated the heading date trait. ZL and RG performed data statistical analysis. XY and JS performed GWAS analyses. ZL, XY, and CL carried out candidate gene association analysis. XmZ and PY performed searching candidate genes/QTLs. ZL wrote the paper. JS revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Liaoning Applied Basic Research Program (2022JH2/101300172), and the Support Program for Young Scientific and Technological Innovation Talents of Shenyang, China (No. RC210408).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1059197/full#supplementary-material

References

Brambilla, V., and Fornara, F. (2013). Molecular control of flowering in response to day length in rice. *J. Integr. Plant Biol.* 55 (5), 410–418. doi: 10.1111/jipb.12033

Brambilla, V., Gomez-Ariza, J., Cerise, M., and Fornara, F. (2017). The importance of being on time: Regulatory networks controlling photoperiodic flowering in cereals. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00665

Brar, D. S., and Khush, G. S. (2018). "Wild relatives of rice: A valuable genetic resource for genomics and breeding research," in *The wild oryza genomes*. Eds. T. K. Mondal and R. J. Henry (Cham: Springer International Publishing), 1–25.

Chen, R., Deng, Y., Ding, Y., Guo, J., Qiu, J., Wang, B., et al. (2022). Rice functional genomics: decades' efforts and roads ahead. *Sci. China Life. Sci.* 65 (1), 33–92. doi: 10.1007/s11427-021-2024-0

Du, A., Tian, W., Wei, M., Yan, W., He, H., Zhou, D., et al. (2017). The DTH8-Hd1 module mediates day-Length-Dependent regulation of rice flowering. *Mol. Plant* 10 (7), 948–961. doi: 10.1016/j.molp.2017.05.006

Ebana, K., Shibaya, T., Wu, J., Matsubara, K., Kanamori, H., Yamane, H., et al. (2011). Uncovering of major genetic factors generating naturally occurring variation in heading date among Asian rice cultivars. *Theor. Appl. Genet.* 122 (6), 1199–1210. doi: 10.1007/s00122-010-1524-1

Gao, H., Jin, M., Zheng, X. M., Chen, J., Yuan, D., Xin, Y., et al. (2014). Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc. Natl. Acad. Sci. U. S. A.* 111 (46), 16337–16342. doi: 10.1073/pnas.1418204111

Guo, Z., Yang, Q., Huang, F., Zheng, H., Sang, Z., Xu, Y., et al. (2021). Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. *Plant Commun.* 2 (6), 100230. doi: 10.1016/j.xplc.2021.100230

Guo, L., Zhang, Z.-h., and Zhuang, J.-y. (2013). Quantitative trait loci for heading date and their relationship with genetic control of yield traits in rice (Oryza sativa). *Rice Sci.* 20 (1), 1–12. doi: 10.1016/s1672-6308(13)60101-8

Hori, K., Nonoue, Y., Ono, N., Shibaya, T., Ebana, K., Matsubara, K., et al. (2015). Genetic architecture of variation in heading date among Asian rice accessions. *BMC Plant Biol.* 15 (1), 115. doi: 10.1186/s12870-015-0501-x

Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., and Yano, M. (2013). Hd16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J.* 76 (1), 36–46. doi: 10.1111/tpj.12268

Imaizumi, T., Ebana, K., Kawahara, Y., Muto, C., Kobayashi, H., Koarai, A., et al. (2021). Genomic divergence during feralization reveals both conserved and distinct mechanisms of parallel weediness evolution. *Commun. Biol.* 4 (1), 952. doi: 10.1038/s42003-021-02484-5

Kwon, C. T., Koo, B. H., Kim, D., Yoo, S. C., and Paek, N. C. (2015). Casein kinases I and 2α phosphorylate oryza sativa pseudo-response regulator 37 (OsPRR37) in photoperiodic flowering in rice. *Mol. Cells* 38 (1), 81–88. doi: 10.14348/molcells.2015.2254

Kwon, C. T., Yoo, S. C., Koo, B. H., Cho, S. H., Park, J. W., Zhang, Z., et al. (2014). Natural variation in early flowering1 contributes to early flowering in *japonica* rice under long days. *Plant Cell. Environ.* 37 (1), 101–112. doi: 10.1111/pce.12134

Lee, Y. S., and An, G. (2015). Complex regulatory networks of flowering time in rice. *J. Rice Res.* 3, 141. doi: 10.4172/2375-4338.1000141

Li, X., Liu, H., Wang, M., Liu, H., Tian, X., Zhou, W., et al. (2015). Combinations of Hd2 and Hd4 genes determine rice adaptability to heilongjiang province, northern limit of China. *J. Integr. Plant Biol.* 57 (8), 698–707. doi: 10.1111/jipb.12326

Lin, H., Liang, Z. W., Sasaki, T., and Yano, M. (2003). Fine mapping and characterization of quantitative trait loci Hd4 and Hd5 controlling heading date in rice. *Breed. Sci.* 53 (1), 51–59. doi: 10.1270/jsbbs.53.51

Li, X., Shi, Z., Qie, Q., Gao, J., Wang, X., and Han, Y. (2020). CandiHap: a toolkit for haplotype analysis for sequence of samples and fast identification of candidate causal gene(s) in genome-wide association study. *BioRxiv*, 967539. doi: 10.1101/2020.02.27.967539

Li, X., Sun, Y., Tian, X., Ren, Y., Tang, J., Wang, Z., et al. (2018). Comprehensive identification of major flowering time genes and their combinations, which determined rice distribution in northeast China. *Plant Growth Regul.* 84 (3), 593–602. doi: 10.1007/s10725-017-0364-2

Liu, X., Liu, H., Zhang, Y., He, M., Li, R., Meng, W., et al. (2021). Fine-tuning flowering time *via* genome editing of upstream open reading frames of heading date 2 in rice. *Rice* 14 (1), 59. doi: 10.1186/s12284-021-00504-w

Liu, T., Liu, H., Zhang, H., and Xing, Y. (2013). Validation and characterization of Ghd7.1, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (Oryza sativa l.). *J. Integr. Plant Biol.* 55 (10), 917–927. doi: 10.1111/jipb.12070

Mao, T., Li, X., Jiang, S.-k., Tang, L., Wang, J.-y., Xu, H., et al. (2017). Discussion on strategy of grain quality improvement for super high yielding *japonica* rice in northeast China. *J. Integ. Agr.* 16 (5), 1075–1083. doi: 10.1016/S2095-3119(16)61563-0

Nemoto, Y., Nonoue, Y., Yano, M., and Izawa, T. (2016). Hd1, a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7. *Plant J.* 86 (3), 221–233. doi: 10.1111/tpj.13168

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T., et al. (2016). Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. *Theor. Appl. Genet.* 129 (4), 805–817. doi: 10.1007/s00122-016-2667-5

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nat. Commun.* 8 (1). doi: 10.1038/ncomms15323

Shibaya, T., Nonoue, Y., Ono, N., Yamanouchi, U., Hori, K., and Yano, M. (2011). Genetic interactions involved in the inhibition of heading by heading date QTL, Hd2 in rice under long-day conditions. *Theor. Appl. Genet.* 123 (7), 1133–1143. doi: 10.1007/s00122-011-1654-0

Shrestha, R., Gómez-Ariza, J., Brambilla, V., and Fornara, F. (2014). Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals. *Ann. Bot.* 114 (7), 1445–1458. doi: 10.1093/aob/mcu032

Sun, J., Ma, D., Tang, L., Zhao, M., Zhang, G., Wang, W., et al. (2019). Population genomic analysis and *De novo* assembly reveal the origin of weedy rice as an evolutionary game. *Mol. Plant* 12 (5), 632–647. doi: 10.1016/j.molp.2019.01.019

Sun, J., Qian, Q., Ma, D.-R., Xu, Z.-J., Liu, D., Du, H.-B., et al. (2013). Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New. Phytol.* 197 (1), 290–299. doi: 10.1111/nph.12012

Sun, J., Zhang, G., Cui, Z., Kong, X., Yu, X., Gui, R., et al. (2022). Regain flood adaptation in rice through a 14-3-3 protein OsGF14h. *Nat. Commun.* 13 (1), 5664. doi: 10.1038/s41467-022-33320-x

Takahashi, Y., and Shimamoto, K. (2011). Heading date 1 (Hd1), an ortholog of arabidopsis CONSTANS, is a possible target of human selection during domestication to diversify flowering times of cultivated rice. *Genes. Genet. Syst.* 86 (3), 175–182. doi: 10.1266/ggs.86.175

Takahashi, Y., Teshima, K. M., Yokoi, S., Innan, H., and Shimamoto, K. (2009). Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. *Proc. Natl. Acad. Sci. U. S. A.* 106 (11), 4555–4560. doi: 10.1073/pnas.0812092106

Tang, L., Ma, D. R., Xu, Z. J., Deng, H. F., Chen, W. F., and Yuan, L. P. (2011). Utilization of weedy rice for development of *japonica* hybrid rice (Oryza sativa l.). *Plant Sci.* 180 (5), 733–740. doi: 10.1016/j.plantsci.2011.02.002

Wang, P., Gong, R., Yang, Y., and Yu, S. (2019). Ghd8 controls rice photoperiod sensitivity by forming a complex that interacts with Ghd7. *BMC Plant Biol.* 19 (1), 462. doi: 10.1186/s12870-019-2053-y

Wang, J., Jiang, T., Zou, D., Zhao, H., Li, Q., Liu, H., et al. (2014). Genetic diversity and genetic relationships of *japonica* rice varieties in northeast Asia based on SSR markers. *Biotechnol. Biotechnol. Equip.* 28 (2), 230–237. doi: 10.1080/13102818.2014.908019

Wu, D., Qiu, J., Sun, J., Song, B. K., Olsen, K. M., and Fan, L. (2022). Weedy rice, a hidden gold mine in the paddy field. *Mol. Plant* 15 (4), 566–568. doi: 10.1016/j.molp.2022.01.008

Xia, H. B., Xia, H., Ellstrand, N. C., Yang, C., and Lu, B. R. (2011). Rapid evolutionary divergence and ecotypic diversification of germination behavior in weedy rice populations. *New. Phytol.* 191 (4), 1119–1127. doi: 10.1111/j.1469-8137.2011.03766.x

Xie, Y., Shen, R., Chen, L., and Liu, Y. G. (2019). Molecular mechanisms of hybrid sterility in rice. *Sci. China Life. Sci.* 62 (6), 737–743. doi: 10.1007/s11427-019-9531-7

Yan, W., Liu, H., Zhou, X., Li, Q., Zhang, J., Lu, L., et al. (2013). Natural variation in Ghd7.1 plays an important role in grain yield and adaptation in rice. *Cell. Res.* 23 (7), 969–971. doi: 10.1038/cr.2013.43

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y., and Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theor. Appl. Genet.* 95 (7), 1025–1032. doi: 10.1007/s001220050658

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., et al. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. *Plant Cell* 12 (12), 2473–2484. doi: 10.1105/tpc.12.12.2473

Zhang, J., Fan, X., Hu, Y., Zhou, X., He, Q., Liang, L., et al. (2021). Global analysis of CCT family knockout mutants identifies four genes involved in regulating heading date in rice. *J. Integr. Plant Biol.* 63 (5), 913–923. doi: 10.1111/jipb.13013

Zhang, B., Liu, H., Qi, F., Zhang, Z., Li, Q., Han, Z., et al. (2019). Genetic interactions among Ghd7, Ghd8, OsPRR37 and Hd1 contribute to Large variation in heading date in rice. *Rice* 12 (1), 48. doi: 10.1186/s12284-019-0314-x

Zhang, Z., Zhang, B., Qi, F., Wu, H., Li, Z., and Xing, Y. (2019). Hd1 function conversion in regulating heading is dependent on gene combinations of Ghd7, Ghd8, and Ghd7.1 under long-day conditions in rice. *Mol. Breed.* 39 (7). doi: 10.1007/s11032-019-1001-8

Zhao, C., Xu, W., Song, X., Dai, W., Dai, L., Zhang, Z., et al. (2018). Early flowering and rapid grain filling determine early maturity and escape from harvesting in weedy rice. *Pest. Manage. Sci.* 74 (2), 465–476. doi: 10.1002/ps.4730

Zong, W., Ren, D., Huang, M., Sun, K., Feng, J., Zhao, J., et al. (2021). Strong photoperiod sensitivity is controlled by cooperation and competition among Hd1, Ghd7 and DTH8 in rice heading. *New. Phytol.* 229 (3), 1635–1649. doi: 10.1111/nph.16946

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fpls.2022.1044420

OPEN ACCESS

EDITED BY
Muhammad Fazal Ijaz,
Sejong University, South Korea

REVIEWED BY
Guy Barker,
University of Warwick, United Kingdom
Darlon V. Lantican,
University of the Philippines Los Baños, Philippines
Debaleena Datta,
University of Engineering and Management, India
Parvathaneni Naga Srinivasu,
Prasad V. Potluri Siddhartha Institute of Technology, India

*CORRESPONDENCE
Tofazzal Islam
tofazzalislam@bsmrau.edu.bd

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 14 September 2022
ACCEPTED 28 November 2022
PUBLISHED 20 December 2022

CITATION
Islam T, Afroz N, Koh C, Hoque MN, Rahman MJ, Gupta DR, Mahmud NU, Nahid AA, Islam R, Bhowmik PK and Sharpe AG (2022) Whole-genome sequencing of a year-round fruiting jackfruit (*Artocarpus heterophyllus* Lam.) reveals high levels of single nucleotide variation.
Front. Plant Sci. 13:1044420.
doi: 10.3389/fpls.2022.1044420

Whole-genome sequencing of a year-round fruiting jackfruit (*Artocarpus heterophyllus* Lam.) reveals high levels of single nucleotide variation

Tofazzal Islam[1]*, Nadia Afroz[1], ChuShin Koh[2], M. Nazmul Hoque[3], Md. Jillur Rahman[4], Dipali Rani Gupta[1], Nur Uddin Mahmud[1], Abdullah Al Nahid[5], Rashedul Islam[6], Pankaj K. Bhowmik[7] and Andrew G. Sharpe[2]

[1]Institute of Biotechnology and Genetic Engineering (IBGE), Bangabandhu Sheikh Mujibur Rahman Agricultural University (BSMRAU), Gazipur, Bangladesh, [2]Global Institute for Food Security (GIFS), University of Saskatchewan, Saskatoon, SK, Canada, [3]Department of Gynecology, Obstetrics and Reproductive Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University (BSMRAU), Gazipur, Bangladesh, [4]Pomology Division, Horticultural Research Center, Bangladesh Agricultural Research Institute, Gazipur, Bangladesh, [5]Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet, Bangladesh, [6]Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada, [7]Cell Technologies and Trait Development, National Research Council of Canada, Saskatoon, SK, Canada

Jackfruit (*Artocarpus heterophyllus* Lam.) is the national fruit of Bangladesh and produces fruit in the summer season only. However, jackfruit is not commercially grown in Bangladesh because of an extremely high variation in fruit quality, short seasonal fruiting (June-August) and susceptibility to abiotic stresses. Conversely, a year-round high yielding (ca. 4-fold higher than the seasonal variety) jackfruit variety, BARI Kanthal-3 developed by the Bangladesh Agricultural Research Institute (BARI) derived from a wild accession found in Ramgarh of Chattogram Hiltracts of Bangladesh, provides fruits from September to June. This study aimed to generate a draft whole-genome sequence (WGS) of BARI Kanthal-3 to obtain molecular insights including genes associated with year-round fruiting trait of this important unique variety. The estimated genome size of BARI Kanthal-3 was 1.04-gigabase-pair (Gbp) with a heterozygosity rate of 1.62%. *De novo* assembly yielded a scaffolded 817.7 Mb genome while a reference-guided approach, yielded 843 Mb of genome sequence. The estimated GC content was 34.10%. Variant analysis revealed that BARI Kanthal-3 included 5.7 M (35%) and 10.4 M (65%) simple and heterozygous single nucleotide polymorphisms (SNPs), and about 90% of all these polymorphisms are in inter-genic regions. Through BUSCO assessment, 97.2% of the core genes were represented in the assembly with 1.3% and 1.5% either fragmented or missing, respectively. By comparing identified orthologous gene groups in BARI Kanthal-3 with five closely and one distantly related species of 10,092 common orthogroups were found across the genomes of the six species. The phylogenetic analysis of the shared

orthogroups showed that *A. heterophyllus* was the closest species to BARI Kanthal-3 and orthogroups related to flowering time were found to be more highly prevalent in BARI Kanthal-3 compared to the other *Arctocarpus* spp. The findings of this study will help better understanding the evolution, domestication, phylogenetic relationships, year-round fruiting of this highly nutritious fruit crop as well as providing a resource for molecular breeding.

KEYWORDS

BARI Kanthal-3, genome sequencing and assembly, heterozygosity, SCOS, SNPs, scaffolds, flowering gene orthologues

1 Introduction

Jackfruit (*Artocarpus heterophyllus* Lam), which belongs to the *Moraceae* family, has attracted the attention of food experts and technologists due to its nutritional health benefits (Zhang et al., 2022). It is a tropical evergreen tree, which produces the largest edible single fruit in the world (up to 50 kg/fruit) (Naik, 1949; Simmonds and Preedy, 2015; Lin et al., 2022). The place of origin of this fruit tree is still unclear, but it is widely grown in tropical countries including China, India, Malaysia, Thailand, Indonesia, Philippines and Bangladesh (Zhang et al., 2021b), and parts of central and eastern Africa, Florida (USA), Latin America and the Caribbean (Rahman et al., 1999; Sidhu, 2012; Sahu et al., 2020). Jackfruit (popularly known as '*Kanthal*') is the national fruit of Bangladesh. Its demand is increasing gradually due to its low price, high nutritious value, diversified uses and potential for commercial cultivation (Sidhu, 2012). The crop is commonly referred to as "poor man's food" due to its lower market price as well as high abundance in the summer season (Rahman et al., 1995; Sahu et al., 2020). Jackfruit flesh, the main edible portion, has a unique aroma and contains high levels of sugars (mainly sucrose, fructose and glucose), carboxylic acids, minerals, vitamins, and dietary fiber (Ong et al., 2006; Lin et al., 2022). The flesh is used as an ingredient in salads, or made into ice cream, jams, nectars, fruit bars, juices, chutney, cakes, jelly and fermented beverages (Zhang et al., 2021a). Notably, jackfruit seeds are rich in starch (60–80% based on dry matter), protein, vitamins and minerals, which may be boiled or roasted and eaten, or boiled and preserved in syrup like chestnuts (Anaya-Esparza et al., 2018).

Bangladesh is one of the largest producers of jackfruit and accounts for about 21% of total fruit production of the country, second only to Mango as the principal fruit crop. During 2019-20, Bangladesh produced 1.1 million tons of jackfruit covering 16,592 hectares area (Statistics., 2020). Despite its numerous advantages, jackfruit trees are not commercially grown as a crop because of an extremely high variation in fruit quality, which is due to its cross-pollinated nature, seed-mediated propagation, short seasonal fruiting and susceptibility to abiotic stresses (Sidhu, 2012). Therefore, the potential of this unique nutritious fruit crop has not yet been utilized in Bangladesh for ensuring food and nutritional security through commercial cultivation and industrial processing. Genetic improvement of existing germplasm to overcome these problems will accelerate jackfruit to become a commercial crop in Bangladesh (Dhar, 1998). A series of earlier studies evaluated the yield, quality and genetic diversity of jackfruit of Bangladesh but none of these studies are systematic and comprehensive (Hossain, 1996; Saha et al., 1996; Rahman et al., 2016) and the underlying molecular mechanisms of the trait diversity in jackfruit is largely unknown. The harvesting period of jackfruit is short (June-August) resulting in a large wastage of this fruit amounting to 20-30% of the crop or even more during some seasons. The Bangladesh Agricultural Research Institute (BARI) developed a year-round jackfruit variety namely, BARI Kanthal-3 in 2014. The number of fruits per plant per year ranges between 219-245 (average = 232), fruits are medium in size (averaging 5.43 kg each) and yield is 1,334.6 (ranges between 1165-1504.2) kg fruit/plant (Table 1). The ripened edible portion contains 35.06 mg/g ß- carotene and 23.6% of total soluble sugar (TSS) (Azad et al., 2007).

Whole-genome sequencing (WGS) provides complete coverage of the coding and noncoding regions of the genome (Galperin and Koonin, 2010), which allows a comprehensive assessment of the genome of any organism including those of plants (Chen et al., 2019). It provides a genetic foundation that enables a greater efficiency to identify genetic diversity at key genes that can be used to enhance, reduce or add certain features to a plant phenotype (Chen et al., 2019). Since the first WGS of the model plant, *Arabidopsis thaliana* in 2000, a large number of plants from diverse taxonomic groups have been sequenced, and genes responsible for various plant traits have been characterized and cloned (Gardner et al., 2016; Chen et al., 2019; Hübner et al., 2019). Recently, WGS of economically important plants and animals such as jute, *Corchorus* spp. (Islam et al., 2017), hilsa (*Tenualosa ilisha*) (Das et al., 2018), and goat (*Capra hircus*) (Siddiki et al., 2019) have created huge public interest in

TABLE 1 Tree and fruit characteristics of *Artocarpus heterophyllus* L. (Year-round vs seasonal jackfruit).

Parameters	Features/traits Year-round (BARI Kanthal-3)	Seasonal jackfruit	Reference (Seasonal jackfruit)
Average No. of fruits/plant	232	57	(Rahman et al., 2016)
Average fruit yield/plant/year (kg)	1,334	308	(Rahman et al., 2016)
Average fruit weight (kg)	5.43	5.8	(Rahman et al., 2016)
Harvesting period	September-June	June-August	(Goswami and Chacrabati, 2016)
Edible portion (%)	52.5	53.43	(Rahman et al., 2016)
pH	5.21	6.0	(Goswami and Chacrabati, 2016)
TSS (%)	23.6	19.87	(Rahman et al., 2016)
Vitamin C (mg/100g)	2.0	6.46	(Goswami and Chacrabati, 2016)
Total sugar (%)	21.74	14.95	(Goswami and Chacrabati, 2016)
Reducing sugar (%)	6.02	6.43	(Goswami and Chacrabati, 2016)
Nonreducing sugar (%)	15.72	8.52	(Goswami and Chacrabati, 2016)
Moisture content (%)	69.10	81.6	(Goswami and Chacrabati, 2016)
Dry matter	30.90	18.4	(Goswami and Chacrabati, 2016)

Bangladesh. Recent advances in genomic analyses have revealed large numbers of single nucleotide polymorphisms (SNPs) as the most common form of DNA sequence variation between alleles in several plant species (Morgil et al., 2020). Because of their high abundance, significant information content, when associated with genes, SNPs have gained the center stage as the principal markers of choice for molecular genetics studies. This includes their application in shortening the time of breeding new varieties in many crops through marker assisted selection (Mammadov et al., 2012; Morgil et al., 2020). SNPs have also been applied for several years to assess diversity in specific genes or genomic regions, revealing the phylogenetic relationships between species. However, the emergence of high throughput sequencing technologies allows the SNP-based genetic diversity studies to be carried out at scale and can be useful in conserving diversity in domesticated populations. Plant phylogenetic and evolutionary studies are conventionally based on variation that exist at genes, and hence the knowledge of SNPs in these regions is essential for this analysis (Lasky et al., 2012). It is also important to know the location of SNPs in the whole genome, because if a SNP is present in the coding or regulatory region of a gene, it can greatly affect the functional activity of the resulting protein, such as an enzyme in a biosynthetic pathway (Somerville and Koornneef, 2002) by affecting gene expression and transcriptional and translational promoter activities. Therefore, SNPs can often be responsible for phenotypic variations that exits between individuals and be utilized as selectable genetic markers for improving agronomic traits.

However until now, only a limited amount of genomic information has been made available for the genus of *A. heterophyllus* (Laricchia et al., 2018; Sahu et al., 2020; Lin et al., 2022). Although the development of a year-round fruiting variety BARI Kanthal-3 from a wild accession offers an opportunity for commercial cultivation and processing of the jackfruit, nothing is known about the underlying molecular mechanism of its year-round fruiting characteristics and other beneficial traits. Molecular understanding of the extremely high phenotypic variabilities in jackfruit would facilitate the future development of high yielding, year-round fruiting, biotic and abiotic stress (e.g., flood, saline, drought, pathogen and pest) tolerant jackfruit varieties through molecular breeding, which is essential for establishing a jackfruit-based processing industry in Bangladesh and elsewhere. We report here an annotated Whole-genome assembly of the year-round fruiting *A. heterophyllus* cv. BARI Kanthal-3 for first time. The results of the promising phenotypic characteristics of the BARI Kanthal-3 variety, together with both the *de novo* and reference-guided assemblies, the identified SNPs, and copy number variation in genes that impact flowering time sheds light on the genetic diversity that exists within the *A. heterophyllus* genome.

2 Materials and methods

2.1 Collection of phenotypic data

The mean values (n = 5) of the phenotypic and biochemical data on tree and fruit characteristics of *Artocarpus heterophyllus* (BARI Kanthal-3) were obtained using the standard protocols, and compared with the published data of seasonal jackfruit (Goswami and Chacrabati, 2016; Rahman et al., 2016).

2.2 Collection of leaf samples, genomic DNA extraction, WGS library construction, and sequencing

The year-round jackfruit cultivar 'BARI Kanthal-3', a superior variety of jackfruit of Bangladesh, was used for

genome sequencing. We collected fresh leaf samples from germplasm repository at BARI, Joydebpur, Bangladesh. The samples were identified as jackfruit by Professor Md. Abdul Baset Mia of Department of Crop Botany of Bangabandhu Sheikh Mujibur Rahman Agricultural University (BSMRAU) in Bangladesh. A voucher specimen was deposited in the herbarium of the Department of Crop Botany of BSMRAU with an accession No. VS (HM) 005/2021. The origin of this germplasm is from a wild accession found in Chattogram Hiltracts (Ramgarh) of Bangladesh. Genomic DNA was extracted from freshly harvested leaves using the QIAGEN DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA) following the manufacturer's protocol. The quality of the DNA was visually inspected by 1% agarose gel electrophoresis. The quantity of the DNA was assessed by a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instruction. Whole-genome sequencing (WGS) library preparation was performed using Nextera XT DNA library preparation kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's protocol. Briefly, after normalization, DNA samples were fragmented and tagged by tagmentation in a single-tube reaction (Hoque et al., 2019). The tagmented DNA was amplified through a limited-cycle PCR program using a unique combination of barcode primers, the Index 1 (i7), Index 2 (i5) and full adapter sequences required for cluster generation. Amplification was followed by a cleanup step that purified the library DNA, and removed small library fragments by using Agencourt AMPure XP beads (Beckman Coulter, Inc.). Finally, prepared libraries were loaded onto a reagent cartridge, clustered on the NextSeq 550 System, and paired-end sequencing (2×150 bp) was performed using the Illumina NextSeq 550 High-Output Kit on the NextSeq 550 desktop sequencer.

2.3 Retrieval of data from the GenBank

In addition to the WGS data of BARI Kanthal-3, we also collected the genomic data (WGS) of five related and one distantly related species viz., *Artocarpus heterophyllus* (family *Moraceae*) i.e., *A. heterophyllus, A. altilis, Morus notabilis, Arabidopsis thaliana* and *Ficus carica* from the AOCC ORCAE platform and the National Center for Biotechnology Information (NCBI) under GenBank accession numbers of CNGB CNP0000486 (https://bioinformatics.psb.ugent.be/orcae/aocc/overview/Arthe), CNGB CNP0000715 (https://bioinformatics.psb.ugent.be/orcae/aocc/overview/Artal), NCBI ASM41409v2 (https://www.ncbi.nlm.nih.gov/genome/?term=ASM41409v2), NCBI TAIR10.1 (https://www.ncbi.nlm.nih.gov/genome/?term=TAIR10.1) and, NCBI Bioproject PRJNA565858, respectively.

2.4 *De novo* genome assembly and annotation

The generated WGS data were filtered through Trimmomatic v0.38 (Bolger et al., 2014) with option "LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:50" parameters to remove Illumina adapter, known Illumina artifacts, phiX, and low-quality regions. The processed reads were assembled by SOAPdenovo2 v2.04 (Luo et al., 2012) with k-mer=39 and subsequently scaffolded using a reference guided approach by RAGTAG (Alonge et al., 2019) software with default parameters. GapCloser v1.12 (Luo et al., 2012) with default parameters ("-l 150 -t 32 -p 31") was utilized for gap closing using the pair-end data. The assembled genome was analyzed by EDTA (Ou et al., 2019) software to create a non-redundant transposable element database which was then used to soft-mask the genome sequence. Genome annotation was performed by using Braker2 (Brůna et al., 2021) software with coding-gene sequence evidence from *A. heterophyllus*.

2.5 Genome assembly validation

The scaffolded sequences were compared against the reference genome by nucmer v4.0.0rc1 (Kurtz et al., 2004) using the default parameters. The genome assembly completeness was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologues), v4.1.4 (Seppey et al., 2019) to evaluate the presence of conserved plant orthologs with the Embryophyta database 10 lineage.

2.6 Genome size estimation

The high-quality data (~50X depth) were provided to Jellyfish v2.2.6 (Marçais and Kingsford, 2011) with "-C -m 21 -s 5G –min-quality=25" parameters to generate k-mer (K=21) frequency distribution. The output histograms were analyzed using GenomeScope (Vurture et al., 2017) to estimate the genome size, heterozygosity level, error rates, and repeat fraction.

2.7 Variant calling

Processed Illumina data were aligned against the draft *A. heterophyllus* genome (Sahu et al., 2020) using Burrows-Wheeler Aligner (BWA) v 0.7.17 (Li and Durbin, 2009). The aligned reads mapped with a quality score 30 or greater were analyzed by samtools/bcftools (Li, 2011) to produce the raw variant calls. Simple single nucleotide polymorphisms (SNPs) were defined as SNPs with greater than 90% ALT allele frequency observed in at

least 30 high quality reads (MAPQ>30). Following variant calling, we used snpEff to annotate variants and predict their effects on genes using a custom database generated using the *A. heterophylus* genome and annotation from https://bioinformatics.psb.ugent.be/gdb/aocc/arthe/ (Sahu et al., 2020).

2.8 Orthologous gene analysis

Gene orthology and orthogroups were inferred by Orthofinder (Emms and Kelly, 2019) based on protein sequence similarity searches using Diamond (Buchfink et al., 2015) software. 306 genes involved in *A. thaliana* flowering-time gene networks were downloaded from the Flowering-Interactive Database (FLOR-ID) (Bouché et al., 2016). Gene expansion or contraction were assessed based on relative gene count within each orthogroup relative to *A. thaliana*.

3 Results and discussion

3.1 Source and phenotypic features of the BARI Kanthal-3 variety

To develop the variety of BARI Kanthal-3, germplasm was collected from locations all over the Bangladesh including in the Chattogram Hiltracts, such as Ramgarh of Khagrachari. In 2014, an accession of Ramgarh was certified for cultivation in Bangladesh with a varietal name of BARI Kanthal-3, representing a new and unique variety of jackfruit in Bangladesh that bear fruits for ten months of the calendar year (September to June) while the seasonal plant gives fruit only for three months (June to August) each year. This year-round fruiting variety produces more than 4-fold higher average number of fruits per plant and fruit yield per plant per year compared to seasonal jackfruit. A mature plant produces an average of 232 (range 219-245) fruits per plant yielding about 1,334.6 kg (1165 -1504.2 kg) fruit/plant/year (Table 1). The fruits of BARI Kanthal-3 were medium (average 5.43 kg each) and average yield was 133.2 t/ha/year. This variety was not affected by any sort of infectious pathogens or pests (data not shown). The tree is erect and medium bushy, and the pulp of the fruit is medium soft, slightly yellow, medium juicy, highly sweet and aromatic (Figure 1). The amounts of ß- carotene and total soluble solid (TSS) in fruits were 35.06 mg/g and 23.6%, respectively. The edible portion of the fruit was 52.5% (Table 1). A large body of literature has revealed that jackfruit is a rich source of carbohydrates, minerals, carboxylic acids, dietary fiber, vitamins and minerals and bioactive compounds (Azad et al., 2007; Khan et al., 2021). Clearly, BARI Kanthal-3 is an extremely high yielding (ca. 4-fold higher than average fruit yield of a seasonal jackfruit per year) and high quality jackfruit.

3.2 Genome sequencing, assembly and annotation

A total of 472 million raw Illumina pair-end sequencing reads from the extracted DNA of *A. heterophyllus* cv. BARI Kanthal-3 leaf tissues were generated. The genome size of BARI Kanthal-3 was estimated to be 1.04 Gb with a heterozygosity rate of 1.62% based on K-mer analysis of the short read data (Figure 2). The estimated size is similar to the recently reported 1.01Gb genome size of seasonal *A. heterophyllus*

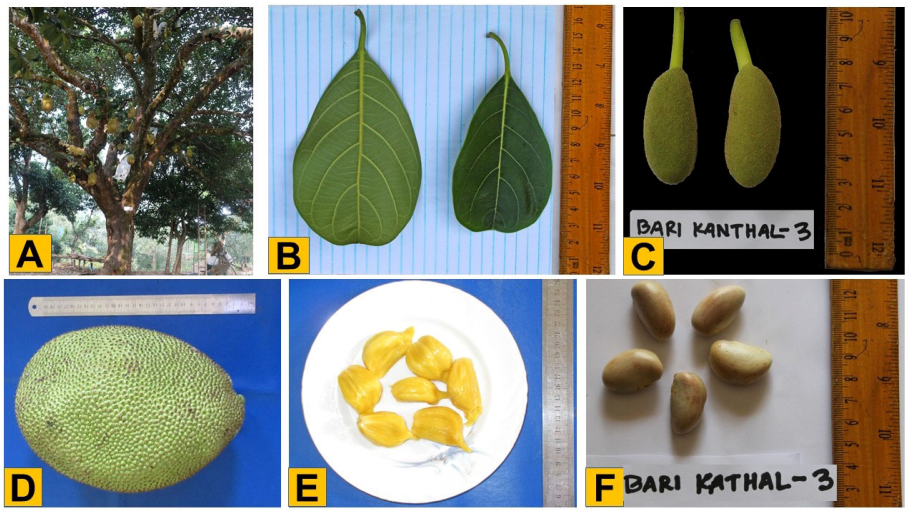


FIGURE 1
Phenotypic characteristics of BARI Kanthal-3. **(A)** Tree *in situ*, **(B)** Leaf, **(C)** Young fruits from the tree, **(D)** Mature fruit, **(E)** Kernel (flesh), and **(F)** Seed.

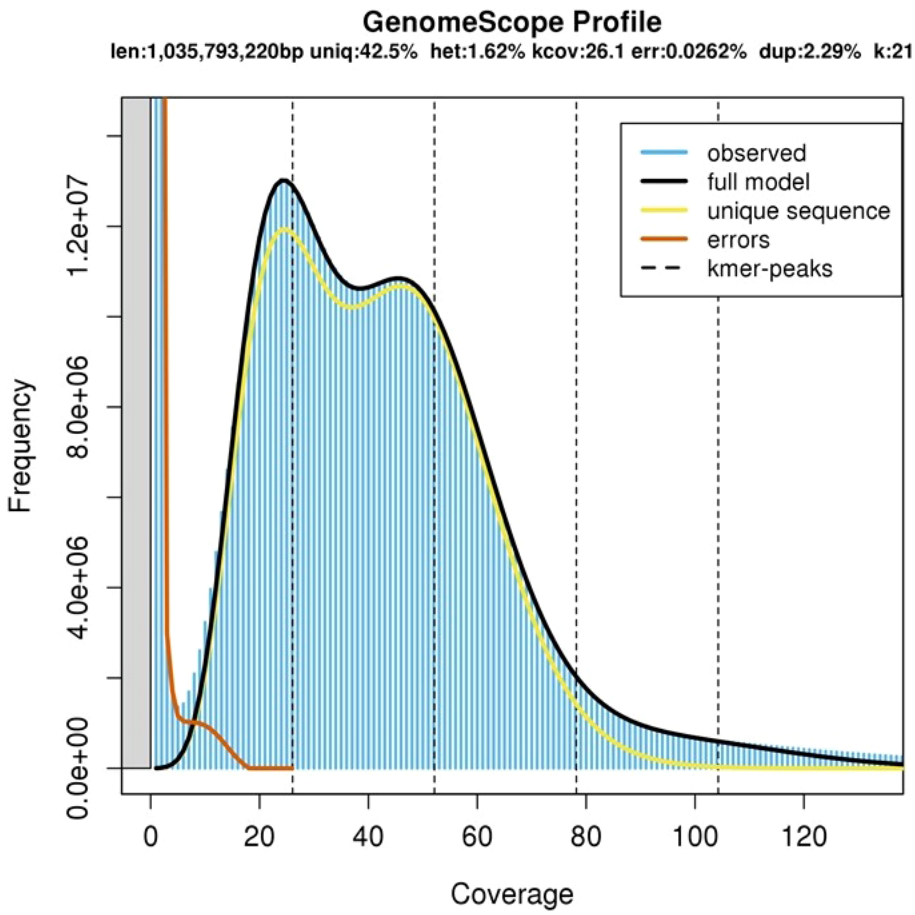


FIGURE 2
Genome size prediction of *Artocarpus heterophyllus* cv. BARI Kanthal-3. The X-axis represents the coverage of the genome while the Y-axis represents the frequency levels.

(Sahu et al., 2020), and is consistent with the c-value of 1.20 pg (Ohri and Kumar, 1986). BARI Kanthal-3 has a higher heterozygosity rate compared to the available reference genomes of seasonal jackfruit recently published, at 0.90 and 0.91, respectively (Sahu et al., 2020; Lin et al., 2022).

After quality filtering of the short reads using Trimmomatic, 439 million clean reads were obtained (Table 2). The high-quality reads assembled into different contigs using SOAPdenovo2, which ultimately yielded a base assembly of 1.36 M scaffolds, totaling 817.7 Mb. The N50s of scaffolds were 1.8 Kb (Table 2). The BARI Kanthal-3 contigs were then further scaffolded together using the reference guided approach using the existing published draft reference genome of jackfruit (Sahu et al., 2020) and the software RAGTAG, and finally gaps in the scaffolds were filled using the same pair-end Illumina data and Gapcloser software. In this case, SOAPdenovo2 + RAGTAG + GapCloser produced a base assembly of N50 size = 425 Kb in 218,562 scaffolds (Table 2). The mapping of paired-end short read

TABLE 2 Result of *de novo* and reference guided genome assembly of *Artocarpus heterophyllus* cv. BARI Kanthal-3.

Assembly	N50 number	N50 size (Kb)	N90 number	N90 size (Kb)	Longest Sequence (Kb)	Num Sequence	Total Bases (Mb)
SOAPdenovo2	88,402	1.8	720,032	0.2	65	1,364,052	817.7
SOAPdenovo2 + RAGTAG + Gapcloser	561	425	14,076	1.4	2,611	218,562	843

data against the *de novo* scaffolded assembly and out of 439 million read pairs, 400 million (91.1%) aligned to the assembly were properly paired indicating a high proportion of the original data is represented in the assembly. The GC content of BARI Kanthal-3 was 34.10% which is comparable to the GC content of a seasonal *A. heterophyllus* from Indian and Chinese origins that were recorded at 32.9% and 34.9%, respectively (Table S1) (Sahu et al., 2020; Lin et al., 2022).

We annotated 41,088 protein-coding genes in BARI-Kanthal-3 assembly using the Braker2 gene prediction pipeline. The gene number is comparable to the two existing *A. heterophyllus* genome assemblies with 35,858 (Sahu et al., 2020) and 41,997 (Lin et al., 2022) annotated genes, respectively.

3.3 Genome assembly validation

To assess the representation of a complete conserved core gene set in the BARI Kanthal-3, assembly, an analysis was carried out to assess the quality and completeness of the draft genome using the Benchmarking Universal Single-Copy Orthologs (BUSCO) datasets and an orthologue data base (Figure 3). We identified 1,614 single copy orthologs (SCOs). Among these SCOs, 97.2% (1569/1614) were complete (single copy = 1094, and duplicated = 475), whereas 21 and 24 were fragmented and missing, respectively (Figure 3). Our present findings aligned with the recently reported findings of Sahu et al. (2020), who reported that out of 1440 BUSCO ortholog groups searched in the *A. heterophyllus* assembly, 95% (1369/1440) were complete BUSCOs, 932 (64.7%) were "complete single-copy", 437 (30.3%) were "complete duplicated", and 15 (1%) were "fragmented", and 56 (4%) were "missing".

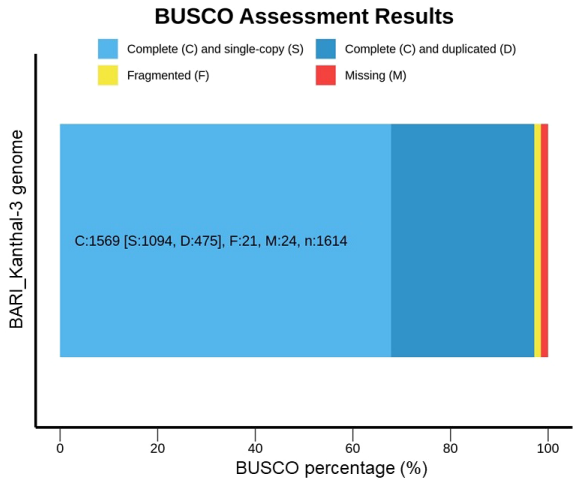


FIGURE 3
BUSCO (v.4.0.4) assessment results of BARI Kanthal-3. BUSCO completeness in *Artocarpus heterophyllus* cv. BARI Kanthal-3 is high indicating that the genome assemblies are of high gene space completeness.

We then carried out an orthologous gene analysis across the BARI Kanthal-3 genome and the genomes of five other species (*A. heterophyllus*, *A. altilis*, *F. carica*, *M. notabilis*, *A. thaliana*) based upon a protein similarity search using Diamond software and the inference of gene orthology and orthogroups using Orthofinder software. In this study, 10,924 orthogroups were found to be shared across the genomes of the six species (Figure 4A) with 399 found to be unique for BARI Kanthal-3 genome. Phylogenetic analysis was performed across the 10,924 shared orthogroups using the rooted gene tree method in Orthofinder (Figure 4B). The analysis revealed that the two genomes of *A. heterophylus* (BARI Kanthal-3 and *A. heterophyllus*) clustered more closely related to the other three *Moraceae* genomes with *A. thaliana* as an outgroup in the phylogenetic tree (Figure 4B). Finally, to assess if there was a relationship between differences in flowering time between BARI Kanthal-3 and seasonal *A. heterophyllus* and differences in the representation of orthogroups related to flowering time between them, we identified the relative numbers of these orthogroups in each of the *Moraceae* genomes compared to Arabidopsis (Figure 4C and Table S2). This identified large scale expansion in the *Artocarpus* spp. likely reflecting the whole genome duplication event previously identified in their evolution (Lin et al., 2022). However, a higher number of duplications were identified in the BARI Kanthal-3 genome compared to the other *Artocarpus* spp. indicating the longer season nature of the accession could be explained by an overall expansion of these gene families, although the more fractured nature of the genome assembly could also be impacting this observation. The availability of more contiguous long read genome assemblies for BARI Kanthal-3 and other *Artocarpus* spp. would enable this observation to be confirmed.

3.4 Variant analysis

The processed WGS reads were aligned against the *A. heterophyllus* draft assembly. Out of a total of 439 M reads, 417 M (95.1%) were found to be aligned in exact pairs. A total of 16 million single-nucleotide polymorphisms (SNPs) were called from the dataset including 5.7 M (35.0%) simple and 10.4 M (65.0%) heterozygous SNPs (Table 3 and Figure S1). Approximately, 90% of all polymorphisms are located in intergenic regions. In this study, 144,787 (2.5%) and 426,997 (7.5%) of the simple SNPs, and 250,715 (2.4%) and 739,288 (7.1%) of the heterozygous SNPs, were found in the exons and introns, respectively (Table 3). We further predicted the effects of variants on genes. As expected, large fraction of the variants was in the intergenic (64.5%), intronic (5%) and up/down-stream regions (29%) of the genes (Figure 5A). There are 232,587 missense mutations and 4,750 gained stop codons

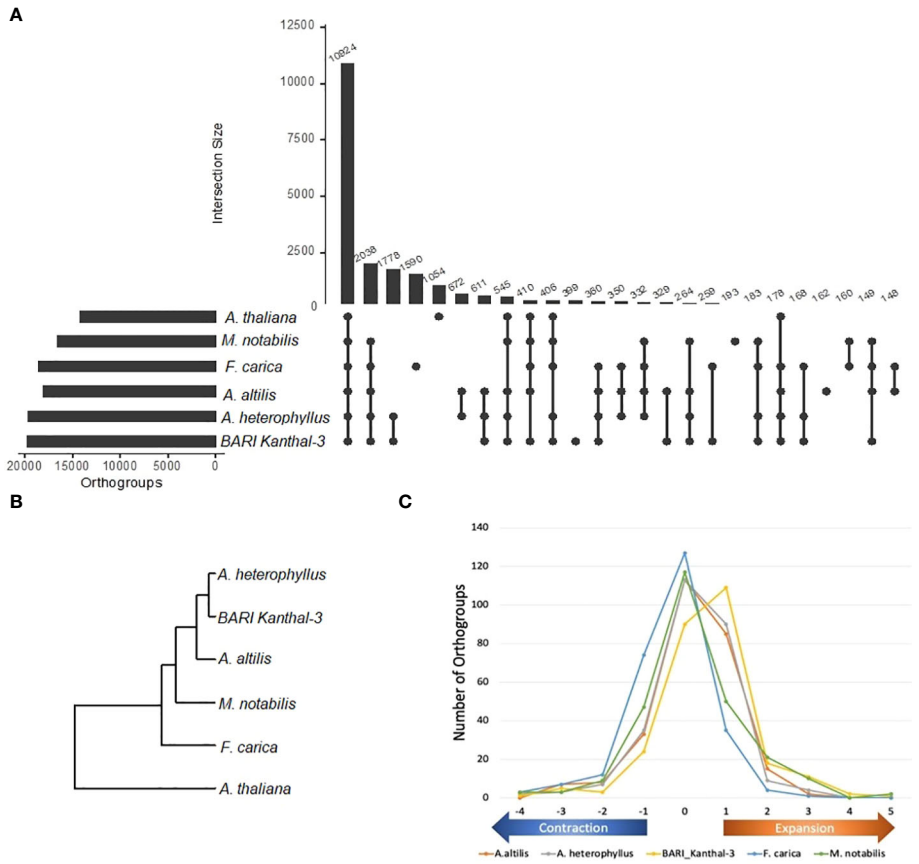


FIGURE 4
Distribution of orthogroups in the genomes of five species (*Artocarpus heterophyllus*, *A. altilis*, *Ficus carica*, *Morus notabilis*, *Arabidopsis thaliana* and BARI Kanthal-3). **(A)** diagram showing unique and shared orthogroups in three genomes of *Artocarpus* species, *M. notabilis*, *F. carica* and *Arabidopsis*. Each number represents the number of orthogroups in common between each pairing. **(B)** Inferred phylogenetic tree constructed with common orthogroups from *A. heterophyllus*, *A. altilis*, *M. notabilis*, *F. carica*, *A. thaliana* and BARI Kanthal-3 (*A. heterophyllus)* genomes. The tree was constructed using rooted gene tree method in OrthoFinder. **(C)** Relative numbers of flowering gene orthogroups in each of the Moraceae genomes compared to *Arabidopsis thaliana*.

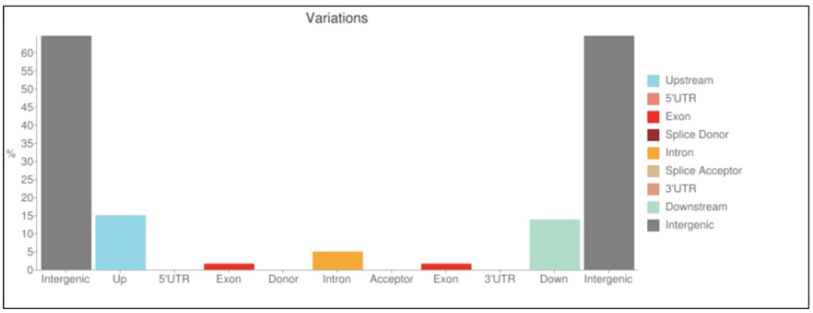
suggesting an altered protein function in BARI Kanthal-3 (Figure 5B). One of the important findings of this study is the high level of heterozygosity in the year-round fruiting jackfruit genome. The high level of heterozygosity in *A. heterophyllus* genome raises the question of which allele, for each heterozygous locus, is represented by the reference genome (Table S2). Therefore, the inherent differences between individual plants should always be considered when utilizing the reference genome to detect SNP variants (Hawkins et al., 2016).

TABLE 3 Results of variant analysis of *Artocarpus heterophyllus* cv. BARI Kanthal-3.

Genomic region	Simple SNPs	Heterozygous SNPs
Exonic	144,787	250,715
Intronic	426,997	739,288
Inter-genic	5,111,742	9,369,507
Total	5,683,526	10,359,510

SNPs have been indicated as the major factors in the creation of phenotypic variation and their effect on functional changes of genes is used as a tool in functional genomics of organisms (Hirakawa et al., 2013). In plants, many traits of interest have been linked with SNPs (Mammadov et al., 2012; Huq et al., 2016; Zhang et al., 2018), including roles in metabolism, cellular processes and signaling, that could have an impact on breeding. This study for the first time identified the SNPs in the genome of a year-round fruiting jackfruit cultivar, which promises the development of genetic markers associated with the important traits of this economically important species including the genes regulating flowering and fruit development. The availability of SNPs within the coding and regulatory sequences also offers the prospect of identifying the causative variations influencing these processes (Varshney,

A



B

Type		
Type (alphabetical order)	**Count**	**Percent**
downstream_gene_variant	3,473,378	13.817%
intergenic_region	16,202,023	64.453%
intron_variant	1,266,335	5.038%
missense_variant	232,587	0.925%
splice_acceptor_variant	1,558	0.006%
splice_donor_variant	1,597	0.006%
splice_region_variant	29,995	0.119%
start_lost	710	0.003%
stop_gained	4,750	0.019%
stop_lost	812	0.003%
stop_retained_variant	376	0.001%
synonymous_variant	162,857	0.648%
upstream_gene_variant	3,760,809	14.961%

FIGURE 5
Variant annotation and effects of variant on genes. **(A)** Distribution and abundance of variants in the genome of BARI Kanthal-3, and **(B)** Number of effects by type and region of the gene variants.

2010). One of the hallmark findings of this study is that majority of the SNPs (47.29%) of BARI Kanthal-3 were localized in the intergenic regions in the proximity of genes (including 5′ UTR, 3′ UTR, and introns). Approximately, 25% of the intergenic SNPs were detected within the region spanning 10 kb upstream of the gene start site and 10 kb downstream of the gene end site (Yamamoto et al., 2007), implying the possibility that some of these SNPs affect the expression of the nearest neighboring genes. It has been reported that a high frequency of genetic variants in the noncoding regions likely results from less selection pressure from natural selection and/or domestication (Barreiro et al., 2008). However, DNA polymorphisms in these regions have been reported to play important roles during evolution and domestication. For example, a mutation in the 5′ regulatory region of the *qSH1* gene, an ortholog of the *Arabidopsis* homeobox gene *REPLUMLESS* (*RPL*) results in the absence of abscission zone formation and thus loss of seed shattering in a subset of temperate *japonica* cultivar of rice (Konishi et al., 2006). Similarly, a considerable number of mutations in introns in pre-harvest sprouting (PHS) genes lead to PHS in rice plant (Lee et al., 2017). Among the 12 PHS mutants (phs), mutations in genes encoding major enzymes of the carotenoid biosynthesis pathway, cause photo-oxidation and ABA-deficiency phenotypes, of which the latter is a major factor controlling the PHS trait in rice (Fang et al., 2008). Interestingly, in jackfruit, MADS-box genes and carotenoid biosynthesis genes, were the primary targets for domestication (Laricchia et al., 2018). However, the role of inter-genic SNPs in *A. heterophylus* in domestication of this horticultural plant needs to be explored further.

SNP markers have become extremely popular in plant molecular genetics due to their genome-wide abundance and amenability for high-throughput detection platforms. For example, SNPs regulating various Quantitative Trait Loci (QTL) responsible for cold and disease resistance such as such as blight, bacterial canker and gray mold have been reported (Zhang et al., 2003; Coaker and Francis, 2004), and SNPs associated with flowering in *Raphanus sativus* have recently been discovered by transcriptome sequencing and computational analysis (Kim et al., 2019). An additional transcriptomics study in BARI Kanthal-3 is

needed for the identification of genes associated with flowering time and year-round fruiting.

4 Conclusions

Although genomic information has been made available for seasonal jackfruit, this is the first report of an assembled and annotated genome for a year-round fruiting jackfruit variety. The annotation yielded a very similar number of annotated genes to a recent genome assembly of a seasonal jackfruit together with a high representation of core othologous genes. An orthologous gene analysis with other Moraceae gennomes confirmed the close realtionship between BARI Kanthal-3 and other *Arctocarpus* spp. but also indicated a higher number of genes related to flowering time were present in the year-round fruiting variety.

This study also reported the distribution of a large collection of nucleotide variation identified across the genome that can be used to identify new functional genes and their regulatory activities specific to BARI Kanthal-3. Furthermore, the genomic data and the identified SNPs of this year-round fruiting jackfruit cultivar could facilitate further genomics and post-genomics studies for detecting other trait-specific genes that are essential for molecular breeding of jackfruit. One of the limitations of this study is lack of information on the role of inter-genic SNPs in BARI Kanthal-3 and the impact on gene function (i.e., changes in protein-coding genes, differential gene expression, and the specific functions of protein-coding genes). To further understand the underlying molecular mechanisms of unique traits of this new variety, a high quality long read genome assembly of BARI-Kanthal-3 and comparative transcriptome analysis with seasonal jackfruit are needed.

The fruit nutritional quality, yield and year-round fruiting properties of BARI Kanthal-3 indicate a unique and valuable genetic material for the improvement of commercial cultivation and development of jackfruit-based processing industry in Bangladesh. The genomic data, associated orthogroups, and SNPs identified in this research will be useful for characterization of trait-specific genes and development of markers for molecular breeding for the improvement of jackfruit, and provides an opportunity to develop this underutilized crop for ensuring food and nutritional security for the increasing population of Bangladesh and other tropical countries.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at https://www.ncbi.nlm.nih.gov/nuccore/JAIQDR000000000.1, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA686208.

Ethics statement

Permission to collect leaf sample of *Artocarpus heteropyllus* Lam. was obtained.

Author contributions

All authors contributed intellectually to this study. TI conceived the study, designed the experiment, coordinated the project, provided reagents and laboratory support, interpreted the results, wrote and revised the manuscript. NA collected the plant samples, extracted DNA, and prepared the draft manuscript; CK assembled and annotated the sequenced data, interpreted results and wrote the manuscript; MNH, interpreted the results, wrote and revised the manuscript; MJH, provided plant samples, collected and interpreted phenotypic and biochemical data and wrote manuscript; NUM and DRG, conducted experiments and prepared library for the DNA sequencing using Illumina Nextseq 550; AAN and RI, analyzed sequenced data, prepared phylogenetic tree and wrote the manuscript; PKB, wrote and revised the manuscript; AGS, coordinated, wrote and revised the manuscript; All authors read, revised, edited and approved the final manuscript.

Funding

This project was funded by the 'BSMRAU Physical Facility and Research Capacity Strengthening Project' under the Ministry of Education of the People's Republic of Bangladesh. Bioinformatics support and analysis was supported with funds from the Bangabandhu Research Chair in Food Security held by AGS at GIFS and supported by GIFS and the Bangladesh Agriculture Research Council.

Acknowledgments

We are thankful to the 'BSMRAU Physical Facility and Research Capacity Strengthening Project' under the Ministry of Education of the People's Republic of Bangladesh for funding this research. The Global Institute for Food Security (GIFS), Canada and BDREN of UGC, Bangladesh also deserve our thanks for offering facilities for bioinformatics analyses of the genomic data of jackfruit cv. BARI Kanthal-3.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1044420/full#supplementary-material

SUPPLEMENTARY TABLE 1
Comparison between the previously reported *Artocarpus heterophyllus* sequence and BARI Kanthal-3 sequence.

SUPPLEMENTARY TABLE 2
Flowering locus C (FLC). Listing of flowering gene orthologues (Enclosed in a separate file).

SUPPLEMENTARY FIGURE 1
Categories of candidate SNPs identified in the genome of BARI_Kanthal-3.

References

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., et al. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* 20, 1–17. doi: 10.1186/s13059-019-1829-6

Anaya-Esparza, L. M., González-Aguilar, G. A., Domínguez-Ávila, J. A., Olmos-Cornejo, J. E., Pérez-Larios, A., and Montalvo-González, E. (2018). Effects of minimal processing technologies on jackfruit (Artocarpus heterophyllus lam.) quality parameters. *Food Bioprocess. Technol.* 11, 1761–1774. doi: 10.1007/s11947-018-2136-z

Azad, A., Jones, J., and Haq, N. (2007). Assessing morphological and isozyme variation of jackfruit (Artocarpus heterophyllus lam.) in Bangladesh. *Agroforestry Syst.* 71, 109–125. doi: 10.1007/s10457-007-9039-8

Barreiro, L. B., Laval, G., Quach, H., Patin, E., and Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nat. Genet.* 40, 340–345. doi: 10.1038/ng.78

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in arabidopsis thaliana. *Nucleic Acids Res.* 44, D1167–D1171. doi: 10.1093/nar/gkv1054

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR. Genomics Bioinf.* 3, lqaa108. doi: 10.1093/nargab/lqaa108

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Chen, F., Song, Y., Li, X., Chen, J., Mo, L., Zhang, X., et al. (2019). Genome sequences of horticultural plants: past, present, and future. *Horticult. Res.* 6, 1–23. doi: 10.1038/s41438-019-0195-6

Coaker, G., and Francis, D. M. (2004). Mapping, genetic effects, and epistatic interaction of two bacterial canker resistance QTLs from lycopersicon hirsutum. *Theor. Appl. Genet.* 108, 1047–1055. doi: 10.1007/s00122-003-1531-6

Das, A., Ianakiev, P., Baten, A., Nehleen, R., Ehsan, T., Ahmed, O., et al. (2018). Genome of tenualosa ilisha from the river padma, Bangladesh. *BMC Res. Notes* 11, 1–3. doi: 10.1186/s13104-018-4028-8

Dhar, M. (1998). Techniques of vegetative and *in vitro* propagation of jackfruit, *Salna, Gazipur Bangladesh:Institute Postgraduate Studies in Agriculture. Ph D Thesis,* pp. 27–31

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 1–14. doi: 10.1186/s13059-019-1832-y

Fang, J., Chai, C., Qian, Q., Li, C., Tang, J., Sun, L., et al. (2008). Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *Plant J.* 54, 177–189. doi: 10.1111/j.1365-313X.2008.03411.x

Galperin, M. Y., and Koonin, E. V. (2010). From complete genome sequence to 'complete'understanding? *Trends Biotechnol.* 28, 398–406. doi: 10.1016/j.tibtech.2010.05.006

Gardner, E. M., Johnson, M. G., Ragone, D., Wickett, N. J., and Zerega, N. J. (2016). Low-coverage, whole-genome sequencing of artocarpus camansi (Moraceae) for phylogenetic marker development and gene discovery. *Appl. Plant Sci.* 4, 1600017. doi: 10.3732/apps.1600017

Goswami, C., and Chacrabati, R. (2016). ""Jackfruit (*Artocarpus heterophylus*)„" in MSJ. Simmonds and V. R. Preedy (Eds) *Nutritional composition of fruit cultivars* (Elsevier), 317–335. doi: 10.1016/B978-0-12-408117-8.00014-3

Hawkins, C., Caruana, J., Schiksnis, E., and Liu, Z. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep29017

Hirakawa, H., Shirasawa, K., Ohyama, A., Fukuoka, H., Aoki, K., Rothan, C., et al. (2013). Genome-wide SNP genotyping to infer the effects on gene functions in tomato. *DNA Res.* 20, 221–233. doi: 10.1093/dnares/dst005

Hoque, M. N., Istiaq, A., Clement, R. A., Sultana, M., Crandall, K. A., Siddiki, A. Z., et al. (2019). Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. *Sci. Rep.* 9, 1–14. doi: 10.1038/s41598-019-49468-4

Hossain, A. (1996). *Status report on genetic resources of jackfruit in Bangladesh* (Singapore: International Plant Genetic Resources Institute Regional Office).

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0

Huq, M. A., Akter, S., Nou, I. S., Kim, H. T., Jung, Y. J., and Kang, K. K. (2016). Identification of functional SNPs in genes and their effects on plant phenotypes. *J. Plant Biotechnol.* 43, 1–11. doi: 10.5010/JPB.2016.43.1.1

Islam, M. S., Saito, J. A., Emdad, E. M., Ahmed, B., Islam, M. M., Halim, A., et al. (2017). Comparative genomics of two jute species and insight into fibre biogenesis. *Nat. Plants* 3, 1–7. doi: 10.1038/nplants.2016.223

Khan, A. U., Ema, I. J., Faruk, M., Tarapder, S. A., Khan, A. U., Noreen, S., et al (2021). A review on importance of artocarpus heterophyllus L. (Jackfruit) *J. Multidiscip. Appl. Natural Science* 1 (2), 106–116. doi: 10.47352/jmans.v1i2.88

Kim, J., Manivannan, A., Kim, D.-S., Lee, E.-S., and Lee, H.-E. (2019). Transcriptome sequencing assisted discovery and computational analysis of novel SNPs associated with flowering in raphanus sativus in-bred lines for marker-assisted backcross breeding. *Horticult. Res.* 6, 1–12. doi: 10.1038/s41438-019-0200-0

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, 1–9. doi: 10.1186/gb-2004-5-2-r12

Laricchia, K. M., Johnson, M. G., Ragone, D., Williams, E. W., Zerega, N. J., and Wickett, N. J. (2018). A transcriptome screen for positive selection in domesticated breadfruit and its wild relatives (Artocarpus spp.). *Am. J. Bot.* 105, 915–926. doi: 10.1002/ajb2.1095

Lasky, J. R., Des Marais, D. L., Mckay, J. K., Richards, J. H., Juenger, T. E., and Keitt, T. H. (2012). Characterizing genomic variation of arabidopsis thaliana: the roles of geography and climate. *Mol. Ecol.* 21, 5512–5529. doi: 10.1111/j.1365-294X.2012.05709.x

Lee, G.-A., Jeon, Y.-A., Lee, H.-S., Hyun, D. Y., Lee, J.-R., Lee, M.-C., et al. (2017). New genetic loci associated with preharvest sprouting and its evaluation based on the model equation in rice. *Front. Plant Sci.* 8, 1393. doi: 10.3389/fpls.2017.01393

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lin, X., Feng, C., Lin, T., Harris, A., Li, Y., and Kang, M. (2022). Jackfruit genome and population genomics provide insights into fruit evolution and domestication history in China. *Horticult. Res.* 9, 1–13. doi: 10.1093/hr/uhac173

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1, 18. doi: 10.1186/2047-217X-1-18

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics*, 728398. doi: 10.1155/2012/728398

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Morgil, H., Gercek, Y. C., and Tulum, I. (2020). "Single nucleotide polymorphisms (SNPs) in plant genetics and breeding," In *The recent topics in genetic polymorphisms* M. Çalışkan et al. eds. (IntechOpen). doi: 10.5772/intechopen.91886

Naik, K. C. (1949) South Indian fruits and their culture. *South Indian fruits their culture* P. Varadachary & Co., Madras. p. 335

Ohri, D., and Kumar, A. (1986). Nuclear DNA amounts in some tropical hardwoods. *Caryologia* 39, 303–307. doi: 10.1080/00087114.1986.10797792

Ong, B., Nazimah, S., Osman, A., Quek, S., Voon, Y., Hashim, D. M., et al. (2006). Chemical and flavour changes in jackfruit (Artocarpus heterophyllus lam.) cultivar J3 during ripening. *Postharvest. Biol. Technol.* 40, 279–286. doi: 10.1016/j.postharvbio.2006.01.015

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R., Hellinga, A. J., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20, 1–18. doi: 10.1186/s13059-019-1905-y

Rahman, A. M., Huq, E., Mian, A., and Chesson, A. (1995). Microscopic and chemical changes occurring during the ripening of two forms of jackfruit (Artocarpus heterophyllus l.). *Food Chem.* 52, 405–410. doi: 10.1016/0308-8146(95)93290-8

Rahman, M. A., Nahar, N., Mian, A. J., and Mosihuzzaman, M. (1999). Variation of carbohydrate composition of two forms of fruit from jack tree (Artocarpus heterophyllus l.) with maturity and climatic conditions. *Food Chem.* 65, 91–97. doi: 10.1016/S0308-8146(98)00175-7

Rahman, M., Patwary, M. A., Barua, H., Nahar, S., and Ahmmed, A. N. F. (2016). Evaluation of yield and quality of three jackfruit (Artocarpus heterophyllus l.) genotypes. *Agriculturists.* 14, 107–111. doi: 10.3329/agric.v14i1.29108

Saha, M., Saha, M., Rahman, M., Nazrul, M., Quasem, A., Halder, N., et al (1996). "Variability in jackfruit," in *Proceedings of the Internal Research Review Workshop* (Gazipur, Bangladesh: Horticulture Research Center, BARI), 1–4.

Sahu, S. K., Liu, M., Yssel, A., Kariba, R., Muthemba, S., Jiang, S., et al. (2020). Draft genomes of two artocarpus plants, jackfruit (*A. heterophyllus*) and breadfruit (*A. altilis*). *Genes* 11, 27. doi: 10.3390/genes11010027

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14

Siddiki, A. Z., Baten, A., Billah, M., Alam, M., Shawrob, K. S. M., Saha, S., et al. (2019). The genome of the black Bengal goat (Capra hircus). *BMC Res. Notes* 12, 1–3. doi: 10.1186/s13104-019-4400-3

Sidhu, A. S. (2012). *Jackfruit improvement in the Asia-pacific region: A status report (APAARI)* FAO Regional Office for Asia & the Pacific (FAO RAP) Maliwan Mansion, 39, Phra Atit Road Bangkok 10200.

Simmonds, M., and Preedy, V. R. (2015). *Nutritional composition of fruit cultivars* (Academic Press). (San Diego, USA: Academic Press).

Somerville, C., and Koornneef, M. (2002). A fortunate choice: the history of arabidopsis as a model plant. *Nat. Rev. Genet.* 3, 883–889. doi: 10.1038/nrg927

Statistics., Y. O. A. (2020). *Bangladesh Bureau of statistics (BBS), statistics and informatics division (SID), ministry of planning, government of the people's republic of Bangladesh* (Government of the People's Republic of Bangladesh).

Varshney, R. K. (2010). "Gene-based marker systems in plants: high throughput approaches for marker discovery and genotyping," in *Molecular techniques in crop improvement* (Dordrecht: Springer), 119–142. doi: 10.1007/978-90-481-2967-6_5

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. doi: 10.1093/bioinformatics/btx153

Yamamoto, Y. Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., et al. (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 8, 1–23. doi: 10.1186/1471-2164-8-67

Zhang, L., Lin, G., Nino-Liu, D., and Foolad, M. R. (2003). Mapping QTLs conferring early blight (Alternaria solani) resistance in a lycopersicon esculentum× l. hirsutum cross by selective genotyping. *Mol. Breed.* 12, 3–19. doi: 10.1023/A:1025434319940

Zhang, Y., Li, B., Xu, F., He, S., Zhang, Y., Sun, L., et al. (2021a). Jackfruit starch: Composition, structure, functional properties, modifications and applications. *Trends Food Sci. Technol.* 107, 268–283. doi: 10.1016/j.tifs.2020.10.041

Zhang, W., Mirlohi, S., Li, X., and He, Y. (2018). Identification of functional single-nucleotide polymorphisms affecting leaf hair number in brassica rapa. *Plant Physiol.* 177, 490–503. doi: 10.1104/pp.18.00025

Zhang, Y., Wang, Q., Zhang, Y., Wu, G., Tan, L., and Zhang, Z. (2022). Effects of moisture content on digestible fragments and molecular structures of high amylose jackfruit starch prepared by improved extrusion cooking technology. *Food Hydrocolloids.* 133, 108023. doi: 10.1016/j.foodhyd.2022.108023

Zhang, Y., Zuo, H., Xu, F., Zhu, K., Tan, L., Dong, W., et al. (2021b). The digestion mechanism of jackfruit seed starch using improved extrusion cooking technology. *Food Hydrocolloids.* 110, 106154. doi: 10.1016/j.foodhyd.2020.106154

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fevo.2022.1057785

OPEN ACCESS

EDITED BY
Mohd. Kamran Khan,
Selçuk University, Türkiye

REVIEWED BY
Manuel De La Estrella,
University of Córdoba, Spain
Tofazzal Islam,
Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

*CORRESPONDENCE
Maria M. Romeiras
✉ mmromeiras@isa.ulisboa.pt

SPECIALTY SECTION
This article was submitted to Conservation and Restoration Ecology, a section of the journal Frontiers in Ecology and Evolution

RECEIVED 03 October 2022
ACCEPTED 13 December 2022
PUBLISHED 09 January 2023

CITATION
Brilhante M, Catarino S, Darbyshire I, Bandeira S, Moldão M, Duarte MC and Romeiras MM (2023) Diversity patterns and conservation of the *Vigna* spp. in Mozambique: A comprehensive study. *Front. Ecol. Evol.* 10:1057785. doi: 10.3389/fevo.2022.1057785

Diversity patterns and conservation of the *Vigna* spp. in Mozambique: A comprehensive study

Miguel Brilhante[1], Sílvia Catarino[1,2], Iain Darbyshire[3], Salomão Bandeira[4], Margarida Moldão[1], Maria Cristina Duarte[5] and Maria M. Romeiras[1,5]*

[1]Linking Landscape, Environment, Agriculture and Food (LEAF), Associated Laboratory TERRA, Instituto Superior de Agronomia (ISA), Universidade de Lisboa, Lisboa, Portugal, [2]Forest Research Center (CEF), Associated Laboratory TERRA, Instituto Superior de Agronomia (ISA), Universidade de Lisboa, Lisboa, Portugal, [3]Royal Botanic Gardens, Kew, Richmond, United Kingdom, [4]Department of Biological Sciences, Eduardo Mondlane University, Maputo, Mozambique, [5]Centre for Ecology, Evolution and Environmental Changes (cE3c) and Change–Global Change and Sustainability Institute, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

Mozambique supports a high diversity of native legume species, including many Crop Wild Relatives (CWRs). Among them, the still understudied genus *Vigna* is a particularly notable and promising donor of favorable traits for crop improvement. This study aims to provide an updated overview of *Vigna* CWRs diversity in Mozambique, defining a conservation strategy for priority target taxa and areas. A checklist of *Vigna* taxa was prepared and using five criteria (taxonomic group, ethnobotanical value, global and regional distributions, and *ex situ* conservation status), the prioritization of each taxon was determined. The distribution of *Vigna* native to Mozambique was studied and diversity hotspots were detected; gaps in *in situ* conservation were analyzed by overlaying species distribution with Mozambique's Protected Areas Network. Maps predicting the differences between future conditions and baseline values were performed to investigate expected changes in temperature and precipitation in *Vigna's* occurrence areas. There are 21 *Vigna* native taxa occurring in Mozambique, with the Chimanimani Mountains and Mount Gorongosa, as diversity hotspots for the genus. Following the IUCN Red List criteria, 13 taxa are of Least Concern, while the remaining eight are currently Not Evaluated. According to their priority level for further conservation actions, 24% of the taxa are of high priority, 67% of medium priority, and 9% of low priority. The important hotspot of Chimanimani Mountains is among the areas most affected by the predicted future increase in temperature and reduction of rainfall. The obtained distribution and species richness maps, represent a relevant first tool to evaluate and improve the effectiveness of Protected Areas and IPAs of Mozambique for the conservation of *Vigna* CWRs. The *in situ* gap analysis showed that 52% of the *Vigna* taxa are unprotected; this could be overcome by establishing reserves in *Vigna* diversity centers, considering the different types of habitats to which the different taxa are adapted, and by increasing *in situ* protection for the

high priority ones. The *ex situ* conservation of *Vigna* is very limited and storing seed collections of these CWRs, is an essential component in global food security, as some taxa seem suitable as donors of genetic material to increase resistance to pests and diseases, or to drought and salinity. Overall, we provide recommendations for future research, collecting, and management, to conserve *Vigna* CWR in Mozambique, providing new data for their sustainable use in crop enhancement, as well as proposing measures for future conservation programs.

KEYWORDS

conservation strategies, cowpea, gap analysis, protected areas, East Africa, species richness, crop wild relatives

1. Introduction

Crop Wild Relatives (CWRs) are undomesticated plants, related to crop species that have the potential to contribute genetic material for crop enhancement (Heywood et al., 2007). Considering the potential negative impact of climate change on biodiversity and food security, strategies for the study and conservation of genetic resources in CWRs must be considered a global priority (Maxted et al., 2010). CWRs offer important sources of useful agronomic traits, which could be incorporated into crop breeding programs, including: tolerance to abiotic and biotic stress; nutraceutical characteristics; and significantly improved crop yields (Pimentel et al., 1997; Tanksley and McCouch, 1997; Meilleur and Hodgkin, 2004; Maxted et al., 2010). Considering the often-limited lifespan of resistant cultivars, the demand for wild genetic breeding material is imperative (e.g., Rocha et al., 2021; Mba and Ogbonnaya, 2022; Vincent et al., 2022). Crop improvement is especially important in areas most likely to be affected by climate changes and with poor management practices of agricultural soils, such as in sub-Saharan Africa (SSA), where food security is particularly alarming (Sintayehu, 2018) and agricultural systems are highly dependent on rainfall rather than on manmade irrigation schemes (Thornton and Herrero, 2015). Sustainable crop production also requires efficient soil fertility management, but soils in wide areas of SSA are highly affected by leaching and erosion, leading to depletion of nutrients (Vidigal et al., 2019). Importantly, more efficient use of nitrogen is urgently needed (Yadav et al., 2017). Crop species from the legume family (Fabaceae or Leguminosae), especially the so-called pulse crops (dry beans, chickpeas, and lentils), can fix atmospheric nitrogen, besides having a high nutritional value as a source of protein (Snapp et al., 2018; Brilhante et al., 2021). African legume species include many CWRs as well as many underutilized locally-adapted crops; thus, promoting and enhancing the productivity of local legume crops will be a valuable contribution to the sustainability of cropping systems and the maintenance of food security in SSA (Vidigal et al., 2019). Many African CWRs are currently threatened, largely due to anthropogenic influences such as habitat loss, unsustainable soil exploitation, and invasive species (e.g., Angola: Catarino et al., 2021a; Malawi: Khaki Mponya et al., 2021; Zambia: Ng'uni et al., 2019). Despite these studies, the CWRs in SSA are still poorly studied, despite being a very rich flora.

Mozambique is recognized as one of the most vulnerable countries to climate change in Africa. More frequent and severe droughts, heatwaves, floods, cyclones, and wildfires are expected in the near future, with a strong negative impact on native flora and food production (INGC, 2009). As an example of climate change effect in Mozambique, it stands out the Cyclone Idai which hit the central area of this country (Sofala) in March 2019 (Charrua et al., 2021a).

A strong candidate genus with high agronomic and food source potential is *Vigna* Savi (Simon et al., 2007), which includes several widely disseminated crops such as cowpea (*Vigna unguiculata*). Most taxa of *Vigna* occur in Africa (Singh, 2020), and ca. 20 are known in Mozambique, where they grow under different ecological conditions (Mackinder et al., 2001; Odorico et al., 2022). In this region, local populations use pulses, particularly of *Vigna*, for human consumption due to their great protein and iron content, medicinal properties, and as a cheaper mean to enrich the soil with nitrogen (N) (Charrua et al., 2021b). This latter factor is particularly important given that the main factor limiting yield production in this region at present is the lack of N, as a result of the poor agricultural practices as slash-and-burn methods used in tropical and sub-tropical conditions (Simon et al., 2007; Catarino et al., 2021b). The conservation of *Vigna* CWRs *in situ* is thus particularly important, but effective conservation of these bioresources requires a detailed knowledge of the species diversity, distribution, and threats, which is lacking in many SSA countries such as Mozambique (Darbyshire et al., 2019).

In this study, we established a comprehensive dataset containing information on the diversity of native *Vigna* CWRs in Mozambique, as well as georeferenced occurrence records for each taxon. Overall, we aim to provide an updated

understanding of the diversity of *Vigna* CWRs, to contribute new data to support their conservation and sustainable use in Mozambique. Specifically, we intend to: (i) characterize the diversity of the *Vigna* CWRs occurring in Mozambique, focusing on their taxonomy, distribution, and main uses; (ii) identify the *Vigna* CWRs taxa prioritary for future conservation measures; and (iii) identify the *Vigna* CWRs centers of diversity in Mozambique and gaps in *in situ* and *ex situ* conservation. Finally, some guidelines to define a conservation strategy for the sustainable use of *Vigna* CWRs in Mozambique are proposed.

2. Materials and methods

2.1. Study area

Mozambique is located in the south-eastern sector of the African continent with an area of 801,509 km^2, bordering Tanzania to the north, the Indian Ocean to the east, Zambia to the northwest, Malawi, and Zimbabwe to the west, and Eswatini and South Africa to the southwest. Its territory is divided into 10 provinces, of which Maputo City, located in the extreme south, encompasses the country's capital. About 70% of the country is covered by forests or woody vegetation and 26% is included within the network of terrestrial Protected Areas (Ministry for the Coordination of Environmental Affairs, 2014).

Mozambique has a tropical climate in general, with a subtropical climate in the southernmost region. In general, the country has two well-defined seasons: a cold and dry season (from April to October) and a warm and wet season (from October to April) (Barbosa et al., 2001). Due to its great extent, the average annual temperature is differentiated as follows: the northern region with 25.5°C in the coastal region and 18.0°C in the higher regions; followed by the central region with 25.0°C in the coastal region and 20.0°C in the higher interior regions, and finally the southern region with 23.0°C in the coastal region and 25.0°C in the interior region. The average annual rainfall reaches 1,030 mm, ranging from 300 mm/year in the southern region (e.g., lower in Gaza inland toward the border with South Africa and higher along the Maputo coast) to 1,400 mm/year in the Zambezi basin (Uamusse et al., 2017).

The country is considered an important area for plant biodiversity, with high species richness (i.e., about 6,157 native and naturalized plant taxa; Hyde et al., 2022a) and a high level of endemism (Darbyshire et al., 2019), resulting from a great diversity of ecosystems, from dry coastal forests to lowland and montane moist forests, savannas, and mangroves, comprising 13 main ecoregions (Burgess et al., 2004) in which seven plant communities are defined, namely: miombo, mopane, and undifferentiated woodland, Afromontane, halophytic and swamp vegetation, and coastal area (Bandeira et al., 1996).

The geographic distribution of plant richness in Mozambique has been demonstrated in several studies (e.g., Darbyshire et al., 2019; Odorico et al., 2022), which report the following centers of plant endemism: Chimanimani-Nyanga (including the highlands of Manica and Sofala), Mulanje-Namuli-Ribáuè (including the highlands of Nampula and Zambezia), Maputaland [subdivided in three subcentres: Maputaland *sensu stricto* (including Maputo and Gaza), Lebombo Mountains (in Maputo province only), and Inhambane Center (including Gaza and Inhambane)], and Rovuma (including the coastal area of Cabo Delgado, Nampula, and Zambezia).

2.2. CWRs inventory and data collection

A comprehensive dataset on the native *Vigna* taxa occurring in Mozambique was collated, which includes the scientific and (where available) common names, taxonomic section, global and regional native distribution, International Union for Conservation of Nature (IUCN) conservation status, and the main uses of the species. These data were obtained by means of a comprehensive review conducted through herbarium collections, relevant publications on Mozambique Flora, and online databases. Therefore, this study was made using three main sources:

(1) Herbarium specimens of *Vigna* collected in Mozambique from the Herbarium of the Instituto de Investigação Científica Tropical, University of Lisbon (LISC) and Herbarium of the Royal Botanic Gardens, Kew (K) were studied in detail to collect data on the main uses, ecology, life-form, and distribution.

(2) A thorough investigation of native *Vigna* from Mozambique data described in the literature. The following relevant publications were consulted: Singh et al. (1997), Mackinder et al. (2001), Da Silva et al. (2004), Maxted et al. (2004), Darbyshire et al. (2019), Singh (2020), and Odorico et al. (2022).

(3) Relevant online databases, namely: (i) POWO–Plants of the World Online (POWO, 2022)[1] for taxonomic and global distribution data; (ii) Flora of Mozambique (Hyde et al., 2022a)[2] for regional distribution data; PROTA–The Plant Resources of Tropical Africa (PROTA, 2022)[3] to access the main uses of *Vigna* species; (iii) IUCN–Red List of Threatened Species (IUCN, 2022)[4] to access data on threats and conservation actions; (iv) Genesys Database (Genesys, 2022)[5] for *ex situ* conservation data; (v) "The Crop Wild Relatives Project" (CWR Project, 2022)[6]

1 https://powo.science.kew.org
2 https://www.mozambiqueflora.com/
3 https://www.prota4u.org/
4 https://www.iucnredlist.org/
5 https://www.genesys-pgr.org/
6 https://www.cwrdiversity.org/

to investigate the *Vigna* wild species identified as CWR of crop species; and (vi) Global Biodiversity Information Facility (GBIF) (GBIF, 2022a)[7] to extract distribution data from sources in addition to the herbaria listed above; and (vii) Germplasm resources information network (GRIN Taxonomy) (USDA, 2022) for data on gene pool, trait type and breeding type.

The scientific names of each species were checked and updated according to Plants of the World Online (POWO, 2022). Moreover, subgenera and sections of *Vigna* into which each species is placed follow the classification of Maxted et al. (2004).

2.3. CWR prioritization

The conservation of CWRs is essential to guarantee the existence of useful genetic resources to supplement the breeding pool of cultivar species. Therefore, conservation planning is necessary, and prioritization is one of the key steps to developing conservation strategies. Here, based on the collated data, we prioritized the 21 *Vigna* taxa native to Mozambique based on the criteria described by Magos Brehm et al. (2010), which were selected and adapted to the context of the present study. Therefore, we applied the following five criteria: (i) potential utilization for crop improvement; (ii) ethnobotanical uses; (iii) regional distribution; (iv) global distribution; and (v) *ex situ* conservation status. Each criterion was applied as follows:

(i) The potential use for crop improvement was based on the concept of the Taxonomic Group (TG) which recognizes the relatedness between crop species and CWRs using the known taxonomic hierarchy. By the application of the TG concept, in decreasing order of affinity with the cultivar, Taxon Group 1 (TG1) includes the cultivar species and other taxa of the same species; Taxon Group 2 (TG2) includes taxa from the same infrageneric section of the cultivar; Taxon Group 3 (TG3) includes taxa of the same subgenus as the cultivar; and Taxon Group 4 (TG4) includes taxa of the same genus as the cultivar (Maxted et al., 2006). We used TG instead of the Gene Pool (GP) concept given that crossbreeding and genetic diversity information was not available for all the studied species. The TG concept was applied based on cowpea (*Vigna unguiculata* subsp. *unguiculata*) due to its importance in Mozambique as it constitutes the most important crop in the country. The potential use for crop improvement was then scored as: TG1 (score 4); TG2 (score 3); TG3 (score 2); TG4 (score 1).

(ii) The ethnobotanical value was based on the main traditional uses of *Vigna* CWRs, with food use as the main focus. Following the criteria defined by Catarino et al. (2021a), we categorized and scored the ethnobotanical value as: used as food (score 4); two or more uses, excluding food (score 3); one use, excluding food (score 2); no recorded uses (score 1).

(iii) The regional distribution was based on the distribution of *Vigna* CWRs, through the 10 Mozambique provinces. Therefore, we categorized and scored as: one or two provinces (score 4); three, four, or five provinces (score 3); six, seven, or eight provinces (score 2); nine or ten provinces (score 1).

(iv) The global distribution was categorized and scored as: restricted to the neighborhood countries of Mozambique (i.e., Malawi, South Africa, Eswatini, Tanzania, and Zimbabwe) (score 4); restricted to Zambezian Region (score 3); confined to the African continent (score 2); also present outside Africa (score 1). Note that there are no endemic taxa of *Vigna* in Mozambique (Darbyshire et al., 2019).

(v) The *ex situ* conservation status considers the number of accessions available in worldwide germplasm banks. Therefore, we categorized and scored this criterion as: zero accessions (score 4); one to four accessions (score 3); five to nine accessions (score 2); more than ten accessions (score 1).

For each *Vigna* CWR taxon, the scores assigned to each criterion were accumulated, with the total score varying from 5 to 20. Considering the obtained scores, prioritization for conservation was established as follows: high priority (scores between 16 and 20); medium priority (scores between 11 and 15); and low priority (scores between 5 and 10).

2.4. Hotspot diversity and conservation analysis

The geographic distribution of *Vigna* species native to Mozambique was obtained based on occurrence data extracted from the GBIF databases (GBIF, 2022b), in total 1,014 records. Duplicate records (i.e., records with the same collector and collection number) were excluded and the location of all records was confirmed using Google Earth Pro 7.3.4.8573 (Serea, 2018). The occurrence database was supplemented with data from herbarium specimens, as noted above. The location data obtained through the labels of the herbarium specimens were georeferenced according to the criteria established by Chapman and Wieczorek (2006). The final database contained 295 occurrence records belonging to 21 taxa, which served as the basis for distribution studies and hotspot identification. Species richness maps were constructed in QGIS v.3.4.4 software (QGIS

7 https://www.gbif.org/

Development Team, 2022) to detect the areas of high diversity of *Vigna* CWRs in Mozambique (i.e., diversity hotspots). To identify the gaps in the *in situ* conservation of *Vigna* CWRs, the areas with high conservation interest were identified based on local species distribution by overlaying the species distribution map with the Protected Areas Network and Important Plant Areas of (IPAs) Mozambique. The geographic boundaries of Mozambique's Protected Areas were obtained as GIS shapefiles from Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM) (UNEP-WCMC and IUCN, 2022). Mozambique's network of Important Plant Areas (hereafter, IPAs) was obtained from Darbyshire et al. (in press).

We produced two maps forecasting the differences between future conditions and the baseline values to analyze the predicted changes in temperature and precipitation in *Vigna* occurrence areas. Baseline data (1970–2000) were retrieved from the WorldClim database (WorldClim, 2020a). Future mean annual temperature (Bio1) and mean annual precipitation (Bio12) for 2061–2080 were obtained from the Coupled Model Intercomparison Project 6 (CMIP6) (Eyring et al., 2016), also available at the WorldClim website (WorldClim, 2020b). The predicted values of future annual temperature and precipitation were obtained as the mean values of the 21 climate models available for the years 2061–2080 (namely, ACCESS-ESM1-5, BCC-CSM2-MR, CanESM5, CanESM5-CanOE, CMCC-ESM2, CNRM-CM6-1, CNRM-CM6-1-HR, CNRM-ESM2-1, EC-Earth3-Veg, EC-Earth3-Veg-LR, GFDL-ESM4, GISS-E2-1-G, INM-CM4-8, INM-CM5-0, IPSL-CM6a-LR, MIROC6, MIROC-ES2L, MPI-ESM1-2-HR, MPI-ESM1-2-LR, MRI-ESM2-0, and UKESM1-0-LL). The data were selected from the Shared Socio-economic Pathway 3-70 (SSP 3-7.0) since this scenario is in the middle of the range [more details in Hausfather (2019)]. The amplitude of the changes was obtained by subtracting the value of baseline temperature and precipitation from future values using QGIS v.3.4.4 software (QGIS Development Team, 2022).

Regarding the *ex situ* conservation, the total number of *Vigna* CWRs accessions collected in Mozambique were accessed through the Genesys Database (Genesys, 2022). Thus, the representativeness of Mozambique's *Vigna* genetic resources in worldwide Genebanks was assessed.

3. Results

3.1. Diversity of *Vigna* CWRs in Mozambique

Our results revealed the occurrence of 21 *Vigna* native taxa in Mozambique, including 13 species, six subspecies, and two varieties (**Table 1**). Among the 21 taxa, most are perennial climbing herbs (eleven taxa), followed by perennial non-climbing herbs (nine taxa) and only one annual non-climbing herb, *V. reticulata*.

Most taxa of *Vigna* occurring in Mozambique (ca. 80%) are confined to the African continent; the exceptions are *V. luteola*, *V. marina*, *V. vexillata* var. *angustifolia*, and *V. vexillata* var. *vexillata*, which have wider distributions outside Africa.

The information gathered from the IUCN Red List indicated that 62% of those taxa were already evaluated according to the IUCN criteria and categories, with 13 of them classified as of Least Concern; eight taxa remain unevaluated (**Table 1**).

Regarding the infrageneric taxonomic classification of *Vigna*, three subgenera (*Haydonia*, *Plectotropis*, and *Vigna*) and seven sections (*Catiang*, *Comosae*, *Haydonia*, *Liebrechtsia*, *Plectotropis*, *Reticulatae,* and *Vigna*) are recorded in Mozambique (**Figure 1**; **Table 1**). The most diverse subgenus is *Vigna*, with 18 taxa, followed by *Plectotropis* with two taxa and *Haydonia* with one. The most diverse section is *Catiang,* which includes the six *V. unguiculata* subspecies and *V. schlechteri*, whilst section *Vigna* contains four taxa, *Reticulatae* and *Plectotropis* three taxa, and *Liebrechtsia* two taxa. Sections with only one taxon are *Comosae* (*V. comosa*) and *Haydonia* (*V. juncea*).

Through the study of infrageneric categories, it was possible to apply the concept of taxonomic group, which allows to know the proximity between the CWRs taxa and the cultivar.

The results (**Figure 1**; **Table 1**) showed that TG1 is composed of six subspecies of *V. unguiculata*; TG2 is represented by only one species, *V. schlechteri*; TG3 includes 11 taxa belonging to the same subgenus; and TG4 gathers the remaining three taxa.

Of the 21 taxa, 11 (52%) are used by local human populations (**Table 1**), each one for various purposes. All the 11 taxa are used as forage, eight as medicines, seven as food, six as fiber, six as ornamental, and three as auxiliary (i.e., cover crop and soil improver with nitrogen). *Vigna vexillata* var. *vexillata* is the taxon with most uses (auxiliary, fiber, food, forage, and medicine). Moreover, except for *V. unguiculata* subsp. *dekindtiana*, with no recorded uses, the remaining five subspecies of *V. unguiculata* also share five uses (fiber, food, forage, ornamental, and medicine); *V. frutescens* and *V. luteola* have three uses, and *V. marina* and *V. reticulata* two uses; *Vigna oblongifolia* has only one recognized use (forage).

3.2. Priority taxa and *in situ* conservation gap analysis

Based on prioritization criteria (**Table 1**), our results revealed that, from the 21 studied taxa, five (24%) are of high priority, 14 (67%) of medium priority, and two (9%) of low priority (**Figure 2**).

TABLE 1 Scientific name, life form, regional and global native distribution, International Union for Conservation of Nature (IUCN) conservation status, main uses, cowpea taxonomic groups, scoring of the prioritization criteria and conservation priority level of native *Vigna* from Mozambique.

Taxon	Life form	Native range	IUCN[1]	Main uses	Taxonomic group (cowpea)	Prioritization criteria					Total	Prioritization
						Potential utilization for crop	Ethnobotanical value	Regional distribution	Global distribution	Germplasm bank accessions	Score	Level
Vigna antunesii Harms	Perennial non-climbing herb	Tanzania to S. Tropical Africa	LC		TG3	2	1	4	3	4	14	Medium
Vigna comosa Baker	Perennial non-climbing herb	Tropical Africa	LC		TG3	2	1	3	2	4	12	Medium
Vigna frutescens A. Rich.	Perennial climbing herb	Tropical and S. Africa	LC	Food, Forage, Ornamental	TG3	2	4	3	2	2	13	Medium
Vigna gazensis Baker f.	Perennial climbing herb	S. Tropical Africa, Madagascar	LC		TG3	2	1	3	3	4	13	Medium
Vigna juncea Milne-Redh.	Perennial non-climbing herb	Tanzania to S. Tropical Africa	LC		TG4	1	1	4	3	4	13	Medium
Vigna kirkii (Baker f.) J. B. Gillett	Perennial non-climbing herb	Tropical Africa	LC		TG4	1	1	4	2	4	12	Medium
Vigna luteola (Jacq.) Benth.	Perennial climbing herb	Tropics and Subtropics	LC	Auxiliary[2], Forage, Medicinal	TG3	2	3	2	1	4	12	Medium
Vigna marina (Burm.) Merr.	Perennial non-climbing herb	Tropical and Subtropical Old World to Pacific	LC	Auxiliary[2], Forage	TG3	2	3	3	1	3	12	Medium
Vigna oblongifolia A. Rich.	Perennial climbing herb	Tropical and S. Africa, Madagascar	LC	Forage	TG3	2	2	4	2	4	14	Medium

(Continued)

10.3389/fevo.2022.1057785

TABLE 1 (Continued)

Taxon	Life form	Native range	IUCN[1]	Main uses	Taxonomic group (cowpea)	Prioritization criteria					Total	Prioritization
						Potential utilization for crop	Ethnobotanical value	Regional distribution	Global distribution	Germplasm bank accessions	Score	Level
Vigna platyloba Hiern	Perennial non-climbing herb	Tanzania to S. Tropical Africa	LC		TG3	2	1	4	3	4	14	Medium
Vigna pygmaea R. E. Fr.	Perennial non-climbing herb	Cameroon to Tanzania and Botswana	LC		TG3	2	1	4	3	4	14	Medium
Vigna reticulata Hook.f.	Annual non-climbing herb	Tropical Africa, Madagascar	LC	Forage, Medicinal	TG3	2	3	3	2	4	14	Medium
Vigna schlechteri Harms	Perennial non-climbing herb	S. Tropical Africa to KwaZulu-Natal	LC		TG2	3	1	4	4	4	16	High
Vigna unguiculata (L.) Walp. subsp. *dekindtiana* (Harms) Verdc.	Perennial climbing herb	Tropical and S. Africa	NE		TG1	4	1	2	2	3	12	Medium
Vigna unguiculata (L.) Walp. subsp. *pawekiae* Pasquet	Perennial climbing herb	Kenya to S. Tropical Africa	NE	Fiber, Food, Forage, Ornamental, Medicinal	TG1	4	4	4	2	4	18	High
Vigna unguiculata (L.) Walp. subsp. *pubescens* (R. Wilczek) Pasquet	Perennial climbing herb	South Sudan to Mozambique	NE	Fiber, Food, Forage, Ornamental, Medicinal	TG1	4	4	3	4	4	19	High

(Continued)

10.3389/fevo.2022.1057785

TABLE 1 (Continued)

Taxon	Life form	Native range	IUCN[1]	Main uses	Taxonomic group (cowpea)	Prioritization criteria					Total	Prioritization
						Potential utilization for crop	Ethnobotanical value	Regional distribution	Global distribution	Germplasm bank accessions	Score	Level
Vigna unguiculata (L.) Walp. subsp. *stenophylla* (Harv.) Maréchal, Mascherpa and Stainier	Perennial climbing herb	S. Tropical and S. Africa	NE	Fiber, Food, Forage, Ornamental, Medicinal	TG1	4	4	4	2	3	17	High
Vigna unguiculata (L.) Walp. subsp. *tenuis* (E. Mey.) Maréchal, Mascherpa and Stainier	Perennial non-climbing herb	S. Tropical Africa to KwaZulu-Natal	NE	Fiber, Food, Forage, Ornamental, Medicinal	TG1	4	4	3	3	2	16	High
Vigna unguiculata (L.) Walp. subsp. *unguiculata*	Perennial climbing herb	Cape Verde, Tropical and S. Africa	NE	Fiber, Food, Forage, Ornamental, Medicinal	TG1	4	4	2	1	3	14	Medium
Vigna vexillata (L.) A. Rich. var. *angustifolia* (Schumach. and Thonn.) Baker	Perennial climbing herb	Tropical and Subtropical Old World	NE		TG4	1	1	3	1	3	9	Low
Vigna vexillata (L.) A. Rich. var. *vexillata*	Perennial climbing herb	Tropics and Subtropics	NE	Auxiliary[2], Fiber, Food, Forage, Medicinal	TG4	1	4	2	1	2	10	Low

[1]IUCN conservation status: CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern; DD, data deficient; NE, not evaluated.
[2]Auxiliary use: Cover crop and soil improver with nitrogen.

10.3389/fevo.2022.1057785

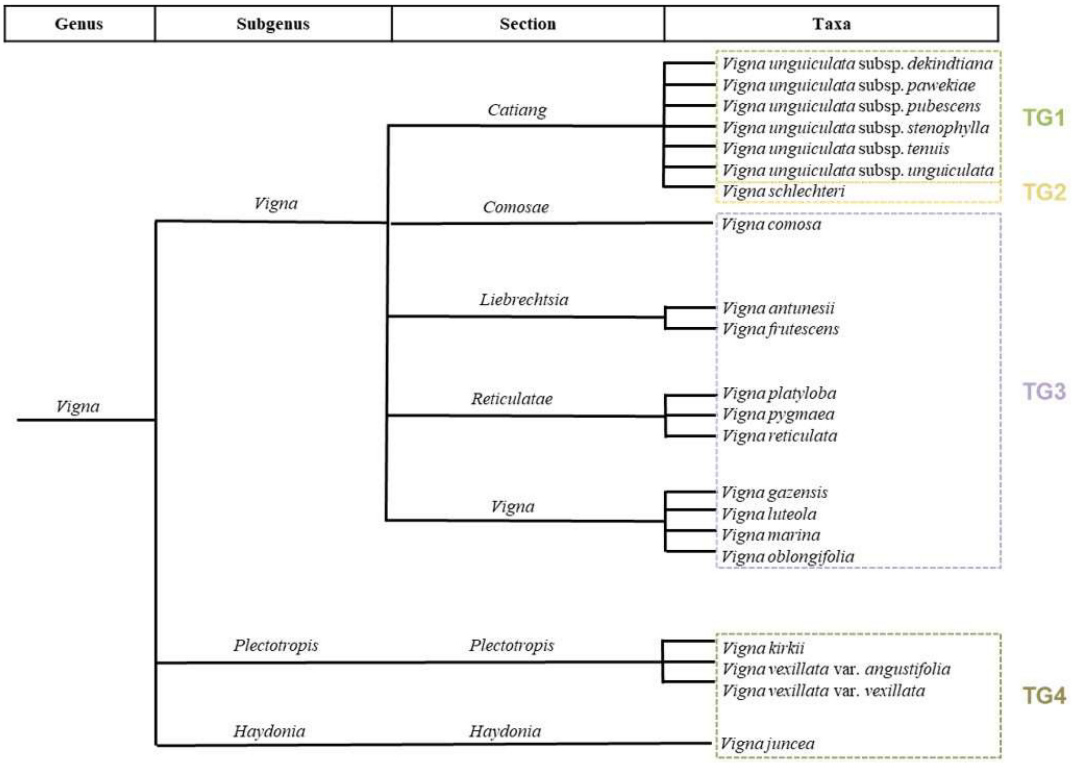


FIGURE 1
Infrageneric classification of *Vigna* taxa native to Mozambique standing out subgenus, sections, and taxonomic groups of each taxon.

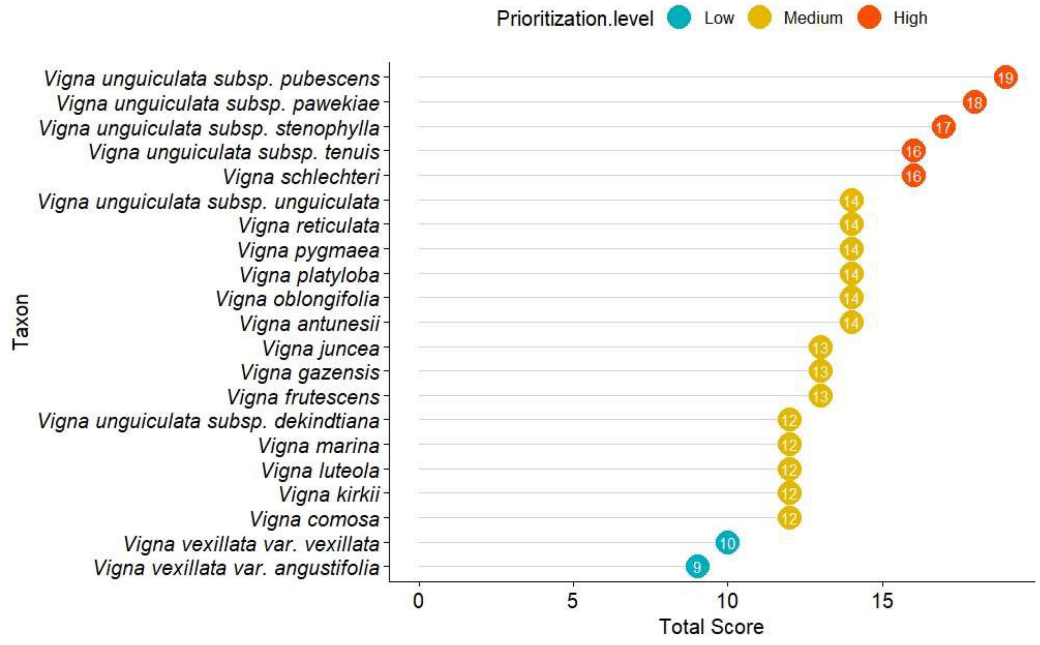


FIGURE 2
Prioritization scores per *Vigna* native crop wild relatives (CWRs) to Mozambique, grouped by prioritization level (high, in orange; medium, in yellow; low, in blue). Taxa are listed by decreasing priority score.

Those classified as highly prioritary for conservation are: *V. unguiculata* subsp. *pubescens*, *V. unguiculata* subsp. *pawekiae*, *V. unguiculata* subsp. *stenophylla*, *V. schlechteri*, *V. unguiculata* subsp. *tenuis*, that is, four of the six subspecies of *V. unguiculata*. Of these, only *V. schlechteri*, has been evaluated according to the IUCN Red List criteria. Among the 14 taxa considered as of medium priority for conservation, 12 are classified as LC by IUCN, while two remain unevaluated (*V. unguiculata* subsp. *dekindtiana* and *V. unguiculata* subsp. *unguiculata*). Only two taxa are considered of low priority for conservation, *V. vexillata* var. *vexillata* and *V. vexillata* var. *angustifolia*, neither of which evaluated for the IUCN Red List. However, they are likely to be of Least Concern since they have a wide distribution, and are not exclusive to the African continent, unlike the other taxa.

The distribution of the genus *Vigna* in Mozambique was studied based on 295 occurrences corresponding to 21 taxa. A species richness map was constructed and the areas of high diversity in *Vigna* CWR in Mozambique were identified (i.e., diversity hotspots of the genus). Based on 25 km circular buffers, the number of taxa ranged from 1 to 7, as shown in **Figure 3**. The diversity hotspots of *Vigna* genus are in the provinces of Manica, Maputo, Sofala, and Tete. The map (**Figure 3**) shows six main centers of diversity, namely the Chimanimani Mountains in Manica, and Namaacha in Maputo, with seven taxa each; and Vila Coutinho in Tete, Mount Gorongosa, and Plateau in Sofala, and Manhiça and Inhaca Island in Maputo, all with six taxa. Therefore, the greatest diversity is found in the central and western areas of the country.

The maps of predicted changes in temperature (**Figure 4A**) and precipitation (**Figure 4B**) for 2061–2080 anticipate the high pressure imposed by climatic changes on Mozambique's biodiversity. The mean annual temperature will significantly increase, with expected changes ranging from 4.5°C to 6°C. The highest increase is predicted to occur in the western provinces, where the only population of *V. schlechteri* is found, a species of high priority level. *Vigna antunesii* and *V. juncea*, both of medium priority level, also occur only in this area. Two important diversity hotspots, namely Vila Coutinho in Tete, and the Chimanimani Mountains in Manica, are located in areas with a predicted increase of more than 5.5°C. The lowest increase in mean temperature is expected to occur near the coast.

Changes in precipitation are predicted to increase the annual mean in the north of the country and to decrease it in the central and southeast regions. The important hotspot of Chimanimani Mountains in Manica is among the most affected areas by such a decrease; this region includes all the records of *V. unguiculata* subsp. *pawekiae*, a taxon of high priority level. A similar decrease is predicted for the eastern region of Inhambane and Bazaruto Archipelago National Park, affecting many species distributed near the coast.

Approximately 9% of the occurrences of *Vigna* are within the boundaries of Protected Areas, including 10 (48%) of the 21 studied taxa. On the other hand, 32% of the occurrences are found in IPAs, comprising 12 (57%) of the 21 studied taxa (**Figure 5**; **Table 2**). The Gorongosa National Park (including Mount Gorongosa, a portion of the Rift Valley and the southern portion of the Cheringoma Plateau) in the province of Sofala hosts the highest number of taxa (*V. gazensis*, *V. luteola*, *V. schlechteri*, *V. unguiculata* subsp. *dekindtiana, V. unguiculata* subsp. *pubescens*, *V. unguiculata* subsp. *stenophylla*, *V. vexillata* var. *vexillata*, and *V. vexillata* var. *angustifolia*). The Chimanimani National Park (Manica province) includes two taxa (*V. gazensis* and *V. vexillata* var. *angustifolia*); the Bazaruto National Park (Inhambane province) includes two taxa (*V. marina* and *V. unguiculata* subsp. *tenuis*); and the Marromeu National Reserve (Sofala province) includes one single taxon (*V. unguiculata* subsp. *tenuis*).

Our results show that all the five taxa categorized as highly prioritary for conservation were found in IPAs, and four of them (*V. schlechteri*, *V. unguiculata* subsp. *pubescens*, *V. unguiculata* subsp. *stenophylla*, and *V. unguiculata* subsp. *tenuis*) were also found in Protected Areas (**Supplementary Figure 1**). Of the 14 taxa categorized as of medium priority for conservation, only four were in Protected Areas (*V. gazensis*, *V. luteola*, *V. marina*, and *V. unguiculata* subsp. *dekindtiana*); these, together with

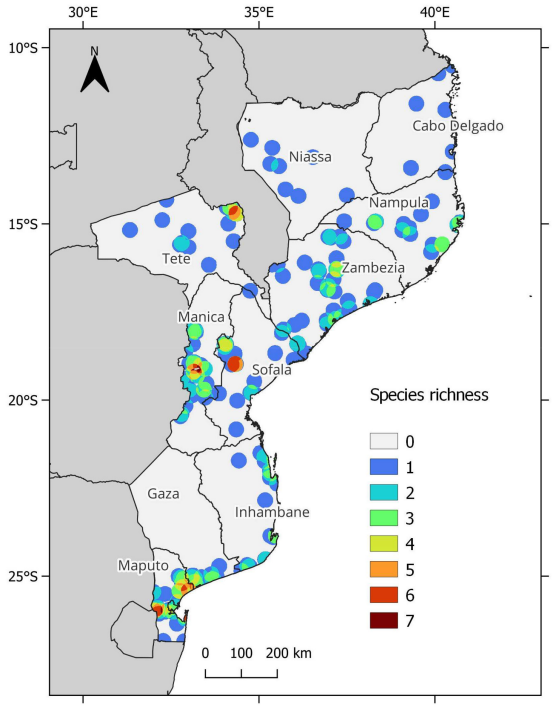


FIGURE 3
Vigna taxa richness in Mozambique (including subspecies), based on occurrence records by applying a circular buffer of 25 km around each occurrence point.

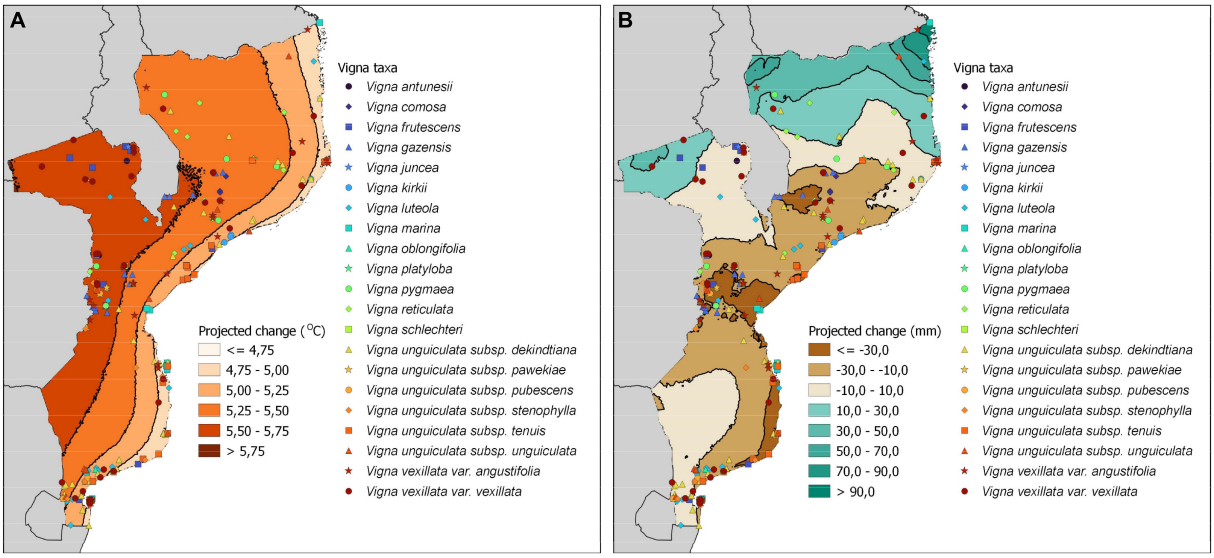


FIGURE 4
Occurrence of *Vigna* taxa in Mozambique with details of predicted changes in temperature **(A)** and precipitation **(B)** for 2061–2080, obtained with the shared socio-economic pathway 3-70 (SSP 3-7.0).

V. unguiculata subsp. *unguiculata*, were also found in IPAs (**Supplementary Figure 2**). The taxa of low priority for conservation (*V. vexillata* var. *angustifolia* and *V. vexillata* var. *vexillata*) were found in different Protected Areas and IPAs (**Supplementary Figure 3**).

3.3. *Ex situ* conservation

The data on the accessions available worldwide revealed that only 30 accessions of native *Vigna* collected in Mozambique are preserved in germplasm banks (**Table 3**). Of the 21 studied taxa, only eight (38%) had at least one accession preserved, being represented by fewer than ten accessions each. The taxa *V. unguiculata* subsp. *tenuis* and *V. vexillata* var. *vexillata* had the highest number of accessions (eight and seven, respectively), followed by *V. frutescens* with five accessions, *V. unguiculata* subsp. *unguiculata* with four accessions, *V. marina* and *V. unguiculata* subsp. *dekindtiana* with two accessions, and *V. unguiculata* subsp. *stenophylla* and *V. vexillata* var. *angustifolia* with one accession each.

Most of the 30 accessions are stored at the International Institute of Tropical Agriculture in Nigeria (33%), the rest being at Meise Botanic Garden in Belgium (23%), Centro Internacional de Agricultura Tropical in Colombia (17%), Australian Grains Genebank in Australia (17%), Australian Pasture Genebank in Australia (7%), and International Livestock Research Institute in Ethiopia (3%) (**Supplementary Table 1**). Notably, Mozambique national germplasm banks lack registered native *Vigna* accessions.

4. Discussion

4.1. Diversity of *Vigna* CWRs in Mozambique

The present study analyzed and detailed for the first time the diversity of the genus *Vigna* in Mozambique, showing that 21 native *Vigna* taxa occur in this country, with no endemic taxa being recorded. Hyde et al. (2022a) had only referred 19 *Vigna* taxa recorded for Mozambique, but the present study adds *V. oblongifolia* and *V. unguiculata* subsp. *dekindtiana* to the list, which is consistent with the new checklist of the plants of Mozambique (Odorico et al., 2022) and Maxted et al. (2004), respectively. In comparison to other surrounding African countries, where the occurrence of *Vigna* is recorded, Mozambique is surpassed by Tanzania (POWO, 2022), Zambia (Bingham et al., 2022), Angola (Catarino et al., 2021a), and Malawi (Hyde et al., 2022b) that have 44, 30, 28, and 24 taxa, respectively. On the other hand, with lower *Vigna* taxa richness, only Zimbabwe (Hyde et al., 2022c) and Botswana (Hyde et al., 2022d) have 19 and nine taxa, respectively.

The provinces with more taxa are Maputo, Manica, Zambezia, and Sofala, with more than 2,000 taxa recorded from different families. Except for Zambezia, these provinces are also among those with the highest number of occurrences of native *Vigna* taxa in Mozambique, unevenly distributed across the country. Specifically, the Chimanimani Mountains (Manica), Mount Gorongosa and Plateau (Sofala), Inhaca island (Maputo), Manhiça (Maputo), Namaacha (Maputo), and Vila Coutinho (Tete) are the centers of diversity for the *Vigna*

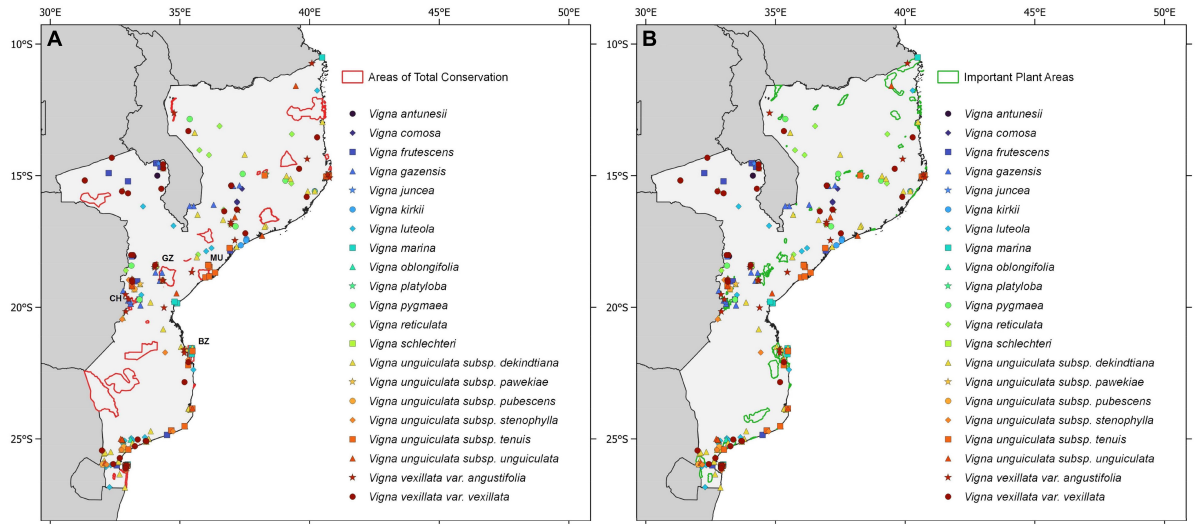


FIGURE 5
Occurrence of *Vigna* taxa in Mozambique with details of **(A)** Protected Areas of total conservation and **(B)** Important Plant Areas (IPAs). Green lines indicate the IPAs. Red lines indicate the Protected Areas and those with *Vigna* taxa are highlighted by the respective acronyms [BZ, Bazaruto National Park; CH, Chimanimani National Park; GZ, Gorongosa National (including Mount Gorongosa, a portion of the Rift Valley and the southern portion of the Cheringoma Plateau); MU, Marromeu National Reserve]. Only Protected Areas of total conservation were included, according to Mozambique's legislation (*Diário da República de Moçambique, law no. 205/2017, p.404*).

TABLE 2 Occurrence of *Vigna* taxa within Protected Areas and Important Plant Areas (IPAs) of Mozambique.

Acronym	Designation	Taxa
BZ	Bazaruto Archipelago National Park	*Vigna marina; Vigna unguiculata* subsp. *dekindtiana, Vigna unguiculata* subsp. *tenuis*
CH	Chimanimani National Park	*Vigna gazensis*, *Vigna vexillata* var. *angustifolia*
GZ	Gorongosa National Park	*Vigna gazensis*, *Vigna luteola*, *Vigna schlechteri*, *Vigna unguiculata* subsp. *dekindtiana*; *Vigna unguiculata* subsp. *pubescens*, *Vigna unguiculata* subsp. *stenophylla*, *Vigna vexillata* var. *vexillata*, *Vigna vexillata* var. *angustifolia*
MU	Marromeu National Reserve	*Vigna unguiculata* subsp. *tenuis*
IPAs	Important Plant Areas	*Vigna gazensis*, *Vigna luteola*, *Vigna marina*, *Vigna schlechteri*, *Vigna unguiculata* subsp. *dekindtiana*, *Vigna unguiculata* subsp. *pawekiae*, *Vigna unguiculata* subsp. *pubescens*, *Vigna unguiculata* subsp. *stenophylla*, *Vigna unguiculata* subsp. *tenuis*, *Vigna unguiculata* subsp. *unguiculata*, *Vigna vexillata* var. *angustifolia*, *Vigna vexillata* var. *vexillata*

genus in Mozambique. In this sense, Gorongosa National Park (including Mount Gorongosa, a portion of the Rift Valley and the southern portion of the Cheringoma Plateau) and Chimanimani Mountains were identified as two of the main centers of diversity for Mozambique flora (e.g., Darbyshire et al., 2019; Odorico et al., 2022). Gorongosa National Park, crossed by the Urema rift, is a site of unique characteristics. The climatic, geological, topographic, and edaphic framework promotes an enormous diversity of different plant communities, with woodlands and grasslands standing out as the main ecosystems (Coates-Palgrave et al., 2007). The Chimanimani Mountains are located on the border between Mozambique and Zimbabwe, on the Great Escarpment of Africa (Clark et al., 2017). These are part of the Eastern Afromontane biodiversity hotspot of Critical Ecosystem Partnership Fund (CEPF; Timberlake et al., 2016) and are also considered a subcenter of endemism associated with the larger Chimanimani-Nyanga Center (Van Wyk and Smith, 2001). In addition, it is where the highest number of endemics in any one site in southern tropical Africa is found (e.g., Wursten et al., 2017), being considered an IPA (Darbyshire et al., 2019). The topography and local geology determine the presence of a high number of plant species, most of which restricted to soils derived from quartzite sandstones that are very poor in nutrients (e.g., deficient in phosphorus) (Wursten et al., 2017; Cheek et al., 2018).

It should be noted that the patterns of distribution and species richness maps presented in this study may be spatially biased due to uneven botanical collection effort, as shown by several studies (e.g., Romeiras et al., 2014). A high floristic diversity is recognized in Mozambique; however, although a lot more studies have been carried out since 2014, there are still significant gaps of knowledge about Mozambique's flora

TABLE 3 Available accessions in worldwide germplasm banks of *Vigna* Crop Wild Relative (CWR) native in Mozambique, assessed through the Genesys Database (Genesys, 2022).

Taxa	Number of accessions
Vigna frutescens	5
Vigna marina	2
Vigna unguiculata subsp. *dekindtiana*	2
Vigna unguiculata subsp. *stenophylla*	1
Vigna unguiculata subsp. *tenuis*	8
Vigna unguiculata subsp. *unguiculata*	4
Vigna vexillata var. *angustifolia*	1
Vigna vexillata var. *vexillata*	7

(especially concerning the geographic coverage), but Flora Zambesiaca is now over 90% complete, and the recent work on IPAs (Darbyshire et al., 2019) and associated expansion of the Red List (IUCN, 2022) have resulted in major advances in the current knowledge. Historically, the botanical expedition to Mozambique that allowed the first major advance in the knowledge of local flora took place between 1942 and 1948, when 7,600 herbarium specimens were collected and several plant species were described (Conde et al., 2014). Furthermore, the period from 1963 to 1973 is added to the aforementioned dates, which was also recognized as very important in terms of botanical missions to Mozambique and, consequently, in increasing the country's botanical knowledge (Martins and Duarte, 2010). After these expeditions, due to the instability caused by two consecutive wars (i.e., war of independence and civil war), botanical studies in the country only restarted from 2000 onward, resulting in the publication of several works, namely: Da Silva et al. (2004), Harris et al. (2011), Timberlake et al. (2011), Wursten et al. (2017), Burrows et al. (2018), Darbyshire et al. (2019), Darbyshire et al. (2020), and Odorico et al. (2022). Even so, more studies and field surveys, to collect occurrence data for future studies on plant diversity and its conservation, are needed to discover new species in Mozambique (Darbyshire et al., 2019). Also, given that the boundaries of most of the proposed centers of endemism are still not fully ascertained, the contribution of botanical collection would be crucial to completely define them. This will make it possible to generate and improve tools for the sustainable use of genetic resources in Mozambique, namely the *Vigna* CWRs.

4.2. *In situ* and *ex situ* conservation of *Vigna* CWRs in Mozambique

Our results showed that 48% of the studied taxa occur in the Protected Areas of total conservation (i.e., *Vigna gazensis*, *V. luteola*, *V. marina*, *V. schlechteri*, *V. unguiculata* subsp. *dekindtiana*, *V. unguiculata* subsp. *pubescens*, *V. unguiculata* subsp. *stenophylla*, *V. unguiculata* subsp. *tenuis*, *V. vexillata* var. *angustifolia*, and *V. vexillata* var. *vexillata*); however, this should be regarded as preliminary information and further field confirmation is needed. It should be noted that one of the high conservation priority taxa, *V. unguiculata* subsp. *pawekiae*, is not protected *in situ*.

The protected area with the highest number of *Vigna* CWRs taxa (i.e., eight taxa) is Gorongosa Natural Park, considered an important area of plant diversity by several previous studies, housing more than 500 taxa (e.g., Da Silva et al., 2004; Müller et al., 2012; Darbyshire et al., 2019). Moreover, this site holds the highest number of *Vigna* taxa with conservation priority (three of high and three of medium priority).

The IPAs of Mozambique highlight key sites for protecting plant diversity and rare and threatened species nationally (Darbyshire et al., 2017), and 15 of them are currently known to support 12 taxa of *Vigna* CWRs, of which five are classified as of medium priority (*V. gazensis*, *V. luteola*, *V. marina*, *V. unguiculata* subsp. *dekindtiana*, *V. unguiculata* subsp. *unguiculata*) and five as of high priority for conservation (*V. unguiculata* subsp. *pubescens*, *V. unguiculata* subsp. *pawekiae*, *V. unguiculata* subsp. *stenophylla*, *V. schlechteri*, and *V. unguiculata* subsp. *tenuis*).

The coverage of plant diversity in Mozambique's Protected Areas network has some gaps. This is supported by the fact that less than 50% of Mozambique's IPAs network coincides with currently Protected Areas (in terms of numbers of sites rather than area). In this respect, IPAs contain more *Vigna* CWRs than Protected Areas, despite their much smaller total area (less than 3% of Mozambique's total land area, in contrast with 30%, in terms of Protected Areas), as they are more targeted to plant diversity. In this sense, and in agreement with the presented data, it is recognized that the *in situ* conservation of CWRs in Protected Areas is scarce (Maxted and Kell, 2009).

One of the diversity hotspots for *Vigna* CWRs, with seven taxa, is located near Vila Coutinho (Tete province), outside Mozambique's network of Protected Areas and IPAs, and there is no conservation support. Although there are few studies on the flora of Mozambique that highlight this area as important in terms of floristic richness, according to Odorico et al. (2022) Tete is among the five regions with the highest species richness for plants. Nonetheless, specifically for endemic/near-endemic plants, it is the poorest in terms of number of taxa (Darbyshire et al., 2019). It is also worth noting that Vila Coutinho includes five taxa classified as of medium priority for conservation (*V. antunesii*, *V. frutescens*, *V. juncea*, *V. reticulata*, and *V. unguiculata* subsp. *dekindtiana*). Furthermore, our analyses revealed that this region will be one of the most affected by climate change in a medium to long term, namely by the increase of mean annual temperature.

Although Mozambique's terrestrial Protected Areas occupy around 30% of its territory (UNEP-WCMC and IUCN,

2022), our data indicate that they are inadequate to protect Mozambique's native CWRs at present. Therefore, management plans must be applied to ensure the sustainable conservation of CWRs *in situ* (Maxted et al., 2008). The establishment of genetic reserves located in the diversity centers of the genus *Vigna*, considering the different types of habitats to which they are adapted, can be an alternative to protect and conserve the diversity of these genetic resources (Dulloo et al., 2008).

The *ex situ* conservation of plant genetic resources in Mozambique, especially CWRs, has been very limited, with little effort to characterize, assess and preserve this genetic heritage. The number of *Vigna* accessions in world germplasm banks, whose place of origin was Mozambique, only represents eight of the 21 (38%) taxa native to this region (Genesys, 2022). Similarly, and supporting the weak representation of the genus in germplasm banks, Catarino et al. (2021a), found only one accession of the 28 *Vigna* CWRs taxa from Angola, preserved in a single European germplasm bank. Furthermore, according to Castañeda-Álvarez et al. (2016), the number of accessions preserved in germplasm banks is unrepresentative of the diversity of CWRs that exist globally. Thus, the inherent genetic diversity of *Vigna* CWRs is neglected and therefore inaccessible to studies for future crop breeding programmes. This is of particular concern for taxa classified as highly prioritary for conservation. To complement the few collections currently conserved *ex situ*, additional collections should be considered a priority for underrepresented taxa and those of high conservation concern. Moreover, as recommended in studies such as that by Contreras-Toledo et al. (2019), all of the 21 taxa under study should be considered priorities for collection since they each have less than 50 preserved accessions in germplasm banks at present. The diversity hotspots detected for native *Vigna* in Mozambique can serve as targeted collection sites to maximize the efficient collection of a high number of taxa, providing the germplasm banks with representative samples of the *Vigna* genetic diversity of Mozambique. The *Vigna* diversity hotspots represent crucial areas for conservation actions, including further seed collection and habitat protection. Furthermore, the high priority taxa should receive a particular attention for increased *in situ* protection, especially given their close relationship with the commercial crops.

4.3. Potential use of *Vigna* CWRs for crop improvement

Studies on genetic diversity and relationships between *Vigna* taxa are scarce and, therefore, the Gene Pool (GP) for some of these is unknown. Due to this fact, we applied the concept of Taxonomic Group (TG) as a proxy for GP to assess the potential use for the improvement of the main crop, cowpea (*V. unguiculata* subsp. *unguiculata*) (**Supplementary Table 2**). However, the use of TG has some limitations, as stated by Catarino et al. (2021a), such as the fact that taxonomic relationships are unclear and there is little consensus between the various studies carried out previously for the genus (Maxted et al., 2004; Tomooka et al., 2011; Pasquet and Padulosi, 2013). There is an imperative need to carry out new studies to clarify the phylogenetic relationships between species.

In order to expand the cowpea breeding pool for tolerance to abiotic (e.g., drought and salinity) and biotic (e.g., pest resistance) stresses, the *Vigna* CWRs of GP/TG 1 and 2 represent the most closely related taxa and, therefore, are the most likely to be used as sources of genetic material associated with beneficial traits for plant breeding (e.g., Maxted et al., 2006). Thus, by studying taxa from different populations and different regions, these CWRs are important sources to evaluate adaptive genetic variation *in situ*, considering the various ecological conditions they are adapted to (Monteiro et al., 2013).

In Mozambique, the wild *Vigna* taxa are ecologically diverse, occupying various habitats, from arid to humid areas (Mackinder et al., 2001; **Supplementary Table 3**). Their growth across a range of environments makes them develop a natural resistance/tolerance to stressful abiotic conditions and guarantees their survival and integrity during their life cycle (e.g., Duque et al., 2013). So, the study of their natural genetic variability *in situ* would promote their role as potential donors of candidate genes involved in adaptive responses associated with these environments.

Previous studies on *Vigna* CWR revealed greater resistance to pests and diseases than to drought and salinity. In fact, according to van Zonneveld et al. (2020), potential resistance to pests and diseases was reported for 16 of the 20 taxa studied in the present work; with lower incidence only three taxa (V. *luteola*, *V. marina*, and *V. vexillata* var. *vexillata*) were reported as having potential of resistance to drought and salinity. Moreover, potential resistance to aphids and pod bugs was reported in *V. unguiculata* subsp. *stenophylla* (Timko and Singh, 2008; Badiane et al., 2014) and, to aphids, bruchid, cowpea mottle carmovirus, flower thrip, pod bug, pod borer, and yellow mosaic virus, in *V. vexillata* (Gomathinayagam et al., 1998; Timko and Singh, 2008; Boukar et al., 2015).

5. Conclusion

To the best of our knowledge, this is the first study focused on crop wild relative species of *Vigna* in Mozambique. The conservation gap analysis revealed that the main *Vigna* centers of diversity are located outside Protected Areas and that most of the Mozambican genetic diversity of *Vigna* is not represented in *ex situ* conservation programmes. Furthermore, the coverage of wild *Vigna* germplasm accessions from Mozambique in

worldwide germplasm banks is insufficient. A targeted seed collecting programme to support future management and *ex situ* conservation of plant genetic resources in Mozambique is therefore recommended. Further studies concerning the ecological and genetic diversity of the genus should also be carried out to assess the potential of those species for crop improvement under a climate change scenario (e.g., drought conditions). Overall, our data contribute to the understand the status of *Vigna* CWR taxa in Mozambique, providing new resources and knowledge for their sustainable use in food crop enhancement as well as for measures for future conservation programmes.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

MB and MR: conceptualization. MB, MD, and MR: methodology. MB and SC: formal analysis. MB, SB, and ID: investigation. MB: writing—original draft preparation. MB, SC, ID, SB, MM, MD, and MR: writing—review and editing. SB and MR: supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Fundação para a Ciência e Tecnologia (FCT) and also by the research unit UID/AGR/04129/2020 (LEAF) through the project "MozDiVigna". MB was supported by the FCT grant (UI/BD/151188/2021).

Acknowledgments

The authors would like to acknowledge the support provided by Fundação para a Ciência e Tecnologia (FCT) and to the research units: UID/AGR/04129/2020 (LEAF) and UID/BIA/00329/2020 (cE3c) funded by Portuguese National Funds through FCT, Portugal. Also, to all the herbarium collection institutes are acknowledged for sharing research material.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.1057785/full#supplementary-material

References

Badiane, F. A., Diouf, M., and Diouf, D. (2014). "Cowpea," in *Broadening the genetic base of grain legumes*, eds M. Singh, I. S. Bisht, and M. Dutta (New Delhi: Springer), 95–114. doi: 10.1007/978-81-322-2023-7_5

Bandeira, S. O., Marconi, L., and Barbosa, F. (1996). "Preliminary study of threatened plants of Mozambique," in *The biodiversity of African plants*, eds L. J. G. van der Maesen, X. M. van der Burgt, and J. M. van Medenbach de Rooy (Dordrecht: Springer), 306–309. doi: 10.1007/978-94-009-0285-5_38

Barbosa, F. M., Cuambe, C. C., and Bandeira, S. O. (2001). Status and distribution of mangroves in Mozambique. *S. Afr. J. Bot.* 67, 393–398. doi: 10.1016/S0254-6299(15)31155-8

Bingham, M. G., Willemen, A., Wursten, B. T., Ballings, P., and Hyde, M. A. (2022). *Flora of Zambia: Genus page: Vigna*. Available online at: https://www.zambiaflora.com/speciesdata/genus.php?genus_id=776 (accessed November 30, 2022).

Boukar, O., Fatokun, C. A., Roberts, P. A., Abberton, M., Huynh, B. L., Close, T. J., et al. (2015). "Cowpea," in *Grain legumes*, ed. A. M. De Ron (New York, NY: Springer), 219–250. doi: 10.1007/978-1-4939-2797-5_7

Brilhante, M., Varela, E., P Essoh, A., Fortes, A., Duarte, M. C., Monteiro, F., et al. (2021). Tackling food insecurity in cabo verde Islands: The nutritional, agricultural and environmental values of the legume species. *Foods* 10:206. doi: 10.3390/foods10020206

Burgess, N., Hales, J. A., Underwood, E., Dinerstein, E., Olson, D., Itoua, I., et al. (2004). *Terrestrial ecoregions of Africa and Madagascar. A conservation assessment*. Washington, DC: Island Press.

Burrows, J. E., Burrows, S., Schmidt, E., Lotter, M., and Wilson, E. O. (2018). *Trees and shrubs Mozambique*. Cape Town: Print Matters Heritage.

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2:16022. doi: 10.1038/nplants.2016.22

Catarino, S., Rangel, J., Darbyshire, I., Costa, E., Duarte, M. C., and Romeiras, M. M. (2021a). Conservation priorities for African *Vigna* species: Unveiling Angola's diversity hotspots. *Glob. Ecol. Conserv.* 25:e01415. doi: 10.1016/j.gecco.2020.e01415

Catarino, S., Brilhante, M., Essoh, A. P., Charrua, A. B., Rangel, J., Roxo, G., et al. (2021b). Exploring physicochemical and cytogenomic diversity of African cowpea and common bean. *Sci. Rep.* 11:12838. doi: 10.1038/s41598-021-91929-2

Charrua, A. B., Padmanaban, R., Cabral, P., Bandeira, S., and Romeiras, M. M. (2021a). Impacts of the tropical cyclone Idai in Mozambique: A multi-temporal landsat satellite imagery analysis. *Remote Sens.* 13:201. doi: 10.3390/rs13020201

Charrua, A. B., Havik, P. J., Bandeira, S., Catarino, L., Ribeiro-Barros, A., Cabral, P., et al. (2021b). Food security and nutrition in Mozambique: Comparative study with bean species commercialised in informal markets. *Sustainability* 13:8839. doi: 10.3390/su13168839

Chapman, A. D., and Wieczorek, J. (2006). *Guide to best practices for georeferencing*. Copenhagen: Global Biodiversity Information Facility, 1–77.

Cheek, M., Chipanga, H., and Darbyshire, I. (2018). Notes on the plant endemics of the quartzitic slopes of Mt Chimanimani (Mozambique & Zimbabwe), and a new, critically endangered species, *Empogona jenniferae* (Rubiaceae-Coffeeae). *Blumea* 63, 87–92. doi: 10.3767/blumea.2018.63.01.08

Clark, V. R., Timberlake, J. R., Hyde, M. A., Mapaura, A., Palgrave, M. C., Wursten, B. T., et al. (2017). A first comprehensive account of floristic diversity and endemism on the Nyanga massif, Manica Highlands (Zimbabwe–Mozambique). *Kirkia* 19, 1–53.

Coates-Palgrave, M., Van Wyk, A. E., Jordaan, M., White, J. A., and Sweet, P. (2007). A reconnaissance survey of the woody flora and vegetation of the Catapú logging concession, Cheringoma district, Mozambique. *Bothalia* 37, 57–73. doi: 10.4102/abc.v37i1.303

Conde, P., Figueira, R., Saraiva, S., Catarino, L., Romeiras, M., and Duarte, M. C. (2014). The botanic mission to Mozambique (1942-1948): Contributions to knowledge of the medicinal flora of Mozambique. *Hist. Ciênc. Saúde Manguinhos* 21, 539–585. doi: 10.1590/S0104-59702014000200007

Contreras-Toledo, A. R., Cortés-Cruz, M., Costich, D. E., de Lourdes Rico-Arce, M., Brehm, J. M., and Maxted, N. (2019). Diversity and conservation priorities of crop wild relatives in Mexico. *Plant Genet. Res.* 17, 140–150. doi: 10.1017/S1479262118000540

CWR Project (2022). *The CWR project*. Available online at: https://www.cwrdiversity.org/project/ (accessed March 25, 2022).

Da Silva, M. C., Izidine, S., and Amude, A. B. (2004). *A preliminary checklist of the vascular plants of Mozamque*. Southern African botanical diversity network report (SABONET) No. 30. Pretoria: SABONET, 192.

Darbyshire, I., Anderson, S., Asatryan, A., Byfield, A., Cheek, M., Clubbe, C., et al. (2017). Important plant areas: Revised selection criteria for a global approach to plant conservation. *Biodivers. Conserv.* 26, 1767–1800. doi: 10.1007/s10531-017-1336-6

Darbyshire, I., Goyder, D. J., Wood, J. R., and Banze, A. E. (2020). Further new species and records from the coastal dry forests and woodlands of the Rovuma Centre of Endemism. *Plant Ecol. Evol.* 153, 427–445. doi: 10.5091/plecevo.2020.1727

Darbyshire, I., Richards, S., Osborne, J., Matimele, H., Langa, C., Datizua, C., et al. (in press). *The important plant areas of Mozambique*. Kew, VI: Royal Botanic Gardens.

Darbyshire, I., Timberlake, J., Osborne, J., Rokni, S., Matimele, H., Langa, C., et al. (2019). The endemic plants of Mozambique: Diversity and conservation status. *PhytoKeys* 136, 45–96. doi: 10.3897/phytokeys.136.39020

Dulloo, M. E., Labokas, J., Iriondo, J. M., Maxted, N., Lane, A., Laguna, E., et al. (2008). "Genetic reserve location and design," in *Conserving plant genetic diversity in protected areas*, eds J. M. Iriondo, N. Maxted, and M. E. Dulloo (Wallingford: CABI), 23–64. doi: 10.1079/9781845932824.0023

Duque, A. S., de Almeida, A. M., da Silva, A. B., da Silva, J. M., Farinha, A. P., Santos, D., et al. (2013). "Abiotic stress responses in plants: Unraveling the complexity of genes and networks to survive", in *Abiotic Stress—Plant Responses and Applications in Agriculture*, eds K. Vahdati, and C. Leslie (London: InTech), 49–101. doi: 10.5772/52779

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9, 1937–1958. doi: 10.5194/gmd-9-1937-2016

GBIF (2022a). *Free and open access to biodiversity data*. Copenhagen: GBIF.

GBIF (2022b). *GBIF occurrence download–Vigna*. Copenhagen: GBIF.

Genesys (2022). Genesys global portal on plant genetic resources. Menlo Park, CA: Genesys.

Gomathinayagam, P., Rathnaswamy, R., and Ramaswamy, N. M. (1998). Interspecific hybridization between *Vigna unguiculata* (L.) Walp. and *V. vexillata* (L.) A. Rich. Through in vitro embryo culture. *Euphytica* 102, 203–209. doi: 10.1023/A:1018381614098

Harris, T., Darbyshire, I., and Polhill, R. (2011). New species and range extensions from Mt Namuli, Mt Mabu and Mt Chiperone in northern Mozambique. *Kew Bull.* 66, 241–251. doi: 10.1007/s12225-011-9277-9

Hausfather, Z. (2019). *CMIP6: The next generation of climate models explained. Carbon brief*. Available online at: https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained (accessed May 28, 2022).

Heywood, V., Casas, A., Ford-Lloyd, B., Kell, S., and Maxted, N. (2007). Conservation and sustainable use of crop wild relatives. *Agric. Ecosyst. Environ.* 121, 245–255. doi: 10.1016/j.agee.2006.12.014

Hyde, M. A., Wursten, B. T., Ballings, P., and Coates Palgrave, M. (2022a). *Flora of Mozambique: Genus page: Vigna*. Available online at: https://www.mozambiqueflora.com/speciesdata/genus.php?genus_id=776 (accessed November 20, 2021).

Hyde, M. A., Wursten, B. T., Ballings, P., and Coates Palgrave, M. (2022b). *Flora of Malawi: Genus page: Vigna*. Available online at: https://www.malawiflora.com/speciesdata/genus.php?genus_id=776 (accessed November 30, 2022).

Hyde, M. A., Wursten, B. T., Ballings, P., and Coates Palgrave, M. (2022c). *Flora of Zimbabwe: Genus page: Vigna*. Available online at: https://www.zimbabweflora.co.zw/speciesdata/genus.php?genus_id=776 (accessed November 30, 2022).

Hyde, M. A., Wursten, B. T., Ballings, P., and Coates Palgrave, M. (2022d). *Flora of Botswana: Genus page: Vigna*. Available online at: https://www.botswanaflora.com/speciesdata/genus.php?genus_id=776 (accessed November 30, 2022).

INGC (2009). *Main report: INGC climate change report: Study on the impact of climate change on disaster risk in Mozambique*, eds K. Asante, G. Brundrit, P. Epstein, A. Fernandes, M. R. Marques, A. Mavume, et al. (Maputo: INGC).

IUCN (2022). *IUCN red list of threatened species. Version 2022.1*. Gland: IUCN.

Khaki Mponya, N., Chanyenga, T., Magos Brehm, J., and Maxted, N. (2021). In situ and ex situ conservation gap analyses of crop wild relatives from Malawi. *Genet. Resour. Crop Evol.* 68, 759–771. doi: 10.1007/s10722-020-01021-3

Mackinder, B., Pasquet, R., Polhill, R., and Verdcourt, B. (2001). "Leguminosae (Papilionoideae)," in *Flora Zambesiaca*, Vol. 3(Pt 5), eds G. V. Pope and R. M. Polhill (Kew, VI: Royal Botanic Gardens), 1–261.

Magos Brehm, J., Maxted, N., Martins-Loução, M. A., and Ford-Lloyd, B. V. (2010). New approaches for establishing conservation priorities for socio-economically important plant species. *Biodivers. Conserv.* 19, 2715–2740. doi: 10.1007/s10531-010-9871-4

Martins, E., and Duarte, M. C. (2010). "Caminhos da botânica tropical nos países lusófonos," in *Viagens e Missões científicas nos Trópicos*, eds A. Martins and T. Albino (Lisboa: Instituto de Investigação Científica Tropical), 128–132.

Maxted, N., and Kell, S. P. (2009). *Establishment of a global network for the. situ conservation of crop wild relatives: Status and needs*. Rome: FAO Commission on Genetic Resources for Food and Agriculture.

Maxted, N., Dulloo, E., Ford-Lloyd, B. V., Iriondo, J. M., and Jarvis, A. (2008). Gap analysis: A tool for complementary genetic conservation assessment. *Divers. Distrib.* 14, 1018–1030. doi: 10.1111/j.1472-4642.2008.00512.x

Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S., and Scholten, M. (2006). Towards a definition of a crop wild relative. *Biodivers. Conserv.* 15, 2673–2685. doi: 10.1007/s10531-005-5409-6

Maxted, N., Kell, S., Toledo, Á, Dulloo, E., Heywood, V., Hodgkin, T., et al. (2010). A global approach to crop wild relative conservation: Securing the gene pool for food and agriculture. *Kew Bull.* 65, 561–576. doi: 10.1007/s12225-011-9253-4

Maxted, N., Mabuza-Diamini, P., Moss, H., Padulosi, S., Jarvis, A., and Guarino, L. (2004). *An ecogeographic study African Vigna. Systematic and ecogeographic studies of crop Genepools 11*. Rome: International Plant Genetic Resources Institute.

Mba, C., and Ogbonnaya, F. C. (2022). "Utilizing plant genetic resources to develop climate resilient crops," in *Agricultural biotechnology, biodiversity and bioresources conservation and utilization*, eds O. O. Obembe, E. O. Ekundayo, A. S. Okoli, A. Gidado, C. O. Adetunji, A. B. Ibrahim, et al. (Boca Raton, FL: CRC Press), 373–404. doi: 10.1201/9781003178880-22

Meilleur, B. A., and Hodgkin, T. (2004). In situ conservation of crop wild relatives: Status and trends. *Biodivers. Conserv.* 13, 663–684. doi: 10.1023/B:BIOC.0000011719.03230.17

Ministry for the Coordination of Environmental Affairs (2014). *Fifth national report on the implementation of convention on biological diversity in Mozambique*. Maputo: Ministry for the Coordination of Environmental Affairs.

Monteiro, F., Romeiras, M. M., Batista, D., and Duarte, M. C. (2013). Biodiversity assessment of sugar beet species and its wild relatives: Linking ecological data with new genetic approaches. *Am. J. Plant Sci.* 4, 21–34. doi: 10.4236/ajps.2013.48A003

Müller, T., Mapaura, A., Wursten, B., Chapano, C., Ballings, P., and Wild, R. (2012). Vegetation survey of Mount Gorongosa. *Occas. Publ. Biodivers.* 23, 1–54.

Ng'uni, D., Munkombwe, G., Mwila, G., Gaisberger, H., Brehm, J. M., Maxted, N., et al. (2019). Spatial analyses of occurrence data of crop wild relatives (CWR) taxa as tools for selection of sites for conservation of priority CWR in Zambia. *Plant Genet. Resour.* 17, 103–114. doi: 10.1017/S1479262118000497

Odorico, D., Nicosia, E., Datizua, C., Langa, C., Raiva, R., Souane, J., et al. (2022). An updated checklist of Mozambique's vascular plants. *PhytoKeys* 189, 61–80. doi: 10.3897/phytokeys.189.75321

Pasquet, R. S., and Padulosi, S. (2013). "Genus vigna and cowpea (V. unguiculata [L.] Walp.) taxonomy: Current status and prospects" in *Proceedings of the fifth world cowpea conference on improving livelihoods in the cowpea value chain through advancement in science*, Dakar.

Pimentel, D., Wilson, C., McCullum, C., Huang, R., Dwen, P., Flack, J., et al. (1997). Economic and environmental benefits of biodiversity. *BioScience* 47, 747–757. doi: 10.2307/1313097

POWO (2022). Available online at: http://powo.science.kew.org (accessed October 28, 2021).

PROTA, (2022). Available online at: https://prota.prota4u.org/ (accessed November 10, 2021).

QGIS Development Team (2022). *QGIS Geographic information system. Open source geospatial foundation project*. Available online at: http://qgis.osgeo.org (accessed April 25, 2022).

Rocha, V., Duarte, M. C., Catarino, S., Duarte, I., and Romeiras, M. M. (2021). Cabo Verde's Poaceae flora: A reservoir of crop wild relatives diversity for crop improvement. *Front. Plant Sci.* 47:630217. doi: 10.3389/fpls.2021.630217

Romeiras, M. M., Figueira, R., Duarte, M. C., Beja, P., and Darbyshire, I. (2014). Documenting biogeographical patterns of African timber species using herbarium records: A conservation perspective based on native trees from Angola. *PLoS One* 9:e103403. doi: 10.1371/journal.pone.0103403

Serea, R. (2018). *Google earth pro 7.3.4.8573*. Available online at: https://www.neowin.net/news/google-earth-pro-7348573/ (accessed April 15, 2022).

Simon, M. V., Benko-Iseppon, A. M., Resende, L. V., Winter, P., and Kahl, G. (2007). Genetic diversity and phylogenetic relationships in *Vigna* Savi germplasm revealed by DNA amplification fingerprinting. *Genome* 50, 538–547. doi: 10.1139/G07-029

Singh, B. (ed.) (2020). *Cowpea: The food legume of the 21st century*, Vol. 164. New York, NY: John Wiley & Sons.

Singh, B. B., Raj, D. M., Dashiell, K. E., and Jackai, L. E. N. (1997). *Advances in cowpea research, copublication of international institute of tropical agriculture (IITA) and Japan international research centre for agricultural sciences (JIRCAS)*. Ibadan: IITA.

Sintayehu, D. (2018). Impact of climate change on biodiversity and as-sociated key ecosystem services in Africa: A systematic review. *Ecosyst. Health Sustain.* 4, 225–239. doi: 10.1080/20964129.2018.1530054

Snapp, S., Rahmanian, M., Batello, C., and Calles, T. (2018). *Pulse crops for sustainable farms in Sub-Saharan Africa*. Rome: Food and Agriculture Organization of the United Nations (FAO). doi: 10.18356/6795bfaf-en

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Thornton, P. K., and Herrero, M. (2015). Adapting to climate change in the mixed crop and livestock farming systems in Sub-Saharan Africa. *Nat. Clim. Chang.* 5, 830–836. doi: 10.1038/nclimate2754

Timberlake, J. R., Darbyshire, I., Wursten, B., Hadj-Hammou, J., Ballings, P., Mapaura, A., et al. (2016). *Chimanimani Mountains: Botany and conservation*. Report produced under CEPF Grant, 63512. Kew, VI: Royal Botanic Gardens.

Timberlake, J., Goyder, D., Crawford, F., Burrows, J., Clarke, G. P., Luke, Q., et al. (2011). Coastal dry forests in northern Mozambique. *Plant Ecol. Evol.* 144, 126–137. doi: 10.5091/plecevo.2011.539

Timko, M. P., and Singh, B. B. (2008). "Cowpea, a multifunctional legume," in *Genomics of tropical crop plants*, eds P. H. Moore and R. Ming (New York, NY: Springer), 227–258. doi: 10.1007/978-0-387-71219-2_10

Tomooka, N., Kaga, A., Isemura, T., and Vaughan, D. (2011). *Vigna. Wild crop relatives: Genomic and breeding resources, legume crops and forages*. Berlin: Springer, 291–311. doi: 10.1007/978-3-642-14387-8_15

Uamusse, M. M., Aljaradin, M., Nilsson, E., and Persson, K. M. (2017). Climate change observations into hydropower in Mozambique. *Energy Procedia* 138, 592–597. doi: 10.1016/j.egypro.2017.10.165

UNEP-WCMC, and IUCN (2022). *Protected planet: The world database on protected areas (WDPA) and world database on other effective area-based conservation measures (WD-OECM)*. Cambridge: UNEP-WCMC.

USDA, Agricultural Research Service, and National Plant Germplasm System (2022). *Germplasm resources information network (GRIN Taxonomy)*. Beltsville, MD: National Germplasm Resources Laboratory.

Van Wyk, A. E., and Smith, G. F. (2001). *Regions of floristic endemism in Southern Africa*. Hatfield: Umdaus Press.

van Zonneveld, M., Rakha, M., Chou, Y. Y., Chang, C. H., Yen, J. Y., Schafleitner, R., et al. (2020). Mapping patterns of abiotic and biotic stress resilience uncovers conservation gaps and breeding potential of *Vigna* wild relatives. *Sci. Rep.* 10:2111. doi: 10.1038/s41598-020-58646-8

Vidigal, P., Romeiras, M. M., and Monteiro, F. (2019). "Crops diversification and the role of orphan legumes to improve the Sub-Saharan Africa farming systems," in *Sustainable crop production*, ed. M. Hasanuzzaman (Rijeka: IntechOpen). doi: 10.5772/intechopen.88076

Vincent, H., Hole, D., and Maxted, N. (2022). Congruence between global crop wild relative hotspots and biodiversity hotspots. *Biol. Conserv.* 265:109432. doi: 10.1016/j.biocon.2021.109432

WorldClim (2020a). *Historical climate data*. Available online at: https://www.worldclim.org/data/worldclim21.html (accessed May 19, 2022).

WorldClim (2020b). *Future climate data*. Available online at: https://www.worldclim.org/data/cmip6/cmip6_clim2.5m.html (accessed May 19, 2022).

Wursten, B., Timberlake, J., and Darbyshire, I. (2017). The Chimanimani Mountains: An updated checklist. *Kirkia* 19, 70–100.

Yadav, M. R., Kumar, R., Parihar, C. M., Yadav, R. K., Jat, S. L., Ram, H., et al. (2017). Strategies for improving nitrogen use efficiency: A review. *Agric. Rev.* 38, 29–40. doi: 10.18805/ag.v0iOF.7306

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 February 2023
DOI 10.3389/fpls.2022.1057317

OPEN ACCESS

EDITED BY
Tofazzal Islam,
Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

REVIEWED BY
Anamika Pandey,
Selçuk University, Türkiye
Romain A. Guyot,
IRD UMR232 Diversité, adaptation, développement des plantes (DIADE), France
Arun Kumar C. Huded,
Central Coffee Research Institute, India

*CORRESPONDENCE
Aaron P. Davis
a.davis@kew.org

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 29 September 2022
ACCEPTED 01 December 2022
PUBLISHED 17 February 2023

CITATION
Davis AP, Kiwuka C, Faruk A, Mulumba J and Kalema J (2023) A review of the indigenous coffee resources of Uganda and their potential for coffee sector sustainability and development. *Front. Plant Sci.* 13:1057317. doi: 10.3389/fpls.2022.1057317

A review of the indigenous coffee resources of Uganda and their potential for coffee sector sustainability and development

Aaron P. Davis[1*†], Catherine Kiwuka[2†], Aisyah Faruk[3], John Mulumba[2] and James Kalema[4†]

[1]Crops & Global Change, Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom, [2]Plant Genetic Resources Centre, National Agricultural Research Organization, Entebbe, Uganda, [3]Partnerships (Conservation), Millennium Seed Bank (Royal Botanic Gardens, Kew), Wakehurst, Sussex, United Kingdom, [4]Department of Plant Sciences, Microbiology and Biotechnology, College of Natural Sciences, Makerere University, Kampala, Uganda

Uganda is a major global coffee exporter and home to key indigenous (wild) coffee resources. A comprehensive survey of Uganda's wild coffee species was undertaken more than 80 years ago (in 1938) and thus a contemporary evaluation is required, which is provided here. We enumerate four indigenous coffee species for Uganda: *Coffea canephora*, *C. eugenioides*, *C. liberica* (var. *dewevrei*) and *C. neoleroyi*. Based on ground point data from various sources, survey of natural forests, and literature reviews we summarise taxonomy, geographical distribution, ecology, conservation, and basic climate characteristics, for each species. Using literature review and farm survey we also provide information on the prior and exiting uses of Uganda's wild coffee resources for coffee production. Three of the indigenous species (excluding *C. neoleroyi*) represent useful genetic resources for coffee crop development (e.g. *via* breeding, or selection), including: adaptation to a changing climate, pest and disease resistance, improved agronomic performance, and market differentiation. Indigenous *C. canephora* has already been pivotal in the establishment and sustainability of the robusta coffee sector in Uganda and worldwide, and has further potential for the development of this crop species. *Coffea liberica* var. *dewevrei* (excelsa coffee) is emerging as a commercially viable coffee crop plant in its own right, and may offer substantial potential for lowland coffee farmers, i.e. in robusta coffee growing areas. It may also provide useful stock material for the grafting of robusta and Arabica coffee, and possibly other species. Preliminary conservation assessments indicate that *C. liberica* var. *dewevrei* and *C. neoleroyi* are at risk of extinction at the country-level (Uganda). Adequate protection of Uganda's humid forests, and thus its coffee natural capital, is identified as a conservation priority for Uganda and the coffee sector in general.

KEYWORDS

genetic resources, climate change, Uganda, excelsa coffee, conservation, robusta coffee, coffee, Crop Wild Relative (CWR)

Introduction

Uganda is the world's seventh largest exporter of coffee, and Africa's second largest exporter, after Ethiopia. In 2019/20 Uganda exported c. 330,540 metric tons (International Coffee Organization (ICO), 2022) of robusta (*Coffea canephora*) and Arabica (*C. arabica*) coffee, at an estimated ratio of around 4:1, respectively (Uganda Coffee Development Authority (UCDA), 2017). Uganda is now the fourth largest robusta producer in the world, after Vietnam, Brazil and Indonesia (Uganda Coffee Development Authority (UCDA) 2017). Coffee accounts for c. 15% of Uganda's annual export revenue, with c. 4.2% of the population (1.7 m people) engaged in coffee farming (Uganda Coffee Development Authority (UCDA), 2017), and c. 20% (8 m people) working in the coffee sector (Kiwuka et al., 2021). Despite its success, the Ugandan coffee sector faces major challenges, which are set to accelerate over the coming decades, due to climate change and other disruptive influences. Uganda is fortunate, however, in possessing key wild (indigenous) coffee genetic resources, which offer promise for coffee crop development, climate-resilience potential (Kiwuka et al., 2021) and commercial enrichment.

Wild coffee (genetic) resources, both from within the two major crop species, *C. arabica* and *C. canephora*, and other species, have played a vital role in sustaining coffee production (farming) and thus the sector as whole (Davis et al., 2019). Examples include the use of wild material for: coffee berry disease (CBD; *Colletotrichum kahawae* J.M.Walter & Bridge) resistance for Ethiopian *C. arabica* (Yonas et al., 2014); coffee wilt disease (CWD; *Gibberella xylarioides* R. Heim & Sacca) resistance for Ugandan *C. canephora* (Kiwuka et al., 2021; Mulindwa et al., 2022); coffee leaf rust (CLR; *Hemileia vastatrix* Berk. & Broome) resistance, globally, for *C. arabica*, through crossing with *C. canephora* (Clarindo et al., 2013; Avelino et al., 2015) and *C. liberica* (Narasimhaswamy, 1960; Surya Prakash et al., 2002); and coffee leaf miner (*Perileucoptera coffeella* Méneville) resistance (Medina Filho et al., 1977a; Medina Filho et al., 1977b) and drought tolerance (Grisi et al., 2008; Melo et al., 2014; Carvalho et al., 2017) in *C. arabica*, through crossing with *C. racemosa* (Davis et al., 2021a). It is worth noting that wild species were used to sustain the global coffee industry in response to the devasting influence of CLR at the end of nineteenth century, firstly using *C. liberica*, from c. 1875–1900, and then *C. canephora* from the early 1900s onwards (McCook, 2014; Davis et al., 2019; McCook, 2019; Davis et al., 2022). Other coffee species are exported on a small-scale, including: *C. congensis* (Congo coffee) and particularly the hybrid 'congusta' (Bharatha Nandhini et al., 2013), *C. eugenioides* (see main text), *C. racemosa* and *C. zanguebariae* (Ibo or Zanzibar coffee) (Davis et al., 2021a). *Coffea stenophylla* was once exported from Upper West Africa (Davis et al., 2020), and may have potential as a crop plant on the basis of being able to provide an Arabica-like flavour at much higher temperatures compared to *C. arabica* (Davis et al., 2021b). Many species are used locally, across Africa, the Indian Ocean Islands (Madagascar and the Mascarene Islands) and Asia, as a substitute for *C. arabica* (Cheney, 1925; Davis et al., 2006). There could be further potential for coffee crop plant development (Davis et al., 2019) amongst the 130 species of *Coffea* now known to science (Davis and Rakotonasolo, 2021).

The most recent (and only) review of Ugandan wild coffee genetic resources was undertaken by A.S. Thomas in 1938, and published six years afterwards (Thomas, 1944). Thomas (1944) included four species in his review *C. eugenioides*, *C. canephora*, *C. excelsa* (now known as *C. liberica* var. *dewevrei*), and *C. spathicalyx*, although the last of these species has since been transferred to the genus *Calycosiphonia* (*C. spathicalyx*) (Davis et al., 2006). Despite the abundance of useful information in the review of Thomas (1944), much has changed since 1938 and more data is now available. In this contribution, we undertake a contemporary review of indigenous coffee resources, with an emphasis on their potential for coffee sector sustainability. We use ground point data from various sources (herbarium specimens and occurrence databases), survey of natural forests (2015 to 2022), and literature reviews, to summarise the taxonomy (and common names), geographical distribution, ecology, conservation, and basic climate characteristics, for each species. Using literature review and farm survey (2015 to 2022), we also provide information on the prior, existing, and potential future uses of Uganda's wild coffee resources for coffee production.

Our objectives were to review: (1) the eco-geography and climate requirements of Uganda's indigenous coffee species; (2) the extinction risk of these species at the country level; and (3) the application of this natural capital for the development and sustainability of the coffee sector.

Methods

Use of scientific names

Scientific names follow the accepted nomenclature for *Coffea*, based on peer-reviewed taxonomic and systematic research, as summarised in global plant name checklists (Govaerts et al., 2022). Synonyms for species names and other taxa are not included here, but are available from other sources (Davis et al., 2006; Govaerts et al., 2022).

Data for mapping, climate profiling and conservation assessments

Occurrence data points derived from herbarium specimens and field surveys (see below) were used to provide the data for the production of distribution maps and basic climate profiling analyses for *C. canephora*, *C. eugenioides*, *C. liberica* var. *dewevrei* and *C. neoleroyi*. We consulted herbarium specimen records from seven herbaria (BM, BR, K, MHU, P, WAG); herbarium codes follow (Holmgren et al., 1990; Thiers, 2019). Location data were georeferenced (if lacking coordinates), manually checked for geolocation accuracy (1 km or less) using GoogleEarth® and corrected if necessary. The herbarium and field surveys provided a dataset of 583 records, comprising, 275 for *C. canephora*, 198 for *C. eugenioides*, 109 for *C. liberica* var. *dewevrei* and 1 for *C. neoleroyi*.

Fieldwork and other data

Study of wild populations of *C. canephora*, *C. liberica* var. *dewevrei*, and *C. eugenioides*, and farm study visits for *C. canephora* and *C. liberica* var. *dewevrei*, were undertaken between 2015 and 2022. Location, habitat

and ecology data were collected from forest sites, and basic agronomy observations were made during the farm visits. Herbarium specimens (see above) and literature were consulted for additional information (including habitat, vegetation, uses and vernacular names). Information on the global distribution of coffee species was taken from published sources (Davis et al., 2006; Davis et al., 2019).

Mapping

A total of 583 data records were used to produce the distribution maps (Figure 1) for the four species (see Data for Mapping, Climate Profiling and Conservation Assessments). The maps were produced in QGIS 3.16 (QGIS Development Team, 2022), using the ESRI Gray (light) basemap available through the QuickMapServices 19.11.1 version plugin (NextGIS, 2019) and administrative area boundaries from GDAM version 1.0 (https://gadm.org/). For mapping the protected areas, we used the World Database on Protected Areas data, obtained *via* the Protected Planet portal (https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA) [accessed November 2022].

Conservation assessments

To produce preliminary national conservation assessments, we followed the IUCN Red List of Threatened Species criteria (IUCN Standards and Petitions Subcommittee, 2022). Area-based conservation metrics were generated from the mapping data set of 583 records (see above) using GeoCat (Bachman et al., 2011) GeoCAT (kew.org) [accessed 1 September 2022] with default settings (Area of Occurrence (AOO) cell width of 2 km).

Climate profiling analyses

We resampled all specimen data to remove duplicates within 1 km of each other, reducing the total number of records used from 583 to 176 (85 for *C. canephora*, 80 for *C. eugenioides*, 10 for *C. liberica* var. *dewevrei* and 1 for *C. neoleroyi*. To understand the fundamental climatic requirements, the statistics package R (R Core Team, 2020) was used to sample specimen data against 19 Bioclim variables (Busby, 1991) from the CHELSA dataset (Karger et al., 2017). For our overview of climatic parameters, we selected the following three Bioclims: Bio1 Annual mean temperature, Bio12 Annual Precipitation, and Bio15 Precipitation Seasonality. Of the 19 Bioclim variables (Busby, 1991), these three have been shown to provide a pragmatic summary of basic climate requirements for *Coffea* (Davis et al., 2021a; Davis et al., 2021b); and are included amongst the key drivers of modelled coffee distribution (Moat et al., 2017; Moat et al., 2019). Scatter (Figure 2) and density (Figure 3) plots were plotted using R (R Core Team, 2020), using the ggplot2 (Wickham, 2016) and ggpubr packages (Kassambara, 2020). These modelling methods have been shown to provide climate metrics that are similar to those provided for coffee species in cultivation (including farmed conditions) and in the wild, produced by direct measurement and other means (Davis et al., 2021b). For validation purposes, our modelled mean annual temperatures (from Bio1), total annual precipitation (Bio12) and precipitation seasonality (Bio15), were compared against publicly available monthly mean temperature precipitation charts for Uganda and published data for cultivated *C. canephora* (DaMatta and Ramalho, 2006; Kath et al., 2020; Venancio et al., 2020); published data are not available for the three other species studied here.

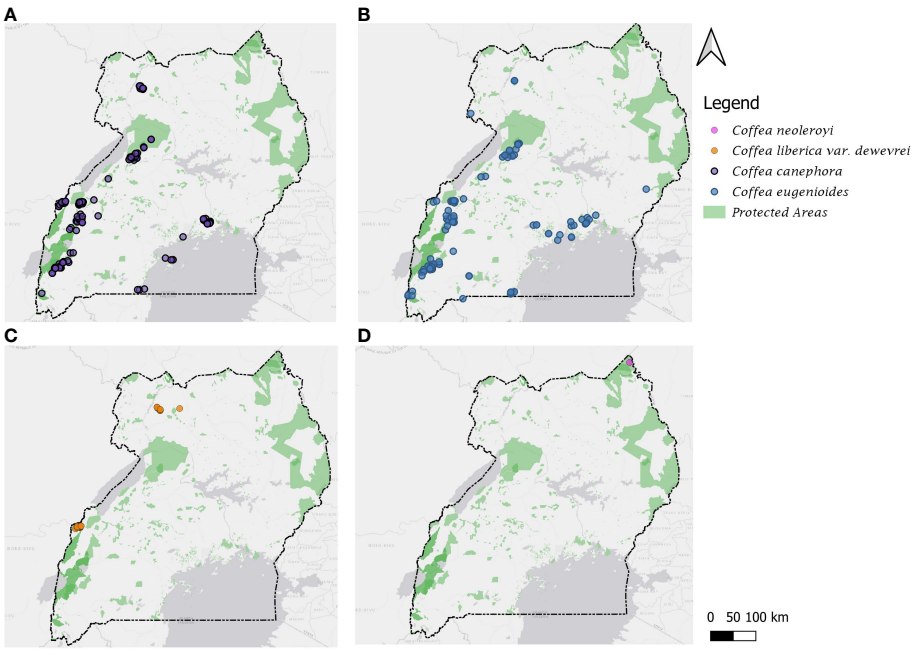


FIGURE 1
Distribution maps for the four indigenous *Coffea* species of Uganda. **(A)** *C. canephora*; **(B)** *C. eugenioides*; **(C)** *C. liberica* var. *dewevrei*; **(D)** *C. neoleroyi*. Protected areas based on World Database on Protected Areas (see Methods).

10.3389/fpls.2022.1057317

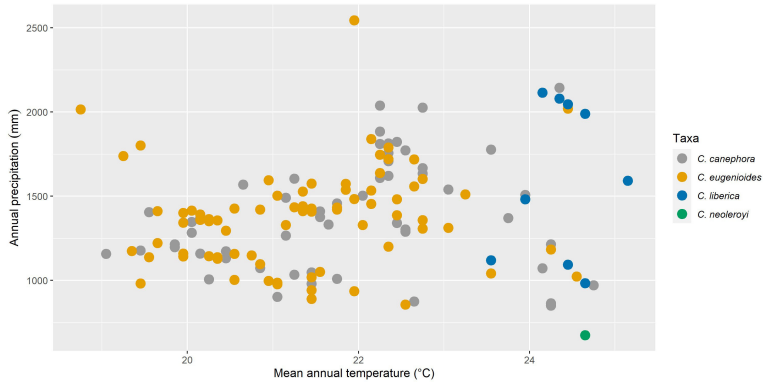


FIGURE 2
Scatter plots of modelled annual mean temperature (CHELSA Bio1) vs. total mean annual precipitation (CHELSA Bio12) for four indigenous Ugandan *Coffea* species (mean values in parentheses): *C. canephora* (21.8°C/1,389 mm); *C. eugenioides* (21.3°C/1,370 mm); *C. liberica* var. *dewevrei* (24.4°C/1,560 mm); and *C. neoleroyi* (data not available).

Results

Survey of wild (indigenous) Ugandan coffee species

We enumerate four indigenous (wild) coffee species for Uganda: *C. canephora*, *C. eugenioides*, *C. liberica* var. *dewevrei* and *C. neoleroyi*. For each species, this survey includes: the global distribution and global IUCN conservation assessment; and for Uganda, the distribution, ecology, elevation, preliminary conservation assessment, protected area occurrence, common names, uses (other than beverage) and miscellaneous notes.

Key to Ugandan coffee (*Coffea*) species

A. Leaves deciduous and/or restricted to short branches; flowers appearing terminal on short shoots; corolla tube much longer than wide …….. 4. *C. neoleroyi*

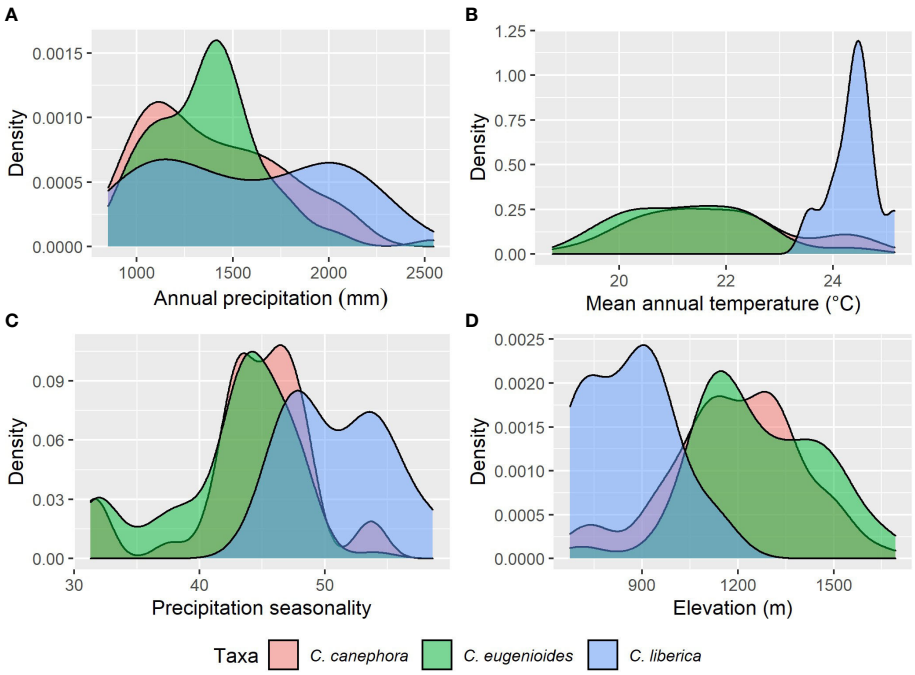


FIGURE 3
Density plots based for key climate variables and elevation (see Methods) for *C. canephora*, *C. eugenioides*, and *C. liberica* var. *dewevrei*. **(A)** total mean annual precipitation (mm year) from CHELSA Bio1; **(B)** mean annual temperature (°C) from CHELSA Bio12; **(C)** precipitation seasonality from CHELSA Bio15; **(D)** elevation (m). Summary statistics: (mean values in parentheses): *C. canephora* (21.8°C/1,389 mm/44/1187 m); *C. eugenioides* (21.3°C/1,370 mm/42/1261 m); and *C. liberica* var. *dewevrei* (24.4°C/1,560 mm/51/857 m). See Table 1.

10.3389/fpls.2022.1057317

TABLE 1 Basic climate profiling and elevations for three indigenous Ugandan coffee species, based on location records for Ugandan distributions, with comparative data from published sources.

Species	Min./max. mean values (data points); and mean of all data points (bold)	Mean annual temperature (°C)	Mean total annual precipitation (mm/year)	Precipitation seasonality	Elevation (m)
		CHELSA Bio1	CHELSA Bio12	CHELSA Bio15	
C. canephora	Min.	18.7	851	31	675
	Max.	24.7	2144	54	1666
	Mean	**21.8**	**1389**	**44**	**1187**
C. eugenioides	Min.	18.7	857	31	700
	Max	24.6	2544	54	1693
	Mean	**21.3**	**1370**	**42**	**1261**
C. liberica var. *dewevrei*	Min.	23.5	983	46	686
	Max	25.2	2115	58	1118
	Mean	**24.4**	**1560**	**51**	**857**
Published mean values for annual temperature and precipitation (Davis et al., 2021b); and precipitation seasonality (Davis et al., 2021a).					
C. arabica		**18.7**	**1614**	**58**	
C. canephora		**23.7**	**1601**	**56**	
C. liberica sensu lato		**23.9**	**1699**		

Mean minimum, mean maximum and mean annual temperatures (CHELSA Bio1); mean total annual rainfall (CHELSA Bio12), and precipitation seasonality (CHELSA Bio15), from the CHELSA modelled dataset (Karger et al., 2017). Numbers in bold indicate mean of all data values. Data for *C. neoleroyi* not given (single datapoint only).

A. Leaves evergreen, distributed evenly along the side branches; flowers distinctly axillary; corolla tube shorter or not much longer than wide …… B

B. Leaves 12–40 × 4.5–22 cm, with 8–17 pairs of secondary veins C

B. Leaves 3–12 × 1–7 cm, with 5–7 pairs of secondary veins 2. *C. eugenioides*

C. Leaves thick (rather leathery), domatia present and usually obvious, on the secondary vein (at the base) or in the secondary vein-midrib axil …… 3. *C. liberica* var. *dewevrei*

C. Leaves thin (paper-like to slightly leathery), domatia absent or inconspicuous, if present in the secondary vein-midrib axil …….. 1. *C. canephora*

1. *Coffea canephora* Pierre ex A.Froehner, Notizbl. Bot. Gart. Berlin-Dahlem 1: 237 (1897) Figures 4A–D, 5C.

Global distribution: west Tropical Africa (western Ghana, Guinea, Ivory Coast, Liberia [inferred or observed; no herbarium specimen data known], Nigeria); west-central Tropical Africa (Cabinda, Cameroon, Congo, Central African Republic, Democratic Republic of Congo, Gabon); north-east Angola. Tropical Africa (southern South Sudan); east Tropical Africa (Tanzania, Uganda); south Tropical Africa (northern Angola). The exact limit of natural distribution is difficult to ascertain owing to introduction and naturalisation. Widely cultivated as robusta coffee across the tropical belt of the world and frequently as Conilon in Brazil; naturalised in Tropical Africa and other tropical areas (not listed here).

Global IUCN conservation assessment: Least Concern (LC) (Chadburn and Davis, 2017a).

Information for Uganda

Distribution: Throughout Uganda but mainly in eastern and western parts of the country. Uganda is a centre of diversity for *C. canephora* (Gomez et al., 2009; Cubry et al., 2013; Merot-L'anthoene et al., 2019; Kiwuka et al., 2021), and is the only country on the eastern side of the Great Rift Valley that holds substantial wild populations of this species. Figure 1A.

Ecology: An exclusively forest-dwelling species, found in the understorey of humid, evergreen forest (rainforest), occurring with a wide range of dominant tree species. Often occurring in the same forests and in close proximity to *C. eugenioides* (Figures 1, 4A) and at low elevations with *C. liberica* var. *dewevrei* (Figure 1).

Elevation: 655–1570 m (observed and recorded); 675–1660 m (modelled).

Preliminary country-level IUCN conservation assessment: Least Concern (LC). The Extent of Occurrence (EOO) is 128,922 km^2 (LC); and Area of Occurrence (AOO) is 392 km^2 (Endangered (EN)). Whilst the EOO for *C. canephora* in Uganda is substantial, the AOO calculation falls within the limits for EN, indicating that the preservation of this species should not be taken for granted and that careful monitoring is required. Over most the eastern part of its distribution in Uganda (Figure 1A), *C. canephora* is now restricted to smaller and increasingly fragmented forests; if these populations were extirpated, the EOO would be severely reduced (c. 33,102 km^2) and the EOO-based rating increased to Near Threatened (NT). Populations in the larger protected areas appear to be healthy, with a high density of individuals, but encroachment, deforestation and disturbance in some protected areas (e.g. Zoka and Itwara Central Forest Reserves) are negatively affecting AOO, number of mature individuals, habitat quality, and population health.

10.3389/fpls.2022.1057317

FIGURE 4
Coffea canephora (robusta coffee) and *C. eugenioides*. **(A)** Dr Robert Acidri: right hand holding *C. canephora* and left hand holding *C. eugenioides*; **(B)** *C. canephora*, fruiting shoot (immature); **(C)** *C. canephora*, leaves; **(D)** *C. canephora*, unripe fruiting branch (immature); **(E)** *C. eugenioides*, fruiting branch (immature); **(F)** *C. eugenioides*, larger leaved variant. **(A–C)**, **(E, F)**: Itwara Forest (wild), western Uganda; **(D)** farmed in Masaka, central Uganda.

FIGURE 5
Seeds (unroasted coffee beans) of three Ugandan coffee species, with some cultivated species for size comparison. **(A)** *C. liberica* var. *dewevrei* (excelsa coffee), cultivated in central Uganda; **(B)** *C. eugenioides*, cultivated in Kampala, Uganda (1921), from RBG Kew Economic Botany Collection; **(C)** *C. canephora* (robusta coffee), cultivated in Uganda; **(D)** *C. arabica* (Arabica coffee), cultivated in Ethiopia; **(E)** *C. liberica* var. *liberica* (Liberica or Liberian coffee), cultivated in Malaysia; **(F)** *C. canephora* (robusta coffee), cultivated in India. Each sample comprises 25 seeds.

Main protected area occurrence: Budongo (CFR), Bugoma (CFR), Itwara (CFR), Kagombe (CFR), Kalinzu (CFR), Kasyoha - Kitomi (CFR), Kibale (NP), Kisangi (CFR), Mabira (CFR), Murchison Falls (NP), Queen Elizabeth National Park (NP/BR), Rwensama (CFR), Semuliki (NP), South Maramagambo (CFR), Tero (CFR), Zoka (CFR). Key: Central Forest Reserve (CFR), National Park (NP), UNESCO-MAB Biosphere Reserve (BR).

Ugandan names: From (Eggeling and Dale, 1952): Mwanyi (Luganda, Lutoro, Kuamba, Lunyoro). From (Kalema and Hamilton, 2020): Mwanyi (Kwamba, Rutooro), Mumwanyi (Luganda), Omwanyi (Runyoro).

Other names: Wild robusta Coffee (Eggeling and Dale, 1952; Kalema and Hamilton, 2020).

Uses (other than beverage): As a masticatory-stimulant (due to the presence of caffeine), and snack, either fresh or dried. A traditional usage is to take a small number (c. 10) of prepared fruits (unripe or semi-ripe and dried whole, sometimes boiled in water) and package them in dried banana leaves, for sale in local shops and at roadsides (personal observation and Thomas (1944). *Coffea canephora* is used in various traditional and ritualistic activities, as an emblem for brotherhood and deep friendship. Even though this activity is steadily declining, it is still used in traditional marriage ceremonies in the Buganda culture.

2. *Coffea eugenioides* S.Moore, J. Bot. 45: 43 (1907) Figures 4E, F, 5C.

Global distribution: west-central Tropical Africa (Burundi [inferred or observed; no herbarium specimen data known], Rwanda, eastern Democratic Republic of Congo); north-east Tropical Africa (southern South Sudan); east Tropical Africa (central & eastern Kenya, eastern Tanzania, Uganda).

Global IUCN conservation assessment: Least Concern (LC) (O'Sullivan et al., 2020).

Information for Uganda

Distribution: Throughout Uganda but mainly in eastern and western parts of the country. Uganda is a centre of diversity for *C. eugenioides* (Thomas, 1944) and is the only country on the eastern side of the Great Rift Valley that holds substantial wild populations of this species. Figure 1B.

Ecology: An exclusively forest-dwelling species, found in the understorey of humid, evergreen forest (rainforest), occurring with a wide range of dominant tree species. Often occurring in the same forests and in close proximity to *C. canephora* (Figures 1, 4A) and at low elevations with *C. liberica* var. *dewevrei* (Figure 1).

Elevation: 910–1828 m (observed and recorded); 700–1693 m (modelled).

Preliminary country-level IUCN conservation assessment: Least Concern (LC). The Extent of Occurrence (EOO) is 133,516 km^2 (LC); and Area of Occurrence (AOO) is 372 km^2 (Endangered (EN). Whilst the EOO for *C. eugenioides* in Uganda is substantial, the AOO calculation returns a rating of EN, indicating that the preservation of this species should not be taken for granted and that careful monitoring is required. Over most the eastern part of its distribution in Uganda (Figure 1B), *C. eugenioides* is now restricted to smaller and increasingly fragmented forests; if these populations were extirpated the EOO would be severely reduced (c. 39,117 km^2) and the EOO based rating increased to Near Threatened (NT). Populations in the larger protected areas appear healthy, with a high density of individuals, but encroachment, deforestation, and disturbance in some protected areas (e.g. Zoka Forest and Itwara Central Forest Reserves) are negatively affecting AOO, number of mature individuals, habitat quality and population health.

Main protected area occurrence: Budongo (CFR), Bugoma (CFR), Bwindi Impenetrable National Park (NP/WH), Itwara (CFR), Kagombe (CFR), Kalinzu (CFR), Kasyoha - Kitomi (CFR), Kibale (NP), Kitubulu (CFR), Luvunya (CFR), Mabira (CFR), Malabigambo (CFR), Mbale (CFR), Mpanga (CFR), Mukambwe (CFR), Namalala (CFR), North Maramagambo (CFR), Nsube (CFR), Queen Elizabeth National Park (BR/NP), Rwensama (CFR), Semuliki (NP), South Maramagambo (CFR), Zoka (CFR). Key: Central Forest Reserve (CFR), National Park (NP), UNESCO-MAB Biosphere Reserve (BR), World Heritage Site (WH).

Ugandan names: From (Katende et al., 1999): Emwanji (Ateso), Mwanji (Adhola, Luganda, Lugwe), Imwanji (Lugisu), Omwani (Rukiga, Runyankore, Runyoro, Rutooro), Nkiga [sic] (Rutooro). From (Kalema and Hamilton, 2020): Mumwanyi (Luganda), Mwanyi (Ganato), Nkinga (Rutooro), Omwanyi (Runyoro).

Other names: Mukono coffee (English; (Bullock, 1930) but not widely applied). The name 'Nandi coffee' is used in Kenya for *C. eugenioides*, but refers to a Kenyan place name and hence a Kenyan variant of this species.

Uses (other than beverage): From (Katende et al., 1999): the ('sweet and tasty') ripe fruits are eaten as a snack, mainly by children; eaten in moderate amounts (Batooro, Bamba, Banyankore, Baganda); the fruits may also be boiled, dried and stored for later use as dry snacks (Baganda); dried leaves are put on hot charcoal and the smoke inhaled to relieve headache; the materials for the above uses are collected from the wild and are not cultivated.

3. *Coffea liberica* W.Bull, Nursery Cat. (William Bull) 97: 4 (1874).

Global distribution: west Tropical Africa (Benin, southern part), Ghana, Guinea, Ivory Coast, Liberia, eastern Sierra Leone, Nigeria); west-central Tropical Africa (Cabinda, Cameroon, Central African Republic, Congo, Democratic Republic of Congo, Gabon); north-east Tropical Africa (southern South Sudan); east Tropical Africa (Uganda); south Tropical Africa (Angola). Naturalised in Tropical Africa and other tropical areas (not listed here).

Global IUCN conservation assessment: Least Concern (LC). (Chadburn and Davis, 2017b).

3a. *Coffea liberica* var. ***liberica*** (not indigenous in Uganda).

Global distribution: west Tropical Africa (Benin, Ghana, Guinea, Ivory Coast, Liberia, Nigeria, eastern Sierra Leone); west-central Tropical Africa (Cabinda, Cameroon, Central African Republic, Congo, Democratic Republic of Congo, Gabon); south Tropical Africa (Angola). Naturalised in Tropical Africa and perhaps other tropical areas (not listed here); widely cultivated at small scale across the tropics.

Global IUCN conservation assessment: Not Evaluated.

3b. *Coffea liberica* var. ***dewevrei*** (De Wild. & T.Durand) Lebrun, Mém. Inst. Roy. Colon. Belge, Sect. Sci. Nat. (8vo) 11(3): 168 (1941).

Global distribution: west-central Tropical Africa (eastern Cameroon, Central African Republic, eastern Democratic Republic

of Congo); north-east Tropical Africa (southern South Sudan); east Tropical Africa (western Uganda).

Global IUCN conservation assessment: Not Evaluated.

Represented in Uganda by the endemic taxon *C. liberica* var. *dewevrei* forma *bwambensis* (see note below).

3b(i). *Coffea liberica* var. *dewevrei* forma *bwambensis* Bridson, Kew Bull. 37: 314 (1982) Figures 5A, 6.

Information for Uganda

Distribution: east Tropical Africa (western Uganda). Endemic to Uganda. Restricted to western Uganda, adjacent to the border with the Democratic Republic of Congo (in Semuliki Forest) and in north-eastern Uganda, in Zoka and at Kilak (Killak) Central Forest Reserves). Comprehensive fieldwork in Itwara Forest (during the years 2020 and 2021) shows that this species does not occur at this location, contrary to previous reports (Thomas, 1940b; Thomas, 1944; Kalema and Beentje, 2012; Kalema and Hamilton, 2020). Figure 1C.

Ecology: In medium elevation, humid, evergreen forest (Zoka Forest) and lowland semi-deciduous humid forest (Semuliki Forest), with a diverse range of dominant tree species and various forest communities. In Semuliki Forest, *C. liberica* var. *dewevrei* occurs mainly (c. 90%) in swamp forest, and even in places that support truly riverine species such as *Pandanus chiliocarpus* (screw pine), although it is not exclusively confined to these habitats in Semuliki Forest, and it grows in drier (soil) areas of this forest reserve that are not associated with water. In Zoka Forest, this species is predominantly found in areas that are not associated with rivers and waterlogged areas.

Elevation: 680–1200 m (observed and recorded); 686–1118 m (modelled).

Preliminary country-level IUCN conservation assessment: Endangered (EN). The Extent of Occurrence (EOO) is 7,716 km^2 (Vulnerable (VU); and Area of Occurrence (AOO) is 64 km^2 (Endangered (EN)). Whilst the EOO for *C. liberica* var. *dewevrei* in Uganda is substantial, a large part of the EOO is without populations of this species, and a substantial area includes Lake Albert (Figure 1C). Populations in Semuliki Forest (National Park) have a high density of individuals, and occur in a large proportion of the areas of the forest that have been surveyed (i.e. the eastern part). Encroachment and deforestation in Zoka Forest are affecting AOO, number of mature individuals, habitat quality and population health. Conservation management improvements are urgently required in Zoka Forest to ensure the Ugandan northernmost populations of this species are protected. Careful monitoring and management of this species *in situ* is urgently required for Uganda.

Main protected area occurrence: Semuliki (NP), Zoka (CFR). Key: Central Forest Reserve (FR), National Park (NP).

Ugandan names: From (Kalema and Hamilton, 2020): Mumwanyi (Luganda). Kisansa coffee (recorded on farms in the Luwero District).

Other names: Coffea excelsa A.Chev. (Botanical Latin; numerous authors). Excelsa coffee (English, numerous authors, widely used); Shari coffee [English; (Eggeling and Dale, 1952)], not widely used.

Uses (other than beverage): Not known.

Notes: The current consensus of taxonomic and systematic study (Lebrun, 1941; Bridson, 1988; N'Diaye et al., 2005; Davis et al., 2006; Baltazar and Buot, 2019; Panaligan et al., 2020; Panaligan et al., 2021) is that *C. liberica* should be divided into two botanical varieties: *C. liberica* var. *liberica* and *C. liberica* var. *dewevrei*. Whilst this view is

FIGURE 6
Coffea liberica var. *dewevrei* (excelsa coffee), cultivated in central Uganda. **(A)** Habit or farmed plant (tree), c. 5 m high, growing with banana; **(B)** Single flower (5-merous), old flowers and flower buds; **(C)** Fruiting branches, with maturing fruits; **(D)** Fruiting branch, with tight axillary clusters; **(E)** Single fruit (cherry), c. 17 × 15 mm; **(F)** sack of fruit (cherry) from a single tree (c. 70 kg).

generally accepted, it is also argued that the taxonomy of *C. liberica* does not fully account for the extreme morphological (Bridson, 1988; Stoffelen, 1998) and potential molecular variation (Charr et al., 2020) within the species, and thus requires further careful critical study. One of the main problems is that *C. liberica* has been introduced and become naturalised throughout tropical Africa, and so sampling for systematic studies may be biased. There also seems to be confusion around the plants identified as 'excelsa', a name that should only be used to refer to var. *dewevrei* but has been used, incorrectly, for variants of var. *liberica* (Davis et al., 2022). The botanical forma (f.) *bwambensis* (i.e. *C. liberica* var. *dewevrei* f. *bwambensis*) has been assigned to represent all indigenous Liberica coffee in Uganda, but the morphological circumscription may possibly also include populations in South Sudan (Bridson, 1988) and perhaps adjoining areas in the Democratic Republic of Congo. Given the uncertainty, in this contribution we refer to the Ugandan populations of Liberica coffee as *C. liberica* var. *dewevrei*.

In 1941, *C. liberica* var. *dewevrei* was found near Kilak (Killak), to the east of Zoka Forest, in riverine forest; fieldwork is required to ascertain whether this species still exists at this location.

4. *Coffea neoleroyi* A.P.Davis, Phytotaxa 10: 43 (2010).

Global distribution: north-east Tropical Africa (south-western Ethiopia, and south-western South Sudan); east Tropical Africa (north-eastern Uganda).

Global IUCN conservation assessment: Endangered (EN) (O'Sullivan et al., 2017).

Information for Uganda

Distribution: Restricted to Mt. Zulia in north-eastern Uganda Figure 1D.

Ecology: On riverbanks, and in seasonally dry *Combretum-Terminalia* savanna woodland, often amongst boulders.

Elevation: c. 1200 m.

Preliminary country-level IUCN conservation assessment: Critically Endangered (CR) or Data Deficient (DD). The Extent of Occurrence (EOO) cannot be calculated owing to a single data point; Area of Occurrence (AOO) is 4 km^2 (CR), based on a single grid cell with a cell width of 2 km. Realistically, at the country-level, this would be better placed in the DD category, given that there is only a single collection, and the area where this species grows is isolated, and has not been the subject of detailed botanical survey; dedicated survey work for this species in north-eastern Uganda is urgently required.

Protected area occurrence: Zulia (FR). Key: Forest Reserve (FR).

Ugandan names: Not known.

Other names: Not known.

Uses (other than beverage): Not known.

Notes: A rare and untypical *Coffea* species, formerly included in *Psilanthus* (Davis et al., 2011), which is characterised by its deciduous habit, and long-tubed flowers (corolla) with very short (included) styles (Davis et al., 2005). All other Ugandan *Coffea* species and most (but by no means all) other *Coffea* species have evergreen leaves, short-tubed flowers, and a long (excluded) style. Like all *Coffea* species, *C. neoleroyi* produces a fruit that contains two seeds, and each seed possesses the typical coffee-bean morphology. The seeds (coffee beans) of *C. neoleroyi* are much smaller than *C. canephora* and *C. liberica*, and smaller than *C. eugenioides* (Figure 5).

Prior and existing uses of wild coffee resources within the Ugandan coffee sector

Coffea canephora (robusta coffee)

The commercial use of *C. canephora* in Uganda dates to at least the mid-1800s, when various observers recorded farming for local use (as a product for chewing and consumption, rather than as a beverage), and national cross-border trade (Thomas, 1935; Thomas, 1940a; Wrigley, 1988; Kiwuka et al., 2021). Traditional and ritualistic uses of this species in Uganda are long-established, although the historical time-line is unclear (Thomas, 1935; Thomas, 1940a). Arabica coffee (*C. arabica*) was believed to have been introduced into cultivation in Uganda in 1900 (Thomas, 1940c), at which time coffee cultivation in general (including *C. canephora*) had started to be promoted as a key agricultural export, gathering increased momentum from the 1910s onwards (Thomas, 1940a).

A recent survey of the genetic diversity of *C. canephora* (Kiwuka et al., 2021) using microsatellite (SSR) markers on a comprehensive sampling of wild and cultivated accessions from Uganda, as well as indigenous populations of this species across Africa, has provided considerable enlightenment regarding the origin of farmed *C. canephora* in Uganda. The analyses of Kiwuka et al. (2021) infer that indigenous populations from the forests of Malabigambo, Mabira, and Kalangala (Ssese Islands), i.e. the southern-central (SC) genetic cluster of Kiwuka et al. (2021), as well as introduced germplasm from other parts of Africa, represent the origin of the farmed robusta used today in Uganda. There is thus agreement with what is known about the early years of coffee cultivation in Uganda, particularly with reference to the development of robusta coffee on the Ssese Islands and surrounding areas on the mainland (Thomas, 1940a; Thomas, 1944; Wrigley, 1988). The study confirms that plants of Congolese origin (i.e. Democratic Republic of Congo) were introduced into Uganda as part of the effort to upscale coffee production in Uganda, even though there were plentiful indigenous *C. canephora* resources present in the forests of Uganda: the transfer of western African *C. canephora* to Uganda is recorded in the literature (Cheney, 1925; Thomas, 1935). The movement of *C. canephora* from other parts of Africa, may have been direct, or indirect, for example *via* Java (Thomas, 1935; Wrigley, 1988). The probable reasons for the introduction of non-Ugandan germplasm were that *C. robusta* (now a synonym of *C. canephora*) was then considered a separate species to *C. canephora*, and that the variation across the two species included specific positive traits and qualities (Cheney, 1925; Thomas, 1935) and would have thus been considered as worthy introductions. It has been suggested (Thomas, 1935) that the Nganda type of *C. canephora* (which has a spreading habit) may have originated from a forest in Uganda, whereas the Erecta type (upright habit) was introduced from the Congo Basin. In their analyses Kiwuka et al. (2021) these morphological types are, however, intermixed with individuals of the SC group, of both non-Uganda and Ugandan origin, indicating that their phenotypic differences are not clearly distinguishable (at least, on the basis of the 19 microsatellite (SSR) markers used), or may have no genetic foundation. There may have been a number of man-made introductions from a range of countries/populations within Africa, although in some cases the genetic similarity may be signalling

natural relationships (i.e. without human intervention) with contiguous and perhaps non-contiguous regions across tropical Africa (Kiwuka et al., 2021). In summary, and based on the evidence at hand, it is inferred that most of the farmed germplasm of Uganda originated from the southern-central forests [the SC cluster of Kiwuka et al. (2021)], including Malabigambo, Kalangala, Mabira, but with admixture of material from other countries, and perhaps intermixing (via spontaneous crosses) between indigenous and introduced genotypes; although the relative contributions of these three factors would be difficult to assess (Kiwuka et al., 2021). The results of Kiwuka et al. (2021) are consistent with the claim that in the latter part of 1800s, plantation owners and smallholder farmers in the Lake Victoria Basin region began cultivating *C. canephora* using directly sourced wild coffee (Thomas, 1935; Thomas, 1940a; Thomas, 1944), although there would have been a considerable amount of selection for the best performing variants at that time (Thomas, 1935). Kiwuka et al. (2021) also revealed that all of the six elite clones (KW13, KW14, KW15, KW16, KW18 and KW19) possessing coffee wilt disease (CWD) resistance and high yield characteristics (Mulindwa et al., 2022), and which have provided the mainstay of modern robusta production in Uganda, are genetically similar to wild SC populations (Kiwuka et al., 2021). It should also be said that a large proportion of globally cultivated robusta originated from Uganda, including those grown by major coffee producers, such as Vietnam and Mexico (Garavito et al., 2016). Kiwuka et al. (2021) also show that the genetic diversity found in Uganda's north-western forests (Zoka, Budongo, Itwara and Kibale) is distinct from the germplasm currently employed in Uganda's coffee farming sector. This is noteworthy because these populations occur in comparatively warmer and drier climatic zones, and may have climate resiliency attributes (Kiwuka et al., 2021), i.e. higher thresholds to abiotic stressors, such as higher temperatures and lower soil moisture.

Coffea eugenioides

Early trials of *C. eugenioides* as a beverage species, in East Africa (including Uganda), were unfavourable due to the small size of its seeds (coffee beans) (Thomas, 1944), susceptibility to coffee leaf rust (Bullock, 1930), poor quality, and very small yields (Thomas, 1940b; Thomas, 1944). Early sensory assessments, were not unfavourable, however, for example: "The liquor was described as pure and entirely free from undesirable flavours, although the strength was not good, probably owing to the presence of immature trees" (Thomas, 1940b); and …"the quality of the bean is mild and agreeable"… (Thomas, 1944). Outside Uganda, and more recently, the flavour of *C. eugenioides* has received praise, for example: "The exception is *C. eugenioides*, which has a very fine aroma, tasting fruity and clean." (Fazouli et al., 2000). Trial plantings of *C. eugenioides* were made in Uganda during the 1920s and 1930s (as evidenced by samples housed at the Economic Botany Collection, RBG Kew), but no further development of the species seems to have been undertaken there. In south-eastern Kenya, small scale production of *C. eugenioides* has been underway for several decades, as Nandi coffee, although the current status of this crop is unclear. In Colombia, *C. eugenioides* is presently grown on a small scale (i.e. on a single estate that produces specialty (high quality) coffee), which sells at a substantial premium, and has been used in national and international coffee making competitions (i.e. the World Barista Championships) on account of its unique, complex flavour and intense natural sweetness. *Coffea eugenioides* is reported to be difficult coffee to grow, and low yielding (e.g. 150 grams per tree of un-milled coffee; https://cafeinmaculada.com/en/blogs/varieties/variedades). Based on the renewed interest, and high market price, preliminary trials of *C. eugenioides* are now underway in Uganda.

Coffea eugenioides has been used as a breeding partner, for imparting flavour qualities and other attributes *via* crosses with other species. Spontaneous diploid (2n=22) crosses between *C. eugenioides* and *C. liberica* have been identified (Maurin et al., 2007) and artificial tetraploid crosses ('Ligenioides') have been made (Reddy et al., 1985; Ganesh et al., 2002), and evaluated (see below). Tetraploid (2n=44) versions of this interspecies hybrid can be readily backcrossed with *C. arabica* (2n=44) and *C. canephora* (using the Timor Hybrid) to produce hybrids with the potential for commercial use (Ganesh et al., 2002). Diploid, artificial hybrids between *C. eugenioides* and *C. canephora* have also been made, which after chromone doubling (2n=44) crossing with *C. arabica* and then backcrossing, have produced a line of high yielding hybrids with acceptable beverage quality and high productivity (Nagai et al., 2008).

Coffea liberica var. *dewevrei* (excelsa coffee)

Coffea excelsa A.Chev. (Chevalier, 1903) and *C. dewevrei* De Wild. & Th. Dur. (Durand and De Wildeman, 1899) are synonyms of *C. liberica* var. *dewevrei* (Davis et al., 2006). *Coffea excelsa* features predominately in early references of this wild plant in Uganda (Thomas, 1940b; Tothill, 1940; Thomas, 1944), and elsewhere (Cheney, 1925; Wellman, 1961; Wrigley, 1988). Indeed, the common name 'excelsa' is frequently applied to this plant, just as Arabica and robusta are applied to *C. arabica* and *C. canephora*. It should be noted, however, that the name 'excelsa' whether used as a common name, or as a species epithet, is often incorrectly applied to small to medium sized seed (coffee bean) variants of *C. liberica* var. *liberica* (Davis et al., 2022), particularly those cultivated in Asia but also parts of Africa. The common name 'excelsa' should be restricted to those plants conforming to *C. liberica* var. *dewevrei* (Bridson, 1988; Davis et al., 2022) originating from the eastern Cameroon, Central African Republic, eastern Democratic Republic of Congo and western Uganda.

Field trials of excelsa coffee in in Uganda, during 1915 and 1916, indicated poor yields (Cheney, 1925; Thomas, 1940b). However, the material used was imported from Java in 1914, and is certainly not *C. liberica* var. *dewevrei* (excelsa) but rather *C. liberica* var. *liberica*, as indicated by the large size of its seeds (Davis et al., 2022). In other countries, cultivated material of *C. liberica* var. *dewevrei* (as *C. excelsa*) was more thoroughly assessed and received favourable reviews. Many identified considerable potential for excelsa as a coffee crop species (Freeman and Chandler, 1907; Cramer, 1913; Cheney, 1925; Chevalier, 1929; Cramer, 1957). For example [translated from French]: "Many farmers consider it to have a great future, as it is very resistant to diseases and insects, and it gives high yields of good quality coffee' (Chevalier, 1929). Some of the information pertaining

to excelsa coffee is likely to be misplaced owing to the confusion between var. *liberica* (Liberica/Liberian coffee) and var. *dewevrei* (excelsa) (Davis et al., 2022), as indicated above.

Over recent decades, the timing of which is not clear but which may date back to at least the 1980s, there has been a dramatic increase in the number of farmers in Uganda (perhaps more than 200) growing *C. liberica* var. *dewevrei*, either with *C. canephora*, or as the dominant coffee crop (Davis et al., 2022). The shift to *C. liberica* var. *dewevrei* has been farmer-led, and has occurred independently of extrernal influences, other than minor interest in purchasing for export as a differentiated coffee (Davis et al., 2022). According to the farmers in lowland Uganda growing *C. liberica* var. *dewevrei*, this plant has been on their farms, in low numbers, for many decades, and was originally gathered from the forest by previous generations (Davis et al., 2022), although this requires confirmation. Preference for farming *C. liberica* var. *dewevrei* over *C. canephora* appears to be the result of production issues with *C. canephora* (robusta), and particularly the increasing occurence and severity of disease (especially coffee wilt disease), pests (particularly stem/twig borers) and droughts. A similar upsacling of *C. liberica* var. *dewevrei* is also occurring in South Sudan (Davis et al., 2022). Farmers in Uganda consistently report high yields for *C. liberica* var. *dewevrei*, which based on yield-per-plant of fresh fruit (e.g. Figure 6F) and an outturn (conversion) ratio (kg of fresh fruit: kg clean coffee) of 7:1, ranges between 877 kg/ha (204 trees/ha) to 3,440 kg/ha (400 trees/ha), for rain-fed, low input (e.g. negligible fertiliser use) farming systems (Davis et al., 2022). As yet, there are no reports of coffee wilt disease (Davis et al., 2022), which is a widespread and devastating disease of *C. canephora* in Uganda. Improved monitoring and further research is required to asess the level of resistance of coffee wilt disease in *C. liberica* var. *dewevrei*, as it was first reported on this species in the Central African Republic in 1927, and later caused widespread damage to Liberica and robusta coffee across large areas of tropical Africa (Gaitán et al., 2015).

Unlike *C. arabica* and *C. canephora*, *C. liberica* var. *dewevrei* grows into a substantial, medium sized tree (Figure 6A), of around 10 m or more. In Uganda, the fruit development period of *C. liberica* var. *dewevrei* is longer than the aforementioned crop species, and the main harvest periods do not overlap: the harvest period for *C. liberica* var. *dewevrei* is January to April. The fruits of *C. liberica* var. *dewevrei* are held in tight axillary clusters (Figures 6C, D), like *C. canephora* and many cultivars of *C. arabica*, but unlike many variants of *C. liberica* var. *liberica*. The fruits and seeds of *C. liberica* var. *dewevrei* are approximately the same size and dimensions (Figures 6C–E) as *C. arabica* (Figure 5E). This presents a distinct advantage over the large-fruited, thick pulped variants of *C. liberica* var. *liberica*, as processing can be carried using standard procedures and the outturn (conversion ratio of fresh fruit to clean coffee) is much more satisfactory (Davis et al., 2022).

Coffea liberica var. *dewevrei* (excelsa coffee) produced in Uganda and South Sudan yields a coffee that is smooth and easy-drinking, with low to medium acidity, low bitterness, possessing a range of positive flavour notes, and a caffeine content similar to *C. arabica* (Davis et al., 2022).

In Uganda, *C. liberica* var. *dewevrei* is mostly exported as, or mixed with, *C. canephora* (robusta). This is partly due to confusion over the identity of excelsa (sometimes considered a large, thick-leaved type of robusta) and because of convenience. A separate market for excelsa coffee (*C. liberica* var. *dewevrei*) does not exist in Uganda, although there have been limited exports of this coffee to Italy in recent years, as Kisansa coffee (https://www.fondazioneslowfood.com/en/ark-of-taste-slow-food/kisansa-coffee/) and in 2022 to the UK (Clifton Coffee, personal communication).

Coffea liberica var. *dewevrei* may also have considerable utility in the Ugandan coffee sector as grafting stock. *Coffea liberica* sensu lato is used to improve resistance to root nematodes, and increase yield and survivability of grafted *C. arabica*, particularly in Hawaii (Myers et al., 2020). In Uganda, *C. liberica* var. *dewevrei* could be used for grafting of CWD resistant *C. canephora* clones. Currently, CWD resistant *C. canephora* is reproduced by cuttings, which although successful means that a tap root is not formed. Plants (scions) of *C. canephora* would likely benefit from the stout tap root and extensive, robust root system of *C. liberica* var. *dewevrei*, particularly under low soil moisture conditions. Further research is warranted, including the identification of the most suitable grafting stock, as undertaken for the grafting of *C. arabica* onto *C. liberica* var. *liberica* stock in Hawaii (Myers et al., 2020).

Coffea liberica var. *liberica* has been used in coffee crop development, *via* hybridization with other beverage species. It has contributed coffee leaf rust resistance to the widely grown Indian cultivar *C. arabica* 'S.795', *via* the progenitor cultivars *C. arabica* 'S.288' and 'S.26' (Narasimhaswamy, 1960; Surya Prakash et al., 2002). In Indonesia, crosses between *C. liberica* var. *liberica* and *C. arabica* have provided a number of tetraploid (and octaploid) hybrids, most notably the 'Kalimas' and 'Kawisari' hybrids, which also have a high degree of resistance to coffee leaf rust, and in some cases high yields and a fair market price (Cramer, 1957). Diploid hybrids between *C. canephora* and *C. liberica* (possibly var. *dewevrei*) have also been documented (Cramer, 1957; Chinnappa, 1970). As mentioned above (for *C. eugenioides*), *C. liberica* hybridizes with *C. eugenioides*, to form diploid and tetraploid hybrids. In India, these hybrids showed good yield potential and coffee leaf rust resistance, although the need for further development was identified (Reddy et al., 1985; Ganesh et al., 2002). In Madagascar, tetraploid *C. liberica* and *C. eugenioides* have been crossed, and then backcrossed with *C. arabica*, to produce high-yielding hybrids with acceptable sensory characteristics, as part of the 'Ratelo Hybrid' programme (Jean Jacques Rakotomalala, personal communication). Thus, in Uganda, indigenous *C. liberica* var. *dewevrei* could offer potential as a breeding partner, for imparting required traits to new coffee crop plants, *via* interspecies crosses.

Coffea neoleroyi

Several factors preclude *C. neoleroyi* from being a coffee crop species, including extremely low yields (due to the production of low numbers of fruits per tree, and small seed (coffee bean) size, diminutive stature and spindly growth form. It is possible that a coffee-like beverage could be made from the seeds of *C. neoleroyi*, but this remains untested. Given the differences in floral morphology between *C. neoleroyi* and coffee beverage crop species (Davis et al., 2005) breeding with other coffee species would be difficult (Couturon et al., 1998). Due to the rarity and geographical isolation of *C. neoleroyi* in Uganda, and indeed in the other known locations for this species (South Sudan and Ethiopia), this species remains poorly known.

Climate profiling

Our basic modelled climate data analysis for the four wild coffee species of Uganda is summarised in Table 1. A scatter plot of annual mean temperature (Bio1) vs. total mean annual precipitation (Bio12) for Ugandan coffee species is given in Figure 2; *C. neoleroyi* is included for illustrative purposes only (as there is only a single data point). Density plots for Bio1, Bio12, precipitation seasonality (Bio15), and elevation are given in Figure 3 (except for *C. neoleroyi*). The modelled mean annual temperatures, annual precipitation and precipitation seasonality for the Ugandan populations are: *C. canephora* (21.8°C/1389 mm/44), *C. eugenioides* (21.3°C/1370 mm/42), and *C. liberica* var. *dewevrei* (24.4°C/1560 mm/51). These data (Table 1; Figures 2, 3) show that *C. canephora*, *C. eugenioides* and *C. liberica* var. *dewevrei* overlap for Bio1, Bio12 and Bio15, which is not surprising given that these species overlap in their distributions (Figure 1), especially *C. canephora* and *C. eugenioides*. A greater density of warmer mean annual temperature (Bio1) for *C. liberica* var. *dewevrei* (compared to *C. canephora* and *C. eugenioides*) is evident because this species is confined to lower elevation forests in western and northern Uganda (Figures 1, 2). *Coffea liberica* var. *dewevrei* has a higher density of occurrences in locations with a higher total mean annual precipitation (Bio12), and precipitation seasonality (Bio15), as shown in Figure 3. The higher density of lower precipitation (Bio1) and lower precipitation seasonality (Bio15) for *C. canephora* and *C. eugenioides* is due to the higher number of datapoints in drier locations, compared to wetter locations, biased by the higher number of datapoints for these species overall. For *C. liberica* var. *dewevrei*, the density distribution for Bio12 (i.e. number of wetter locations) is bimodal (Figure 3A) due to the disparity in rainfall for the north-western locations (Zoka Forest and Kilak (Killak); 938–1580 mm per year), vs. those in central-western (Semuliki National Park; c. 2200 mm), and because the number of ground-point data records are the same in each area (five). The values for precipitation seasonality (Bio 15) are higher in *C. liberica* var. *dewevrei* (51), compared to *C. canephora* and *C. eugenioides* (44 and 42), but given the above-stated considerations (including the low number of samples) any firm interpretations are inadvisable. The global values (i.e. across the entire indigenous distribution) for *C. canephora* are higher [56; (Davis et al., 2021b)] than for Uganda alone (44). Given the higher precision expected with a national (Uganda only) rather than global (across the African continent) the lower figure is likely to be more meaningful.

Discussion

Survey of wild coffee species and conservation priorities

There are four indigenous (wild) coffee species in Uganda: *C. canephora*, *C. eugenioides*, *C. liberica* var. *dewevrei* and *C. neoleroyi*. *Coffea canephora* and *C. eugenioides* are widespread in western Uganda (Figures 1A, B) where suitable forest habitat exists, although in central Uganda many populations occur in small and often degraded forest parcels, which require improved safeguarding. Uganda represents important centres of diversity for *C. canephora*, *C. eugenioides* and *C. liberica* var. *dewevrei*, and all three are priority species for coffee crop plant development (i.e. Coffee Crop Wild Relative (CWR) Group 1; (Davis et al., 2019). *Coffea liberica* var. *dewevrei* is restricted to three populations in western Uganda (Figure 1C). In Semuliki Forest (National Park) the population appears to be quite extensive, healthy, and with a reasonably high density individuals. The forested area covers most of the Semuliki protected area boundary (219 km²). Conversely, in Zoka Forest the population is under threat from encroachment and is chronically suffering from reduced forest cover and poor forest health. The third population, at Kilak (Killak), has not been surveyed for *C. liberica* var. *dewevrei* since it was last recorded there (Thomas, 1944); dedicated fieldwork in this area is required. All of the records used by Thomas (1940b); Thomas (1942); Thomas (1944) for his surveys of indigenous coffee species were based on, or vouchered, using herbarium specimens, except the records used for *C. liberica* var. *dewevrei* at Itwara, which were based on observation only. It is likely that he, or his informants, mistakenly identified large-leaved variants of *C. canephora* (which the authors have seen in Itwara) as *C. liberica* var. *dewevrei*. It could be argued that the *C. liberica* var. *dewevrei* once occurred in this forest and has since been extirpated due to partial deforestation, but this seems unlikely. There are plentiful herbarium specimens for *C. canephora* and *C. eugenioides* collected from Itwara, and it seems likely that at least one specimen of *C. liberica* var. *dewevrei* would have been collected, especially given the relatively easy access to this forest. *Coffea neoleroyi* is only known from a single collection (Figure 1D) in a remote area of north-eastern Uganda. Further dedicated field survey for this species is required, to fully understand the number and density of populations in Uganda, and across its natural range (i.e. Uganda, Ethiopia and South Sudan).

Our preliminary country-level IUCN Red List conservation assessments (IUCN Standards and Petitions Subcommittee, 2022) for the four indigenous coffee species of Uganda are: *C. canephora* (Least Concern), *C. eugenioides* (Least Concern), *C. liberica* var. *dewevrei* (Endangered), and *C. neoleroyi* (Critically Endangered, or Data Deficient). Under a Least Concern rating, individual populations may still be at risk of extirpation, as is the case for *C. canephora* and *C. eugenioides*. In terms of genetic resources, and their value to the Ugandan coffee sector, this review shows that potentially useful attributes (diseases resistance, climate resiliency, etc.) are distributed across populations (as well as the species as a whole) and thus require conservation. The rating of Endangered for *C. liberica* var. *dewevrei* is of considerable concern, particularly given the level of forest clearance and land use change at Zoka Forest. Further data is required before a confident extinction assessment can be made for *C. neoleroyi*, but this is also undoubtedly also a species of concern.

Climate profiling

In Uganda, indigenous *C. canephora*, *C. eugenioides* and *C. liberica* var. *dewevrei* occur (grow and reproduce) over the same range of basic climate variables (Bio1, mean annual temperature;

Bio12, mean annual precipitation; and Bio15, precipitation seasonality) as summarised in Table 1 and Figures 2, 3. *Coffea liberica* var. *dewevrei* occurs in much warmer locations (mean annual temperature 24.4°C) than *C. canephora*, *C. eugenioides* (21.8°C and 21.3°C, respectively) on account of it being restricted to low elevation forests. The elevation restriction may not be due to inability, or lack of opportunity, for this species to exist at higher elevations, although there could be intrinsic factors in play. There is also the possibility that the distribution and elevation range of *C. liberica* var. *dewevrei* in Uganda may have been more extensive historically, prior to forest clearance by humankind. The bimodality in mean precipitation (Bio12) and higher precipitation seasonality (Bio 15) for *C. liberica* var. *dewevrei* (compared to *C. canephora* and *C. eugenioides*; Figure 3) may infer further climatic differences (other than a mean temperature difference) but based on the data at hand no firm assumptions can be made. Field observations made by us show that *C. liberica* var. *dewevrei* is often associated with high water tables. It occurs in swamp forest (at Semuliki National Park) and can be close to rivers (Zoka Forest), and outside Uganda it has been often recorded in gallery forest in native habitats (Chevalier, 1929). However, it is by no means exclusive to these wetter habitats, as observed by us in Zoka Forest. Published observations (Thomas, 1940b; Thomas, 1944) and farmer feedback (from farm observation and farmer feedback during dry spells in 2021 and 2022) indicate that *C. liberica* var. *dewevrei* is more drought tolerant than *C. canephora*. Drought tolerance assumptions and observations for *C. liberica* var. *dewevrei* (as *C. excelsa*) have been made by other workers (Anon, 1890; Cheney, 1925; Cramer, 1957). The modelling approach used here neither supports nor refutes drought tolerance (e.g. higher precipitation seasonality), particularly as the Bio15 values for *C. canephora* (precipitation seasonality (PS) value = 44) *C. liberica* var. *dewevrei* (PS = 51) are not that far apart (Table 1; Figure 3E). However, the values for indigenous *C. arabica* (Ethiopia and South Sudan (PS = 58; (Davis et al., 2021b); and indigenous Ugandan *C. canephora* (PS = 44) infer that there are precipitation seasonality differences between these two species. This is supported by observation of wild and farmed populations of *C. canephora*, which generally occur in locations with lower precipitation seasonality than *C. arabica* (A. Davis personal observation), although these relationships are complex (and probably fine-scaled) and require careful evaluation. *Coffea racemosa* and *C. zanguebariae*, two species occurring in areas of extremely seasonal rainfall, can have Bio15 values of 90 (Davis et al., 2021a), indicating the scale of difference in precipitation seasonality between *Coffea* species. Multi-location variety trials (MLVTs) are required to substantiate the analyses and observations presented here, and better understand the climatic tolerances of indigenous Ugandan coffee species, and cultivated Arabica coffee. Experimentation of this nature would be critical for understanding the value of Uganda's wild species resources for crop development, across different agroecological conditions under a changing climate.

Prior, existing and future uses of wild coffee resources

Molecular analyses (Kiwuka et al., 2021) have substantiated the assumption that indigenous populations of *C. canephora* (robusta coffee) have provided [after selection (Thomas, 1944)] Uganda with the bulk of the germplasm on which their robusta farming sector is based (Thomas, 1940a; Thomas, 1944), as elaborated above (see Results). Indigenous coffee natural capital has also provided the resources for sustaining robusta production in Uganda and other countries (Garavito et al., 2016). In particular, wild germplasm (Kiwuka et al., 2021) has played a key role in developing cultivars (several clones) to combat the devastating effects of coffee wilt disease (Rutherford, 2006; Musoli et al., 2009; Musoli et al., 2013; Mulindwa et al., 2022), which ravaged robusta production in Uganda in the 1990s and remains an ongoing issue. Despite the considerable contribution already made by using indigenous *C. canephora* resources, much remains untapped (Ngugi and Aluka, 2019; Kiwuka et al., 2021) and may prove to be of value for addressing sustainability issues. For example: germplasm from wild populations of *C. canephora* may offer climate resiliency potential, particularly those from warmer and drier forests, such as Zoka Forest (Kiwuka et al., 2021); it has been shown that wild collections of wild robusta from Kalangala and Itwara forests have a high level of resistance to coffee wilt disease (Phiri and Baker, 2009); and selections from Itwara have shown promise under plantation conditions (Thomas, 1944).

The recent farming uptake of *C. liberica* var. *dewevrei* (excelsa coffee) represents an interesting development, and may offer considerable potential, as a third coffee crop species for Uganda (after *C. arabica* and *C. canephora*). Excelsa coffee fulfils farmers requirements, as it is easy to grow, appears to possess resistance to many of the major pests and diseases of coffee, is high yielding, and has an acceptable conversion (outturn) ratio from fresh fruit to clean coffee, particularly compared to the large fruited, thickly pulped types of *C. liberica* var. *liberica* (Davis et al., 2022). *Coffea liberica* var. *dewevrei* may be more tolerant of higher temperatures, compared to *C. canephora* in Uganda (mean annual temperature 24.4°C vs. 21.8°C; Table 1), but this requires careful assessment *via* field trials. Farmers in Uganda have reported better performance (personal communication) of *C. liberica* var. *dewevrei* over *C. canephora* during drought conditions (in 2021 and 2022). *Coffea liberica* var. *dewevrei* is certainly more heat tolerant than wild and cultivated *C. arabica*, which has a mean annual temperature range of 18–22 °C (Alègre, 1959; DaMatta and Ramalho, 2006; Davis et al., 2021b). It is clear that *C. arabica* cannot be grown successfully alongside *C. liberica* var. *dewevrei* in most lowland conditions in Uganda, even though it may persist over the short-term (Thomas, 1940c; Haarer, 1962). Whilst our climate analyses are not in conflict with these observations they do not support them with any degree of confidence. The thick, fleshy leaves, stout drunk and extensive root system of *C. liberica* var. *dewevrei* are features that are likely to constitute drought (and heat) tolerance advantages over *C. canephora*. Under wet soil conditions in Uganda, the root-system of *C. canephora* has been reported to be superficial and shallow (Thomas, 1944), which is likely to make this species susceptible during drought periods. It has also been noted that *C. liberica* var. *dewevrei* is tolerant of a wide range of soils (Thomas, 1944). *Coffea liberica* var. *dewevrei* (excelsa) is capable of producing economically viable, good quality coffee (see above) but further assessment is required to see how it will perform across the value chain. This species may also offer useful prospects as a rootstock for *C. canephora* and perhaps other coffee crop species, and for coffee crop development *via* breeding (see Results for the potential of both uses).

Due to its highly desirable flavour qualities, *C. eugenioides* may offer potential as a niche crop for the high-value sector of the coffee market, as it has in Central America. The development of *C. eugenioides* in Uganda would require investment and a proof-of-concept period, to test for commercial viability. The small seed (coffee bean) size and low yields represent key constraints, unless better performing variants can be found within wild populations, although this seems unlikely based on the field surveys we have carried out so far. *Coffea eugenioides* is likely to offer better potential as a breeding partner, for imparting flavour qualities and other attributes *via* interspecies crosses.

The wild diversity of *C. eugenioides* and *C. canephora* in Uganda could be of paramount interest, since they are the progenitors of *C. arabica* and could be used to produce Arabica analogues. Moreover, it has been shown that the *C. canephora*-derived sub-genome of *C. arabica* is closely related to the *C. canephora* accessions from northern Uganda, and in particular Zoka Forest (Merot-L'anthoene et al., 2019).

Conclusion

In this contribution we enumerate four indigenous coffee species for Uganda (*C. canephora*, *C. eugenioides*, *C. liberica* (var. *dewevrei*) and *C. neoleroyi*) and provide new ecogeographical data summaries (and other information) for each of these species. Climate profiling, *via* simple modelling methods, shows overlap for basic climate requirements for three of these species (*C. neoleroyi* was excluded due to lack of data), although *C. liberica* var. *dewevrei* has a higher density of individual records in locations of higher temperature, and a higher precipitation seasonality. At the national level, a draft IUCN Red List assessment indicates that *C. liberica* var. *dewevrei* is Endangered, and that *C. neoleroyi* could be Critically Endangered. Many wild *Coffea* populations in Uganda are compromised due to land use change (e.g. deforestation and agricultural encroachment) and some populations may be threatened with extirpation. The considerable indigenous diversity reported for *C. canephora*, and assumed diversity for *C. eugenioides* and *C. liberica* var. *dewevrei*, based on number of populations and range, represents valuable natural capital for crop development (e.g. *via* breeding) and the sustainability of the Uganda coffee sector in general, particularly under changing climatic conditions. Wild populations of *C. canephora* have provided Uganda with the bulk of diversity for the establishment and sustainability of its thriving robusta coffee sector. *Coffea liberica* var. *dewevrei* (excelsa coffee) shows potential as stand-alone crop species, and as a source of grafting stock for *C. canephora* (robusta coffee) and other coffee species. The coffee natural capital of Uganda requires improved protection, in order to avoid the loss of genetic diversity and coffee crop development options.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials. Further inquiries can be directed to the corresponding author.

Author contributions

AD designed the research project, undertook the research and some of the analyses, and wrote the paper; CK designed the research project and undertook the research; AF designed the research, undertook the research and most of the analyses; JM provided ideas, guidance and logistic support; JK designed the research project and undertook the research. All authors contributed to the article and approved the submitted version.

Funding

Funding was provided by the Darwin Initiative, UK (Project 27-014), the Amar-Franses Foster-Jenkins Trust, and the Bentham-Moxon Trust.

Acknowledgments

We gratefully acknowledge the National Forest Authority (NFA) and the Ugandan Wildlife Authority (UWA), for granting permission to study wild populations of *Coffea* in Uganda. We thank Moses Lumu and Robert Acidri at the National Agricultural Research Organization (NARO) for fieldwork support. We thank the directors and staff at the Royal Botanic Gardens, Kew (K); Natural History Museum, London (BM); Makerere University (MHU); Muséum National d'Histoire Naturelle, Paris (P); National Botanic Garden of Belgium, Meise (BR); and National Herbarium Nederland, Wageningen (WAG), for access to herbarium collections. We thank Kyagalanyi Coffee Ltd. (ED&F Man/Volcafe, Uganda), Equatoria Teak Company (Kenya/South Sudan), and Clifton Coffee Roasters (Bristol, UK), for information concerning excelsa coffee.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alègre, C. (1959). Climates et caféiers d´Arabie. *Agron. Trop.* 14, 23–58.

Anon (1890). Liberian Coffee. *Bull. Misc. Inform. (Royal Bot. Gard. Kew).* 47, 245–253.

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., et al. (2015). The coffee rust crises in Colombia and central America, (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur.* 7, 303–321. doi: 10.1007/s12571-015-0446-9

Bachman, S., Moat, J., Hill, A., de Torre, J., and Scott, B. (2011). Supporting red list threat assessments with GeoCAT: geospatial conservation assessment tool. *ZooKeys.* 126, 117–126. doi: 10.3897/zookeys.150.2109

Baltazar, M. D., and Buot, I. E. (2019). Leaf architectural analysis of taxonomic confusing coffee species: *Coffea liberica* and *Coffea liberica* var. *dewevrei*. *Biodiversitas.* 20, 1560–1567. doi: 10.13057/biodiv/d200611

Bharatha Nandhini, R. M., Rahul, R. N., Thilaga, S., Surya Prakash Rao, N., and Ganesh, D. (2013). Molecular distinction of C×R hybrid (*Coffea congensis*×*Coffea canephora*) from morphologically resembling male parent using rbcL and matK gene sequences. *S. Afr. J. Bot.* 88, 334–340. doi: 10.1016/j.sajb.2013.08.011

Bridson, D. M. (1988). "*Coffea*," in Flora of Tropical East Africa, Rubiaceae, part 2. Eds. R. M. Polhill, D. M. Bridson and B. Verdcourt (Rotterdam/Brookfield: Balkema), 703–723.

Bullock, A. A. (1930). Nandi coffee. *Bul. Misc. Inform. (Royal Botanic Gardens Kew).* 1930 (9), 401–402. doi: 10.2307/4107660

Busby, J. R. (1991). "BIOCLIM – a bioclimatic analysis and prediction system," in *Nature conservation: cost effective biological surveys and data analysis.* Eds. C. R. Margules and M. P. Austin (Canberra, Australia: Commonwealth Scientific and Industrial Research Organisation (CSIRO), 64–68.

Carvalho, F. G., Sera, G. H., Andreazi, E., Sera, T., de Barista Fonseca, I. C. B., Carducci, C., et al. (2017). Drought tolerance in seedlings of coffee genotypes carrying genes of different species. *Coffee Sci.* 12, 9–16.

Chadburn, H., and Davis, A. P. (2017a). *Coffea canephora*. In: *The IUCN red list of threatened species 2017.* Available at: https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T18290186A18539466.en (Accessed 19 September 2022).

Chadburn, H., and Davis, A. P. (2017b). *Coffea liberica*. In: *The IUCN red list of threatened species 2017.* Available at: https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T18537594A18539526.en (Accessed 19 September 2022).

Charr, J.-C., Garavito, A., Guyeux, C., Crouzillat, D., Descombes, P., Fournier, C., et al. (2020). Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on *Coffea canephora* (robusta coffee). *Mol. Phyl. Evol.* 151, 106906. doi: 10.1016/j.ympev.2020.106906

Cheney, R. H. (1925). *A monograph of the economic species of the genus Coffea L.* (The New York University Press).

Chevalier, A. (1903). Notes préliminaires sur qualques caféiers sauvages noveaux ou peu connus de L'Afrique centrale. *Rev. Cult. Colon.* 12, 256–259.

Chevalier, A. (1929). Les Caféiers du globe, fasc. 1: Généralités sur les caféiers. *Ency. Biol.* 5, 1–196.

Chinnappa, C. C. (1970). Interspecific hybrids of *Coffea canephora* and *C. liberica*. *Genetica.* 41, 141–146. doi: 10.1007/BF00958900

Clarindo, W. R., Carvalho, C. R., Eveline, T. C., and Koehler, A. D. (2013). Following the track of "Híbrido de Timor" origin by cytogenetic and flow cytometry approaches. *Genet. Resour. Crop Evol.* 60, 2253–2259. doi: 10.1007/s10722-013-9990-3

Couturon, E., Lashermes, P., and Charrier, A. (1998). First intergeneric hybrids (*Psilanthus ebracteolatus* Hiern × *Coffea arabica* L.) in coffee trees. *Can. J. Bot.* 76, 542–546. doi: 10.1139/b98-017

Cramer, P. J. S. (1913). Gevens over de variabiliteit van de in nederlandsch-indië verbouwde koffie-sorten. *Batavia. Uitgaande Dept. Landbouw Mededeeling.* 1, 1–696.

Cramer, P. J. S. (1957). *A review of literature of coffee research in Indonesia.* (Turrialba: SIC Editorial, Inter-American Institute of Agricultural Sciences).

Cubry, P., De Bellis, F., Pot, D., Musoli, P., and Leroy, T. (2013). Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. *Genet. Res. Crop Evol.* 60 (2), 483–501. doi: 10.1007/s10722-012-9851-5

DaMatta, F. M., and Ramalho, J. D. C. (2006). Impacts of drought and temperature stress on coffee physiology and production: a review. *Braz. J. Plant Physiol.* 18, 55–81. doi: 10.1590/S1677-04202006000100006

Davis, A. P., Bridson, D. M., and Rakotonasolo, F. (2005). "A reexamination of *Coffea* subgenus *Baracoffea* and comments on the morphology and classification of *Coffea* and *Psilanthus* (Rubiaceae-Coffeeae)," in A festschrift for William G. D'Arcy - The legacy of a taxonomist. *Monographs in Systematic Botany.*, 104 (St. Louis: MBG Press).

Davis, A. P., Chadburn, H., Moat, J., O'Sullivan, R., Hargreaves, S., and Nic Lughadha, E. (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. *Sci. Adv.* 5, eaav3473. doi: 10.1126/sciadv.aav3473

Davis, A. P., Gargiulo, R., Almeida, I. N. d. M., Caravela, M. I., Denison, C., et al. (2021a). Hot coffee: the identity, climate profiles, agronomy, and beverage characteristics of *Coffea racemosa* and *C. zanguebariae*. *Front. Sust. Food Syst.* 5, 740137. doi: 10.3389/fsufs.2021.740137

Davis, A. P., Gargiulo, R., Fay, M. F., Sarmu, D., and Haggar, J. (2020). Lost and found: *Coffea stenophylla* and *C. affinis*, the forgotten coffee crop species of West Africa. *Front. Plant Sci.* 11, 616. doi: 10.3389/fpls.2020.00616

Davis, A. P., Govaerts, R., Bridson, D. M., and Stoffelen, P. (2006). An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot. J. Linn. Soc.* 152, 465–512. doi: 10.1111/j.1095-8339.2006.00584.x

Davis, A. P., Kiwuka, C., Faruk, A., Walubiri, M. J., and Kalema, J. (2022). The re-emergence of Liberica coffee as a major crop plant. *Nat. Plants.* 8, 1322–1328. doi: 10.1038/s41477-022-01309-5

Davis, A. P., Mieulet, D., Moat, J., Sarmu, D., and Haggar, J. (2021b). Arabica-like flavour in a heat-tolerant wild coffee species. *Nat. Plants.* 7, 413–418. doi: 10.1038/s41477-021-00891-4

Davis, A. P., and Rakotonasolo, F. (2021). Six new species of coffee (*Coffea*) from northern Madagascar. *Kew Bull.* 76, 497–511. doi: 10.1007/s12225-021-09952-5

Davis, A. P., Tosh, J., Ruch, N., and Fay, M. F. (2011). Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.* 167, 357–377. doi: 10.1111/j.1095-8339.2011.01177.x

Durand, T., and De Wildeman, E. (1899). Matériaux pour le flore du Congo. *Bull. Soc R. Belgique.* 38, 171–200.

Eggeling, W. J., and Dale, I. R. (1952). *The indigenous trees of the Ugandan protectorate* (Glasgow: Robert MacLehose & Co. Ltd. The University Press).

Fazouli, L. C., Perez Maluf, P., Guerreiro Filho, O., Medina Filho, H. P., and Silvarolla, M. B. (2000). "Breeding and biotechnology of coffee," in *Coffee biotechnology and quality.* Eds. T. Sera, C. R. Soccol, A. Pandey and S. Roussos (Dordrecht: Springer), 27–45.

Freeman, W., and Chandler, S. (1907). *The world's commercial products.* (London: Sir Isaac Pitman and Sons, Ltd).

Gaitán, A. L., Cristancho, M. A., Castro Caicedo, B. L., Rivillas, C. A., and Cadena Gómez, G. (2015). *Compendium of coffee diseases and pests.* (St. Paul: APS Press).

Ganesh, D., Ram, A. S., Prakash, N. S., Mishra, M. K., Jamsheed, A., Jagadeesan, M., et al. (2002). "Evaluation of *Coffea liberica* × *Coffea eugenioides* and its progenies for yield, leaf rust tolerance and quality," in *Proceedings of the 15th Plantation Crops Symposium Placrosym XV*, Mysore, India, 10–13 December, 2002. 72–77.

Garavito, A., Montagnon, C., Guyot, R., and Bertrand, B. (2016). Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biol.* 16, 242. doi: 10.1186/s12870-016-0933-y

Gomez, C., Dussert, S., Hamon, P., Hamon, S., De Kochko, A., and Poncet, V. (2009). Current genetic differentiation of *Coffea canephora* Pierre ex A.Froehn. in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol. Biol.* 9, 167. doi: 10.1186/1471-2148-9-167

Govaerts, R., Rhusam, M., Frodin, D., Bridson, D., Dawson, S., and Davis, A. P. (2022). "*Coffea* in world checklist of Rubiaceae," *The Royal Botanic Gardens, Kew.* Available at: http://wcsp.science.kew.org/. Retrieved 18 July 2022.

Grisi, F. A., Donizeti, J., de Castro, E. M., de Oliveira, C., Biagiotti, G., and de Melo, L. A. (2008). Leaf anatomical evaluations in 'catuaí' and 'siriema' coffee seedlings submitted to water stress. *Ciênc. Agrotec.* 32, 1730–1736. doi: 10.1590/S1413-70542008000600008

Haarer, A. E. (1962). *Modern coffee production.* (London: Leonard Hill [Books] Limited).

Holmgren, P. K., Holmgren, N. H., and Barnett, L. C. (1990). "Index herbariorum. part 1," in *The herbaria of the world., 8th edn* (New York Botanical Garden: Regnum Vegetabile).

International Coffee Organization (ICO) (2022). *Trade statistics.* Available at: http://www.ico.org/trade_statistics.asp (Accessed 7 Aug 2022).

IUCN Standards and Petitions Subcommittee (2022). *Guidelines for using the IUCN red list categories and criteria.* (Prepared by the Standards and Petitions Committee of the IUCN Species Survival Comission). Available at: https://www.iucnredlist.org/resources/redlistguidelines. Version 15.1 (July 2022).

Kalema, J., and Beentje, H. (2012). *Conservation checklist of the trees of Uganda.* (Kew: Royal Botanic Gardens).

Kalema, J., and Hamilton, A. (2020). *Field guide to the forest trees of Uganda: for identification and conservation.* (Croydon: CABI).

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci. Data.* 4, 170122. doi: 10.1038/sdata.2017.122

Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0.* Available at: https://CRAN.R-project.org/package=ggpubr.

Katende, A. B., Ssegawa, P., Birnie, A., Holding, C., and Tengnäs, B. (1999). *Wild food plants and mushrooms of Uganda.* (Nairobi: Regional Land Management Unit/Sida).

Kath, J., Byrareddy, V. M., Craparo, A., Nguyen-Huy, T., Mushtaq, S., Cao, L., et al. (2020). Not so robust: robusta coffee production is highly sensitive to temperature. *Glob. Change Biol.* 26, 3677–3688. doi: 10.1111/gcb.15097

Kiwuka, C., Goudsmit, E., Tournebize, R., de Aquino, S. O., Douma, J. C., Bellanger, L., et al. (2021). Genetic diversity of native and cultivated Ugandan robusta coffee (*Coffea canephora* Pierre ex A. Froehner): climate influences, breeding potential and diversity conservation. *PloS One.* 16, e0245965. doi: 10.1371/journal.pone.0245965

Lebrun, J. (1941). Recherches morphologiques et systématiques sur les caféiers du Congo. *Congo Belge: Publ. L'instit. Nation. L'etude Agron.* 11, 1–186.

Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y., and Fay, M. F. (2007). Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. *Ann. Bot.* 100, 1565–1583. doi: 10.1093/aob/mcm257

McCook, S. (2014). "Ephemeral plantations: The rise and fall of Liberian coffee 1870–1900," in *Comparing apples, oranges, and cotton. Environmental histories of the plantation.* Ed. F. Uekötter (Frankfurt/New York: Campus Verlag), 85–112.

McCook, S. (2019). *Coffee is not forever. A global history of coffee leaf rust* (Ohio: Ohio University Press).

Medina Filho, H. P., Carvalho, A., and Medina, D. M. (1977a). Germoplasma de *C. racemosa* e seu potencial no melhoramento do cafeeiro. *Bragantia.* 36, 43–46. doi: 10.1590/S0006-87051977000100040

Medina Filho, H. P., Carvalho, A., and Monaco, L. (1977b). Melhoramento do cafeeiro. XXXVII – observações sobre a resistência do cafeeiro ao bicho mineiro. *Bragantia.* 36, 131–137. doi: 10.1590/S0006-87051977000100011

Melo, E. F., Fernandes-Brum, C. N., Pereira, F. J., de Castro, E. M., and Chalfun-Júnior, A. (2014). Anatomic and physiological modifications in seedlings of *Coffea arabica* cultivar Siriema under drought conditions. *Ciênc. Agrotec.* 38, 25–33. doi: 10.1590/S1413-70542014000100003

Merot-L'anthoene, V., Tournebize, R., Darracq, O., Rattina, V., Lepelley, M., Bellanger, L., et al. (2019). Development and evaluation of a genome-wide coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. *Pl. Biotechn. J.* 17, 1418–1430. doi: 10.1111/pbi.13066

Moat, J., Gole, T. W., and Davis, A. P. (2019). Least concern to endangered: applying climate change projections profoundly influences the extinction risk assessment for wild Arabica coffee. *Glob. Change Biol.* 25, 390–403. doi: 10.1111/gcb.14341

Moat, J., Williams, J., Baena, S., Wilkinson, T., Gole, T. W., Challa, Z. K., et al. (2017). Resilience potential of the Ethiopian coffee sector under climate change. *Nat. Plants.* 3, 17081. doi: 10.1038/nplants.2017.81

Mulindwa, J., Kaaya, A. N., Muganga, L., Paga, M., Musoli, P., Sseremba, G., et al. (2022). Cup quality profiles of robusta coffee wilt disease resistant varieties grown in three agro-ecologies in Uganda. *J. Sci. Food Agric.* 102, 1225–1232. doi: 10.1002/jsfa.11460

Musoli, P. C., Cilas, C., Pot, D., Nabaggala, A., Nakendo, S., Pande, J., et al. (2013). Inheritance of resistance to coffee wilt disease (*Fusarium xylarioides* Steyaert) in robusta coffee (*Coffea canephora* Pierre) and breeding perspectives. *Tree Genet. Genomes.* 9, 351–360. doi: 10.1007/s11295-012-0557-9

Musoli, C. P., Kangire, A., Leroy, T., Nabaggala, A., Nakendo, S., Olal, S., et al. (2009). *Towards a variety resistant to coffee wilt disease (CWD) : a case for robusta coffee (Coffea canephora) in Uganda 22nd InternationalConference on coffee science; 14–19 september 2008; campinas, SP,Brazil.* (Paris: Association Scientifique Internationale du Café (ASIC), 1472–1479.

Myers, R., Kawabata, A., Cho, A., and Nakamoto, S. T. (2020). Grafted coffee increases yield and survivability. *Hortic. Tech.* 30, 428–432. doi: 10.21273/HORTTECH04550-20

N'Diaye, A. N., Poncet, V., Louran, J., Hamon, S., and Noirot, M. (2005). Genetic differentiation between *Coffea liberica* var. *liberica* and *C. liberica* var. *dewevrei* and comparison with *C. canephora*. *Plant Syst. Evol.* 253, 95–104. doi: 10.1007/s00606-005-0300-1

Nagai, C., Rakotomalala, J.-J., Katahira, R., Li, Y., Yamagata, K., and Ashihara, H. (2008). Production of a new low-caffeine hybrid coffee and the biochemical mechanism of low caffeine accumulation. *Euphytica.* 164, 133–142. doi: 10.1007/s10681-008-9674-9

Narasimhaswamy, R. L. (1960). Arabica selection S.795-its origin and performance – a study. *Indian Coffee.* 24, 197–204.

NextGIS (2019). *QuickMapServices version 19.11.1.* Available at: https://plugins.qgis.org/plugins/quick_map_services/ (Accessed 12 September 2022).

Ngugi, K., and Aluka, P. (2019). Genetic and phenotypic diversity of robusta coffee (*Coffea canephora* L.). *Caffein. Cocoa Based Bever.* 8, 89–130. doi: 10.1016/B978-0-12-815864-7.00003-9

O'Sullivan, R. J., Chadburn, H., and Davis, A. P. (2017). *Coffea neoleroyi (errata version published in 2020). the IUCN red list of threatened species 2017.* Available at: https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T18290348A176942734.en (Accessed 19 September 2022).

O'Sullivan, R. J., Duarte, A., and Davis, A. P. (2020). *Coffea eugenioides (amended version of 2017 assessment). the IUCN red list of threatened species 2020.* Available at: https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T18290318A174150624.en (Accessed 03 August 2022).

Panaligan, A. C., Baltazar, M. D., and Alejandro, G. J. D. (2020). Genetic polymorphism of registered and popularly cultivated coffee (*Coffea* spp.) in the Philippines using inter-simple sequence repeats markers. *Biodiversitas.* 21, 4228–4233. doi: 10.13057/biodiv/d210938

Panaligan, A. C., Baltazar, M. D., and Alejandro, G. J. D. (2021). Molecular authentication of commercially cultivated coffee (*Coffea* spp.) in the Philippines using DNA barcodes. *Int. J. Agric. Biol.* 25, 27–230. doi: 10.17957/IJAB/15.1660

Phiri, N., and Baker, P. S. (2009). *Coffee wilt in Africa. Final technical report of the regional coffee wilt programme 2000–07).* (CAB International).

QGIS Development Team (2022). *QGIS geographic information system.* (Open Source Geospatial Foundation Project). Available at: http://qgis.osgeo.org.

R Core Team (2020). *R: A language and environment for statistical computing.* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Reddy, A. G. S., Raju, K. V. V. S. N., and Dharmaraj, P. S. (1985). Breeding behaviour of ligenioides — a spontaneous amphidiploid between *Coffea liberica* and *C. eugenioides*. *J. Coffee Res.* 15, 33–37.

Rutherford, M. A. (2006). Current knowledge of coffee wilt disease, a major constraint to coffee production in Africa. *Phytopathology.* 96, 663–666. doi: 10.1094/PHYTO-96-0663

Stoffelen, P. (1998). *Coffea and Psilanthus in Tropical Africa: A systematic and palynological study, including a revision of the West and central African species.* (Katholieke Universiteit Leuven).

Surya Prakash, N., Combes, M.-C., Somanna, N., and Lashermes, P. (2002). AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica.* 124, 265–271. doi: 10.1023/A:1015736220358

Thiers, B. (2019). *Index herbariorum: A global directory of public herbaria and associated staff.* (New York Botanical Garden's Virtual Herbarium). Available at: http://sweetgum.nybg.org/science/ih/ (Accessed 11 September 2019).

Thomas, A. S. (1935). Types of robusta coffee and their selection in Uganda. *East Afr. Agr. J.* 1, 193–197. doi: 10.1080/03670074.1935.11663646

Thomas, A. S. (1940a). "Section IX. a. robusta coffee," in *Agriculture in Uganda.* Ed. J. D. Tothill (London: Oxford University Press), 289–311.

Thomas, A. S. (1940b). "Section IX. b. other coffees," in *Agriculture in Uganda.* Ed. J. D. Tothill (London: Oxford University Press), 311–313.

Thomas, A. S. (1940c). "Section x. Arabica coffee. i. history and general," in *Agriculture in Uganda.* Ed. J. D. Tothill (London: Oxford University Press), 314–325.

Thomas, A. S. (1942). The wild Arabica coffee on the Boma Plateau. Anglo-Egyptian Sudan. *Empire J. Expt. Agric.* 10, 207–212.

Thomas, A. S. (1944). The wild coffees of Uganda. *Empire J. Expt. Agric.* 12, 1–12.

Tothill, J. D. (1940). *Agriculture in Uganda.* (London: Oxford University Press).

Uganda Coffee Development Authority (UCDA) (2017). *Uganda — country coffee profile.* Available at: https://ugandacoffee.go.ug/resource-center/publications.

Venancio, L. P., Filgueiras, R., Mantovani, E. C., do Amaral, C. H., da Cunha, F. F., dos Santos Silva, F. C., et al. (2020). Impact of drought associated with high temperatures on *Coffea canephora* plantations: a case study in espírito Santo state, Brazil. *Sci. Rep.* 10, 19719. doi: 10.1038/s41598-020-76713-y

Wellman, F. L. (1961). *Coffee: Botany, cultivation and utilization*. (London: Leonard Hill [Books] Limited; New York: Interscience Publishers, Inc).

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

Wrigley, G. (1988). *Coffee – tropical agriculture series.* (Harlow: Longman Scientific & Technical).

Yonas, B., Bayetta, B., and Chemeda, F. (2014). Assessments of resistances of indigenous Arabica coffee genotypes for multiple diseases. *J. Plant Breed. Crop Sci.* 2, 54–61. doi: 10.15580/GJPBCS.2014.3.101013898

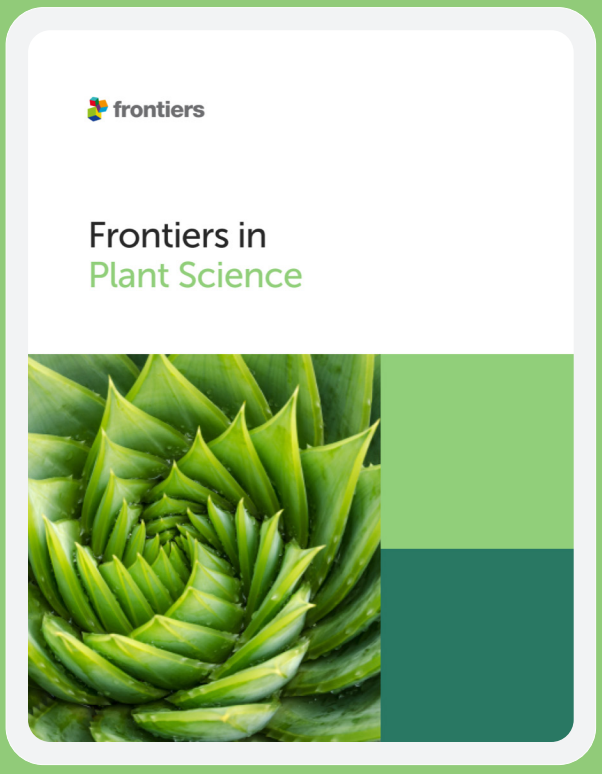
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact