

Population Genomics Reveals Demographic History and Climate Adaptation in Japanese *Arabidopsis halleri*

Ryo A. Suda¹, Shosei Kubota², Vinod Kumar³, Vincent Castric⁴, Ute Krämer³, Shin-Ichi Morinaga⁵ and Takashi Tsuchimatsu^{1,*}

¹Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan, ²Fasmac Co., Ltd., 3088 Okada, Atsugi, Kanagawa, 243-0021 Japan, ³Molecular Genetics and Physiology of Plants, Ruhr University Bochum, Universitätsstrasse 150, ND3/30, Bochum D-44801, Germany, ⁴University Lille, CNRS, UMR 8198—Evo-Eco-Paleo, Lille F-59000, France, ⁵Department of Natural & Environmental Science, Teikyo University of Science, 2-2-1 Senju-Sakuragi, Adachi-ku, Tokyo, 120-0045 Japan

*Corresponding author: E-mail, tsuchimatsu@bs.s.u-tokyo.ac.jp

Received 8 April 2024; Accepted 25 September 2024

Climate oscillations in the Quaternary forced species to major latitudinal or altitudinal range shifts. It has been suggested that adaptation concomitant with range shifts plays key roles in species responses during climate oscillations, but the role of selection for local adaptation to climatic changes remains largely unexplored. Here, we investigated population structure, demographic history and signatures of climate-driven selection based on genome-wide polymorphism data of 141 Japanese *Arabidopsis halleri* individuals, with European ones as outgroups. Coalescent-based analyses suggested a genetic differentiation between Japanese subpopulations since the Last Glacial Period (LGP), which would have contributed to shaping the current pattern of population structure. Population demographic analysis revealed the population size fluctuations in the LGP, which were particularly prominent since the subpopulations started to diverge (~50, 000 years ago). The ecological niche modeling predicted the geographic or distribution range shifts from southern coastal regions to northern coastal and mountainous areas, possibly in association with the population size fluctuations. Through genome-wide association analyses of bioclimatic variables and selection scans, we investigated whether climate-associated loci are enriched in the extreme tails of selection scans, and demonstrated the prevailing signatures of selection, particularly toward a warmer climate in southern subpopulations and a drier environment in northern subpopulations, which may have taken place during or after the LGP. Our study highlights the importance of integrating climate associations, selection scans and population demographic analyses for identifying genomic signatures of population-specific adaptation, which would also help us predict the evolutionary responses to future climate changes.

Keywords: *Arabidopsis halleri* • Climate change • Demographic history • Glacial cycles • Local adaptation • Population genomics

Introduction

Climate oscillations that occurred in the Quaternary period (2.6 million years ago to present) forced species to shift to major latitudinal or altitudinal ranges (Comes and Kadereit 1998, Hewitt 2004, Schmitt 2007, Qiu et al. 2011). In glacial periods, some locations served as climate refugia, where species were allowed to persist in local habitats, and they expanded from there when climate conditions improved (Provan and Bennett 2008, Gavin et al. 2014). Repeated environmental changes would lead to geographic or distribution range shifts, contractions, expansions and subdivisions of populations that profoundly shaped the present-day distributions and genetic structures of many plant and animal species (Hewitt 2000, 2004).

It has been suggested that adaptation concomitant with geographic or distribution range shifts plays key roles in species responses during climate oscillations in the Quaternary (Davis and Shaw 2001, Excoffier et al. 2009, Keller et al. 2011, De Lafontaine et al. 2018, Luqman et al. 2023). Understanding the mechanisms of local adaptation to past climate changes is also crucial for predicting the evolutionary responses to future environmental changes (Waldvogel et al. 2020). However, the influence of past climate oscillations has mainly been studied through the lens of neutral genetic variation, and knowledge is still limited about the role of selection in the local adaptation of populations to climate changes and associated range shifts (De Lafontaine et al. 2018, Luqman et al. 2023), except for a few pioneering studies (The 1001 Genomes Consortium 2016, Luqman et al. 2023).

Whole-genome polymorphism data from large population samples can be powerful for studying local adaptation in response to climate change in various ways (Weigel and Nordborg 2015, Bamba et al. 2019, Bourgeois and Warren 2021). First, the advance of statistical modeling based on coalescent theory enables us to infer the demographic history of populations, including historical changes in effective population size and the timing of population divergence during glacial cycles. Second, genome-wide association studies (GWAS) can identify genes underlying phenotypes or other characteristics, such as the climate of sampled locations, enabling us to investigate genotype–environment associations (Hancock et al. 2011, Savolainen et al. 2013, Rellstab et al. 2015, The 1001 Genomes Consortium 2016, Lee et al. 2017). Third, genome-wide selection scans can detect loci under various selection modes, including selective sweeps and divergent selection. Importantly, climate adaptation is suggested to be typically a polygenic process, in which multiple genes are involved in the course of adaptation to a climatic condition (Healy et al. 2018, Barghi et al. 2020, Yang et al. 2022, Ma et al. 2024). Consequently, selection signals of each locus are often not strong, making it difficult to detect climate-driven selection. A common feature of empirical tests for polygenic adaptation including climate adaptation is thus to detect genome-wide signals by combining independent information from polymorphism data, such as selection scans and genotype–environment associations (Barghi et al. 2020, Yang et al. 2022).

The Japanese archipelago is an ideal area to study the mechanisms of how populations adapt to local climatic changes and associated range shifts, as it stretches 3,000 km from northeast to southwest over a wide range of climatic zones from subarctic to subtropical and has a complex mountainous topography in Central Japan (Ohsawa and Ide 2011). Population structure analyses of several widely distributed species in the Japanese archipelago suggested range shifts during glacial cycles (Ohsawa and Ide 2011, Iwasaki et al. 2012, Sakaguchi et al. 2018, Magota et al. 2021), but investigations on the climate adaptation at a region-wide scale are still scarce. This study focused on *Arabidopsis halleri*, a perennial herb distributed in East Asia and Europe. This species has been used extensively for evolutionary and ecological studies (reviewed in Koch 2018, Honjo and Kudoh 2019), such as self-incompatibility (Castric and Veke­mans 2004, Durand et al. 2020), heavy metal hyperaccumulation (Krämer 2010, Stein et al. 2017) and adaptation to high altitudes (Kubota et al. 2015, Yumoto et al. 2021, Yoshida et al. 2023). Five subspecies are recognized in *A. halleri*: *A. halleri* subsp. *halleri*, *A. halleri* subsp. *tatrica*, *A. halleri* subsp. *dacica*, *A. halleri* subsp. *ovirensis*, and *A. halleri* subsp. *gemmaifera* (Honjo and Kudoh 2019). *Arabidopsis halleri* subsp. *gemmaifera* is distributed throughout Japan, Korea, north-eastern China and the Russian Far East, while the other four subspecies are distributed in Europe. In Japan, *A. halleri* subsp. *gemmaifera* is observed on the four main islands (Kyushu, Honshu, Shikoku and Hokkaido). Given its wide distribution in the diverse climates across the Japanese archipelago, *A. halleri* can serve as

a model species to study how plant populations are locally adapted to diverse climates and their changes during glacial cycles. Population structure and demographic histories are the fundamental basis for studying local adaptation, but the information is currently limited except for a few studies, including one based on microsatellite markers (Sato and Kudoh 2014) and two restricted to a microgeographic scale (Kubota et al. 2015, Yoshida et al. 2023).

In this study, using the whole-genome resequencing data of 141 Japanese *A. halleri* individuals and an outgroup of 16 European individuals, we aimed to understand how *A. halleri* adapted to local climates during glacial cycles through a population genomics approach. Specifically, we addressed the following questions:

- 1) How are Japanese populations of *A. halleri* genetically structured, and what demographic processes, such as population size fluctuations and divergence, formed the current pattern of population structure?
- 2) Which loci are associated with bioclimatic variables of sampled locations?
- 3) Do climate-associated loci tend to be under selection? In other word, are there any signatures of polygenic adaptation? If so, in which population selection has acted?
- 4) How have the potential habitats of *A. halleri* shifted since the Last Glacial Period (LGP)?

Results

Population structure of *A. halleri* in Japan

We resequenced the genomes of 141 individuals from 135 *A. halleri* wild populations distributed across Kyushu, Honshu and Hokkaido islands in Japan using short reads (Supplementary Table S1). As outgroups, we also obtained sequence data of 12 and four individuals from 10 Central European populations and four Romanian populations, respectively. Four of the Central European individuals were classified as *A. halleri* subsp. *tatrica* and the others as *A. halleri* subsp. *halleri* (Supplementary Table S1). We mapped short reads to the *A. halleri* v2.03 assembly (DOE-JGI, <http://phytozome.jgi.doe.gov/>), obtaining 1,506,831 single-nucleotide polymorphisms (SNPs) in total after variant calling and filtering. Network analysis using this SNP set revealed a genetic structure in accordance with geography: the Japanese *A. halleri* formed a cluster distinct from Central European and Romanian clusters, and among the Japanese accessions, Northern Japan (NJ) individuals formed a single group, separated from other individuals of Western Japan (WJ), Kansai Region (KR) and Central Japan (CJ) subpopulations (Fig. 1A). A clustering analysis using the non-negative matrix factorization (sNMF) also revealed a population structure that reflects geography in Japan (Fig. 1B, Supplementary Fig. S1A). The clustering pattern in $K = 5$ was largely concordant with the network-based analysis (Fig. 1B), while the best-supported number of clusters was $K = 7$ based on the cross-entropy criterion (Supplementary Fig. S1B). Overall, these results suggest that Japanese *A. hal-*

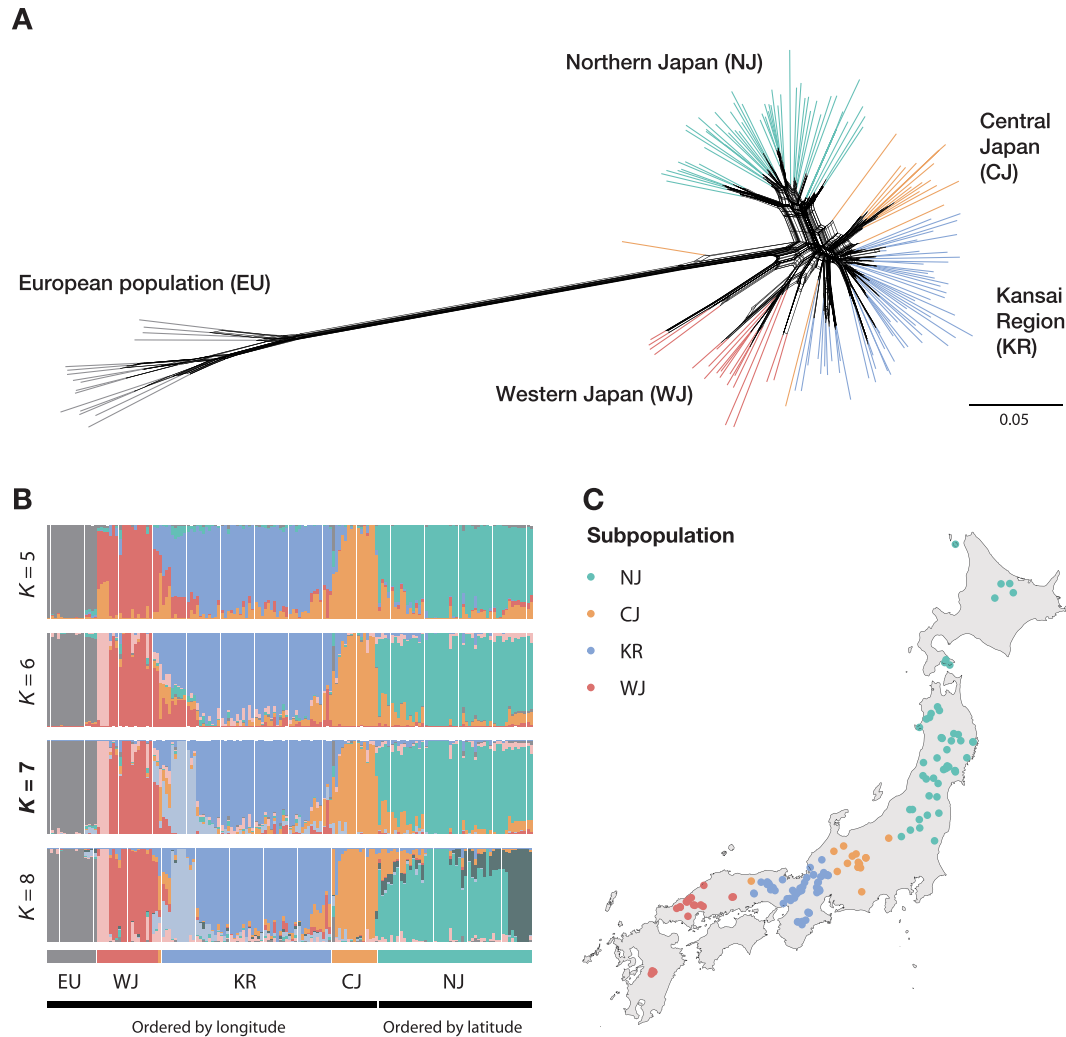


Fig. 1 Population structure of *A. halleri*. (A) Phylogenetic network generated by SplitsTree4 v.4.19.2 (Huson and Bryant 2006). External branches were colored according to the classification based on a maximum factor under $K = 5$ of the clustering analysis. The scale bar indicates the proportion of positions with different alleles. (B) Clustering by non-negative matrix factorization using the R package LEA v.3.10.2 (Frichot et al. 2014, Frichot and François 2015). Results for $K = 2$ –12 are available in [Supplementary Fig. S1](#). Individuals are ordered by longitude for the EU, WJ, KR, and CJ subpopulations and by latitude for the NJ subpopulation. (C) Geographic distribution of Japanese individuals with the classification based on a maximum factor under $K = 5$.

leri are geographically differentiated across the archipelago. For the following analyses of demographic history, individuals were classified into five subpopulations based on a maximum factor under $K = 5$ (**Fig. 1C**): Europe (EU), WJ, KR, CJ, and NJ.

Demographic history and population subdivisions

We employed the approximate Bayesian computation (ABC) analysis to infer demographic history and divergence time among subpopulations. In this analysis, our primary focus was to obtain the approximate timescale of divergence time between subpopulations. Therefore, while more complex scenarios involving migrations and hybridizations could be considered, we examined four relatively simple scenarios consistent with the patterns of hierarchical clustering of population

structure ([Supplementary Fig. S2A](#); **Fig. 1**). Among them, scenario 3 was chosen as the best scenario with a posterior probability of 0.814, in which the WJ subpopulation first diverged from the other subpopulations from Japan. Under scenario 3, the split time between the European and Japanese subpopulations (t_4) was estimated to be 71,800 [95% credible interval of 35,400–140,900] generations ago (**Fig. 2A**, [Supplementary Fig. S2](#), [Table S2](#)). Assuming a generation time of 2 years (Roux et al. 2011, Sato and Kudoh 2014, Honjo and Kudoh 2019), this estimate corresponds to 143,600 [70,800–281,800] years ago. This time estimate of the Europe–Japan split within *A. halleri* was significantly smaller than the estimated divergence time between *A. halleri* and its closest relative, *A. lyrata*, which was about 337,000 [272,800–438,200] years ago (Roux et al. 2011). All the split events within Japanese subpopulations were inferred

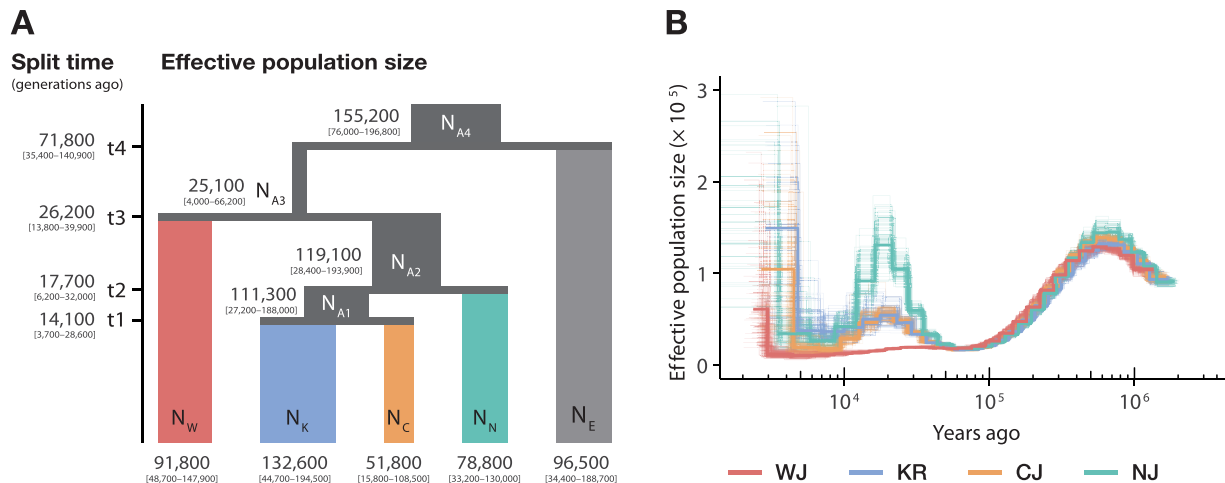


Fig. 2 Demographic history of Japanese *A. halleri*. (A) The best demographic model suggested by the approximate Bayesian computation analysis. The DIYABC-RF v.1.2.1 software was used (Collin et al. 2021). Effective population sizes of current and ancestral populations and split times are shown with 95% credible intervals. (B) The pairwise sequential Markovian coalescent (PSMC) analysis of four Japanese subpopulations (Li and Durbin 2011). Thick lines represent the PSMC estimates on the original sequence for each individual, and thin lines indicate the results for 100 bootstrap runs.

to have occurred during the LGP (about 115,000–11,700 years ago) (EPICA community members 2004): $t_3 = 52,400$ [27,600–79,800], $t_2 = 35,400$ [12,400–64,000] and $t_1 = 28,200$ [7,400–57,200] years ago, assuming a generation time of 2 years. The ancestral population of all Japanese subpopulations was estimated to have the smallest effective population size (N_e) ($N_{A3} = 25,100$ [4,000–66,200]).

Although scenario 3 was the best model, scenario 2 was comparable in terms of the classification vote, the number of times a scenario is selected in 1,000 random forest trees (Supplementary Fig. S2B). Nevertheless, both scenarios provided similar split times and effective population sizes, and all the split events within Japanese subpopulations were inferred to have occurred during the LGP under scenario 2 as well. (Supplementary Table S2).

We also performed the pairwise sequential Markovian coalescent (PSMC) analysis to gain further insights into the demographic history within Japanese populations (Li and Durbin 2011). We chose representative individuals for each Japanese subpopulation based on the result of clustering analysis to infer the historical changes in N_e (see 'Materials & Methods' section for details). Assuming a generation time of 2 years, all four subpopulations showed similar dynamics of N_e before ~50,000 years ago, but the more recent patterns were different between subpopulations: We observed an increase of N_e between 50,000 and 20,000 years ago followed by a decrease until 7,000 years ago in CJ, KR and NJ subpopulations, while N_e of the WJ subpopulation declined for the period between 30,000 and 3,000 years ago with a slight expansion before 30,000 years ago (Fig. 2B). While we note that the PSMC-based estimate of the recent past is relatively less reliable due to the limited number of independent coalescent events, different population size

dynamics after 50,000 years ago suggests genetic differentiation between subpopulations, in line with the ABC analysis.

Genomic signatures of climate adaptation in Japan

Next, we investigated the signatures of climate adaptation in the genomes of Japanese *A. halleri*. Our approach is summarized in Fig. 3A. First, we identified loci associated with local climate through genome-wide association mapping. Second, we performed genome-wide selection scans based on the cross-population extended haplotype homozygosity (XP-EHH) to detect subpopulation-specific selection (Sabeti et al. 2007, Gautier and Vitalis 2012, Gautier et al. 2017, Klassmann and Gautier 2022). Then, we asked if climate-associated loci are enriched in the extreme tails of selection scans. The statistical significance of the enrichments was assessed by permutation tests, with a null hypothesis that the fraction of the extreme tails of XP-EHH scores in climate-associated loci is the same as the fraction in the whole genome (climate-associated and climate-non-associated loci) (see 'Materials & Methods' section for details). Thus, high enrichment scores indicate that climate-associated loci tend to overlap with the positively selected regions more than expected by chance. Since the XP-EHH-based selection scan detects the direction of selection, we performed the enrichment analyses for all the pairwise comparisons of the four subpopulations in Japan to detect climate-associated subpopulation-specific local adaptation.

To perform these selection analyses, information on the population structure is required. Because these analyses focus on Japanese populations, we re-analyzed population structure only with Japanese individuals (Supplementary Fig. S3A). We confirmed that the pattern of clustering was fairly close to the one when including European populations, and the

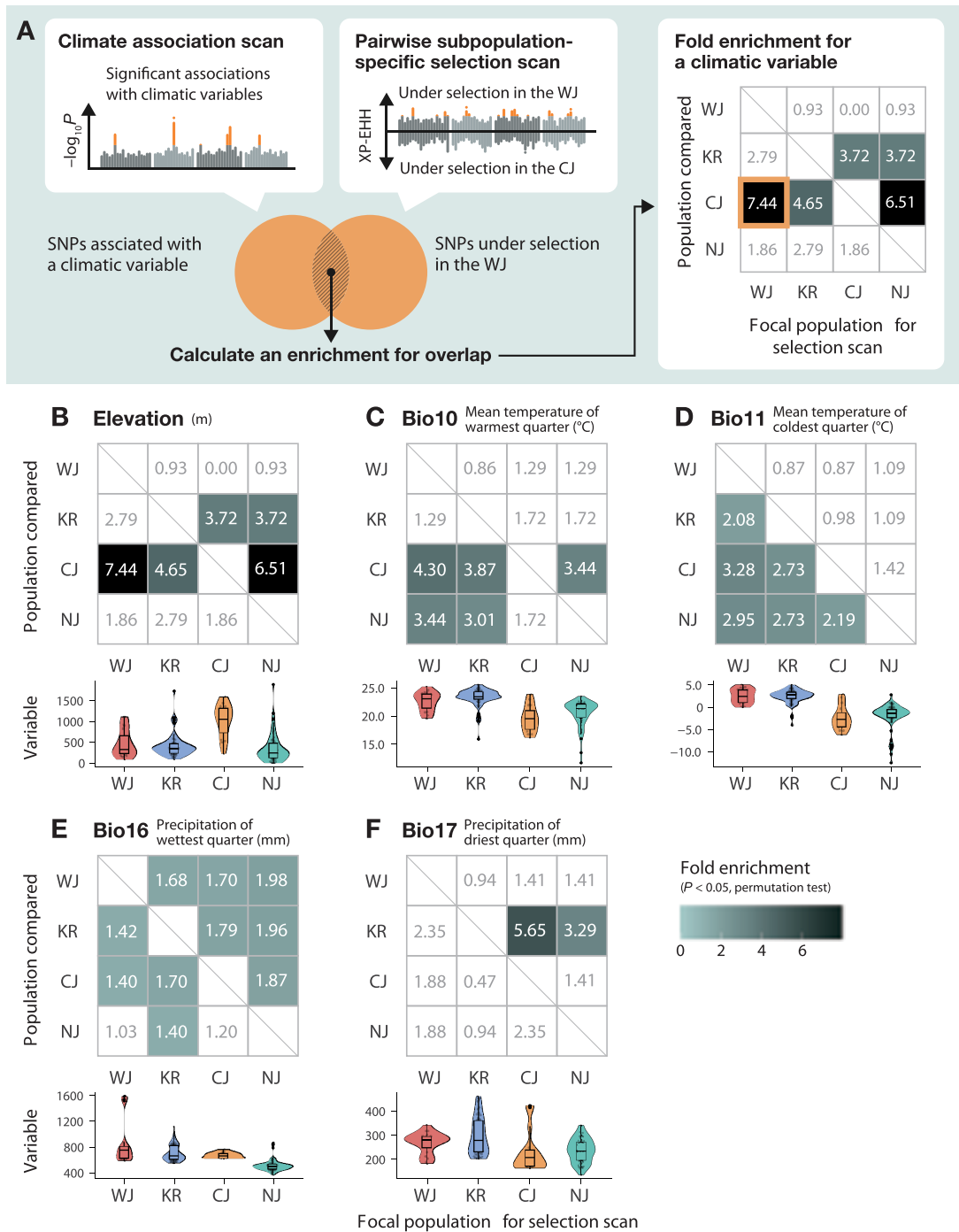


Fig. 3 Selection enrichment analysis for climate adaptation. (A) Scheme of the analysis. A genome-wide association mapping was performed for each bioclimatic variable to find associated SNPs using the R package LEA v.3.10.2 (Gain and François 2021). We also performed genome-wide selection scans based on the cross-population extended haplotype homozygosity (XP-EHH) for all pairwise combinations of subpopulations using the R package rehh v.3.2.2 (Gautier et al. 2017). Positive and negative values of XP-EHH indicate the direction of selection, providing information about which subpopulation selection may have acted. Then, fold enrichments of the overlap were calculated with their statistical significance based on permutation tests. High enrichment scores indicate that more climate-associated loci tend to overlap with the positively selected regions than expected by chance. Enrichments and distributions of variables between subpopulations for elevation (B), bio10 (mean temperature of warmest quarter [°C]) (C), bio11 (mean temperature of coldest quarter [°C]) (D), bio16 (precipitation of wettest quarter [mm]) (E), and bio17 (precipitation of driest quarter [mm]) (F). (B–F) Because XP-EHH scores are calculated for all pairwise combinations of subpopulations, enrichments are also obtained for each focal subpopulation for selection scan (horizontal axis) and each subpopulation in comparison (vertical axis).

individuals were classified into four Japanese subpopulations based on the maximum factor under $K = 4$: WJ, KR, CJ and NJ (**Supplementary Fig. S3B**). Taking the information of population structure into account, we performed GWAS for elevation, 19 bioclimatic variables, and seven principal components of bioclimatic variables (PC1–PC7) using latent factor mixed models (LFMM) (Caye et al. 2019, Gain and François 2021). These environmental variables were highly variable between the four subpopulations (**Fig. 3, Supplementary Fig. S4**). The number of peaks satisfying genome-wide significance [4-kb windows containing SNPs with $P < 0.05$ after Benjamini–Hochberg false-discovery rate (FDR) correction] ranged from 0 to 2,518 among the variables (**Supplementary Figs. S5, S6**).

To detect subpopulation-specific selection, we performed genome-wide selection scans based on the XP-EHH for all pairwise comparisons of the four subpopulations in Japan (Sabeti et al. 2007, Gautier and Vitalis 2012, Gautier et al. 2017, Klassmann and Gautier 2022). We first calculated the integrated extended haplotype homozygosity of a SNP site (iES), representing the average length of shared haplotypes (Tang et al. 2007). We then calculated cross-population extended haplotype homozygosity (XP-EHH) based on the cross-comparison of two populations to discover alleles with near-fixation in one population (Sabeti et al. 2007). XP-EHH values were calculated for each SNP with pairwise comparisons between the four subpopulations (**Supplementary Fig. S7**). Then, we performed enrichment analyses to test whether environmentally associated peaks are enriched in regions under subpopulation-specific selective sweeps. We found significant enrichments of 4-kb windows containing SNPs associated with elevation and 14 bioclimatic variables in the top or bottom 2.5% tails of the XP-EHH scores in at least one subpopulation comparison (**Fig. 3, Supplementary Figs. S8, S9**). We also calculated F_{ST} between the four subpopulations as an alternative criterion for population differentiation, and detected significant enrichments of windows containing SNPs associated with elevation and 15 bioclimatic variables in the top 5% of F_{ST} tails, providing robust support for adaptive differentiation between subpopulations on climate-associated loci (**Supplementary Figs. S10, S11, S12**). An advantage of using XP-EHH is that we can identify not only adaptive differentiation between subpopulations but also the direction of selection, i.e. in which subpopulations selection might have acted. We therefore scrutinized the enrichments of XP-EHH detailed further. Elevation-associated SNPs showed strong enrichments of the XP-EHH scores when focusing on WJ, KR, NJ subpopulations in comparison with the CJ subpopulation, which had the highest elevation level (**Fig. 3B**). However, we note that elevation showed strong correlations with some of the bioclimatic variables and shared many significant windows with them, suggesting that enrichment of the XP-EHH scores in elevation-associated SNPs may not necessarily reflect the selection for elevation *per se* but for other associated bioclimatic variables (**Supplementary Figs. S13, S14**). SNPs associated with bio10 (mean temperature of warmest quarter) were also enriched in the XP-EHH scores when focusing on WJ

and KR subpopulations in comparison with CJ and NJ subpopulations that had lower values of bio10 (**Fig. 3C**). Similarly, the association with bio11 (mean temperature of coldest quarter) showed significant enrichments in the XP-EHH scores when more southern subpopulations were compared with northern ones (**Fig. 3D**). In contrast, SNPs associated with bio16 (precipitation of wettest quarter) and bio17 (precipitation of driest quarter) tended to be enriched when focusing on more northern subpopulations compared with southern ones (**Fig. 3E, F**). These enrichment analyses demonstrate the prevailing selection signatures for climate-associated loci, particularly selection toward a warmer climate in southern subpopulations (WJ and KR) and a drier environment in northern subpopulations (CJ and NJ).

Here, we highlight a few genes showing associations with bioclimatic-associated SNPs and signatures of selection, particularly focusing on those that showed enrichments in more than two comparisons of subpopulations (**Fig. 4A, Supplementary Table S3**). Among the peaks associated with bio10 (mean temperature of the warmest quarter), two windows on chromosome 7 were commonly included in the 2.5% tail of XP-EHH scores in WJ, KR and NJ subpopulations when compared with CJ populations (**Fig. 4A**). The SNP with the strongest association was located ~1.2 kbp downstream of the stop codon of the gene Ah7G32100, annotated as a homolog of *A. thaliana* HEAT INDUCIBLE LIPASE-1, and showed a clear geographic cline in the allele distribution (**Fig. 4B, C**). We also found three peaks associated with bio17 (precipitation of driest quarter) and under selection in the CJ and NJ subpopulations in comparison to the KR subpopulation (**Fig. 4A**). The SNP with the strongest association on chromosome 8 was located ~1.9 kbp upstream from the start codon of Ah8G12760, which also showed a geographic cline in the allele distribution (**Fig. 4B, C**). This gene was annotated as a homolog of *A. thaliana* RAF10, which is involved in abscisic acid (ABA) response (Lee et al. 2015). In addition, we found a peak associated with bio11 (mean temperature of coldest quarter), which was under selection in the KR and CJ subpopulations compared with the NJ subpopulation. The most associated SNP was located ~2.8 kbp upstream from the start codon of Ah1G24310, a homolog of EMBRYO DEFECTIVE 1507 in *A. thaliana*. Its mutant was reported to show reduced expression levels of FLOWERING LOCUS C (FLC) and early flowering (Mahrez et al. 2016). These genes are promising candidates for climatic adaptation, particularly for warmer temperatures and lower precipitation.

Historical changes in the suitable distribution area

As selection enrichment analysis suggested climate-driven selection, we next investigated the post-glacial changes in the distribution of Japanese *A. halleri* with a framework of ecological niche modeling (also called species distribution modeling). Ecological niche modeling is a numerical tool that combines observations of species occurrence with environmental estimates intended to predict distributions across landscapes,

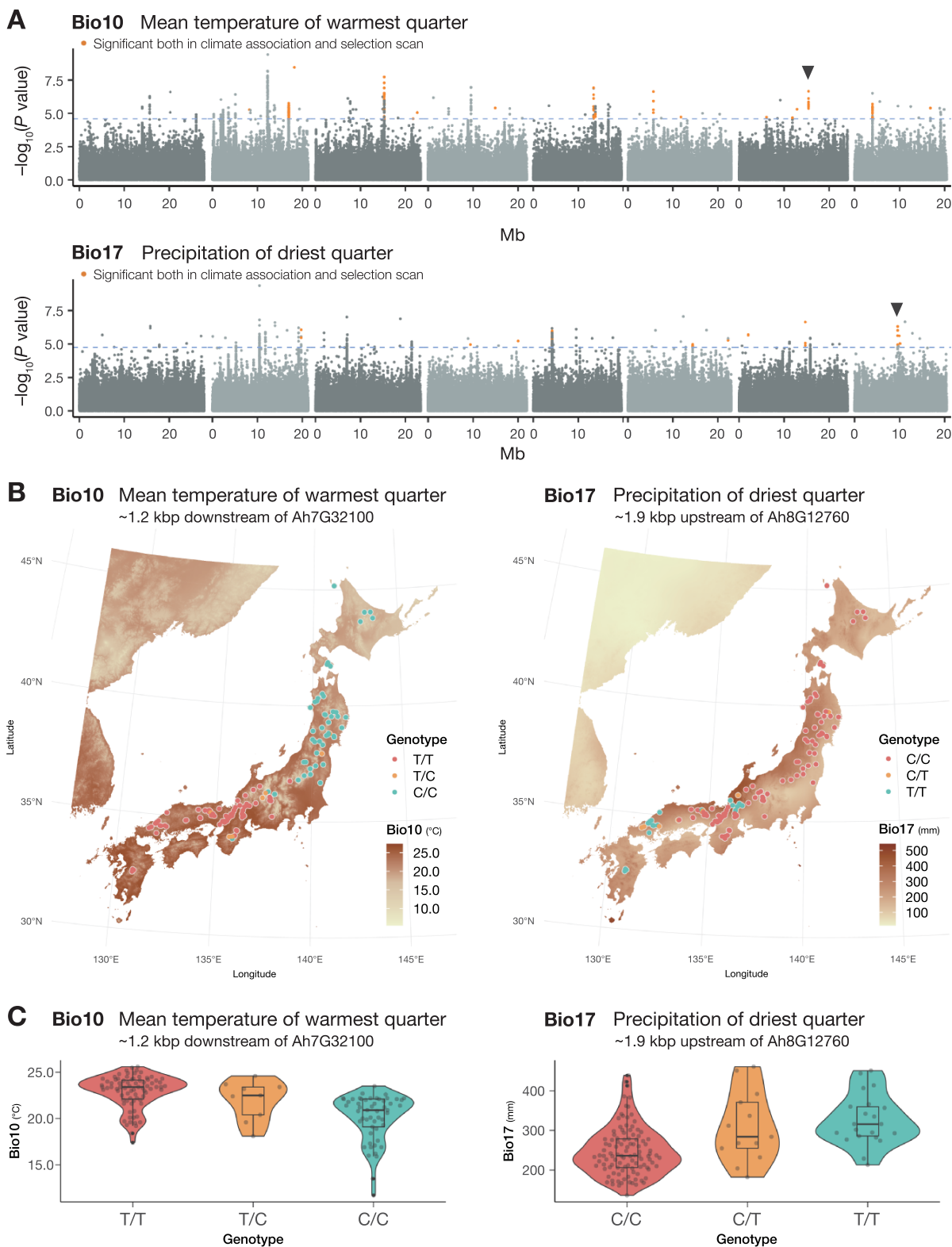


Fig. 4 Climate-associated loci and their geographic distributions. (A) Manhattan plots for bio10 (mean temperature of warmest quarter) and bio17 (precipitation of driest quarter). Horizontal dashed lines indicate $P = 0.05$ after Benjamini–Hochberg FDR correction. Highlighted dots indicate SNPs associated significantly with bioclimatic variables and included in the 4-kb windows of selection scans that were significant in more than two combinations of subpopulations. SNPs indicated by arrowheads were used for plots of (B) and (C). SNPs on the assembled chromosomes are shown. (B) Geographic distributions of the alleles strongly associated with bio10 and bio17. These alleles are located ~1.2 kbp downstream of the stop codon of Ah7G32100 and ~1.9 kbp upstream from the start codon of Ah8G12760, respectively. (C) Distribution of bioclimatic variables for each allele.

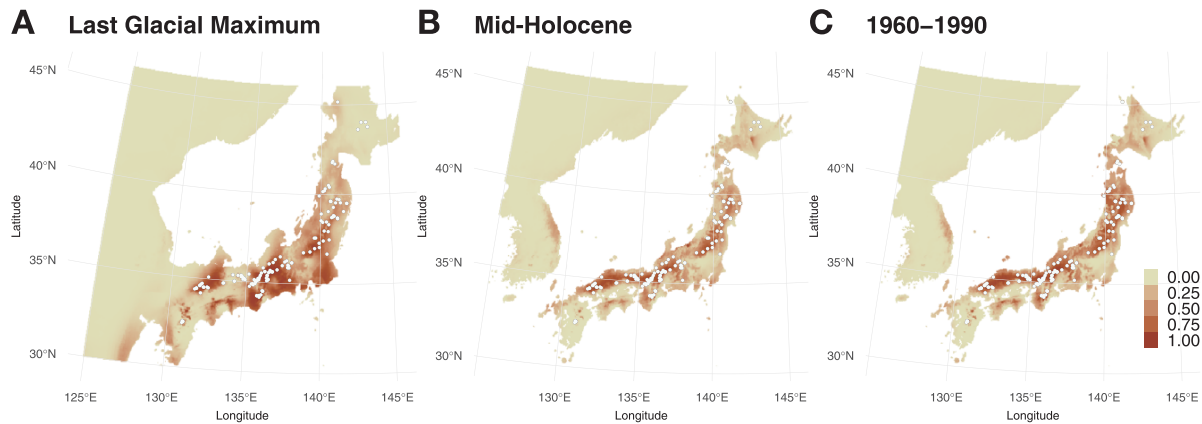


Fig. 5 Ecological niche modeling of the potential habitat of *A. halleri*. MaxEnt v.3.4.4 was used (Phillips et al. 2004, 2024). Sample locations are indicated by open circles. The occurrence probabilities under the LGM (about 22,000 years ago) (A), Mid-Holocene (about 6,000 years ago) (B), and the current (1960–1990) climate conditions (C) are shown.

including extrapolation in space and time (Elith and Leathwick 2009). Here, we first obtained distribution probabilities under the current climate conditions based on the sampling locations of Japanese *A. halleri* and six principal components of bioclimatic variables (see ‘Materials & Methods’ section for details). Then, we also developed distribution models under the Last Glacial Maximum (LGM; about 22,000 years ago) and Mid-Holocene (about 6,000 years ago). We found that potential habitats were predicted mainly in the lowlands of southern coastal regions in the LGM (Fig. 5A). In contrast, predictions based on Mid-Holocene and current climate conditions showed that potential habitats were shifted toward northern coastal and mountainous areas (Fig. 5B, C).

Discussion

Genomic signatures of climatic adaptation during glacial cycles

In this study, we investigated population structure, demographic history and signatures of climate-driven selection by exploiting genome-wide polymorphism data of 141 Japanese *A. halleri* individuals and European accessions as outgroups. The ABC and the PSMC analyses consistently suggested genetic differentiation among Japanese subpopulations since the LGP (Fig. 2), shaping the current pattern of population structure. Clonal reproduction and a lack of specific morphological structure for long-distance dispersal of seeds may have facilitated the formation and maintenance of population structure (Sato and Kudoh 2014, Honjo and Kudoh 2019). Population demographic analysis revealed population size fluctuations in the LGP, which was particularly prominent since the subpopulations started to diverge (~50,000 years ago). Together with the prediction by the ecological niche modeling that the potential habitat may have shifted from southern coastal regions to northern coastal and mountainous areas (Fig. 5), our analyses

overall suggest that subpopulations in Japanese *A. halleri* have experienced population subdivisions, fluctuations and distribution range shifts accompanied by the climatic change during and after the LGP, as also reported in other plants and animals (e.g. Comes and Kadereit 1998, Hewitt 2004, Qiu et al. 2011).

The combination of GWAS for bioclimatic variables and XP-EHH-based selection scans enabled us to detect signatures of subpopulation-specific selection with explicit climatic contexts (Fig. 3, Supplementary Figs. S8, S9). An advantage of using XP-EHH is that we can identify not only the adaptive divergence between subpopulations but also the direction of selection, i.e. in which subpopulations selection might have acted. For example, temperature-associated GWAS peaks were enriched in the XP-EHH tail when focusing on southern subpopulations (WJ and KR), suggesting adaptation toward higher temperatures in those subpopulations (Fig. 3C, D). Similarly, the enrichment of GWAS peaks associated with precipitation of the driest quarter was observed in the XP-EHH scores when focusing on Central Japan (CJ) and Northern Japan (NJ) populations, suggesting adaptation toward lower precipitations (Fig. 3F). Although our study aimed at detecting polygenic adaptation for local climates rather than screening genes, it is worth mentioning that several promising genes appeared to be linked to the associated SNPs, such as homologs of *HEAT INDUCIBLE LIPASE-1* and *RAF10*, involved in ABA response, associated with the mean temperature of the warmest quarter and precipitation of the driest quarter, respectively. Given the subpopulation divergence since the LGP, such subpopulation-specific climatic adaptation for higher temperature or lower precipitation would have occurred during or after the LGP, possibly accompanied by population contraction and expansions. In *A. thaliana*, which experienced an expansion of its distribution after the LGP, a study using more than 1,000 natural strains also identified SNPs associated with precipitation-related bioclimatic variables, and they showed a significant geographic gradient (The 1001 Genomes Consortium 2016, Lee et al. 2017). Except for those pioneering studies,

attempts to identify loci potentially involved in adaptation are still scarce. Our study highlights the importance of integrating climate associations, selection scans and population demographic analyses to identify genomic signatures of population-specific adaptation when comparing several genetic clusters. Such combined approaches would also be important for reducing false positives of environmental association analysis that are suggested to be high (Rellstab et al. 2015).

Population structure and demography history of *A. halleri*

While there have been a few studies on population structure in Japanese *A. halleri*, they either relied on a limited number of loci and sampling locations (Sato and Kudoh 2014) or focused on a microgeographic scale (Kubota et al. 2015, Yoshida et al. 2023). In this study, by using full-genome resequencing data from region-wide samples, we revealed a population structure that showed a clear geographic pattern (Fig. 1). Based on the data of population structure, European *A. halleri* were inferred to have been distributed in multiple refugia in the LGP (Šrámková-Fuxová et al. 2017, Šrámková et al. 2019). Our population demography analysis suggests population subdivision during the LGP, consistent with the persistence of multiple refugia. Our ecological niche modeling also identified multiple potential habitats along the southern coastal regions during the LGM, which might have served as glacial refugia. Species occurring in similar geographic regions sometimes share similar patterns of population structure, which may reflect shared history during glacial cycles. The pattern of geographical differentiation of Japanese *A. halleri* appeared to be at least partially comparable with those of other plants widely distributed in Japan. The genetic boundary between northern and southern clusters similar to that of CJ and NJ subpopulations was previously identified in several plant species in Japan (Ohsawa and Ide 2011, Iwasaki et al. 2012, Magota et al. 2021). The phylogeographic analysis of the goldenrod *Solidago virgaurea* complex identified similar genetic clusters across the Japanese archipelago (Sakaguchi et al. 2018). These shared population structures imply similarity of the distribution shifts since the LGP and survival in the same refugia (Iwasaki et al. 2012). The smallest N_e of the ancestral population of all Japanese subpopulations in *A. halleri* suggests that a demographic bottleneck occurred upon colonization into the Japanese archipelago, consistent with migration to the archipelago through land bridges, as was suggested in some species (Ohsawa and Ide 2011, Sakaguchi et al. 2018). It should be noted that our demographic analyses considered relatively simple bifurcating scenarios because our main purpose was to obtain the approximate timescale of divergence time between subpopulations; more complex scenarios involving admixture between subpopulations or introgressions from related species could be considered, which may better explain the whole pattern of population structure of *A. halleri*.

Conclusion and future perspectives

In this study, the integrated analysis of population demography, climate associations, selection scans, and ecological niche modeling revealed the signatures of subpopulation-specific climatic adaptation, particularly for higher temperature or lower precipitation, which would have taken place during or after LGP, possibly accompanied by population contraction and expansions in Japanese *A. halleri*. Knowledge of the adaptation to past and current environments should help us predict the evolutionary responses to future climate changes. Recent population genomic studies focusing on drought tolerance in *A. thaliana* predicted its distribution under future climate models based on the geographic information of adaptive alleles and field experiments (Exposito-Alonso et al. 2017, 2019). While there is a growing interest in using population genomics to predict future evolutionary responses, the importance of validating genomic predictions has been suggested (Capblancq et al. 2020). The functional relevance of several genes that appeared in our climate GWAS needs to be further validated in *A. halleri*, and common garden experiments in multiple locations in Japan would be helpful for such validations. We would also like to note that not only adaptation to temperature or precipitation, but the evolution of mating systems and polyploidy are also often suggested to be associated with population demographic changes during LGP and post-glacial expansions (Tsuchimatsu et al. 2010, 2012, Durvasula et al. 2017, Novikova et al. 2017, Willi et al. 2022). Investigation of further phenotypic variation and its genetic basis will be valuable for a better understanding of the mechanisms of adaptation to a changing climate.

Materials and Methods

Sampling and whole-genome resequencing

In this study, we used whole-genome resequencing data of 141 Japanese and 16 European individuals (Supplementary Table S1). The 141 locations of Japanese samples largely cover the distribution range of *A. halleri* in Japan, according to the Global Biodiversity Information Facility (GBIF Backbone Taxonomy 2023). Among these individuals, eight were sequenced by Kubota et al. (2015). For the other individuals, whole-genome resequencing was performed by the following methods, depending on samples.

For 90 Japanese individuals, we sampled about 30 mg dried biomass of leaves per individual from herbarium specimens or from samples collected in each location, and genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) (Supplementary Table S1). DNA quality, quantity and integrity were assessed with a Nanodrop spectrophotometer (ThermoFisher Scientific, Darmstadt, Germany). After fragmentation by the NEBNext dsDNA Fragmentase kit (Ipswich, MA, USA), libraries were prepared using the TruSeq DNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, and whole-genome sequencing with 100-bp single-end reads were performed using an Illumina HiSeq 2000 instrument (Illumina, San Diego, CA, USA). For herbarium samples, the fragmentation procedure was not performed.

For another set of 43 Japanese individuals, we sampled about 30 mg dried biomass of leaves per individual, and genomic DNA was extracted using a modified cetyltrimethylammonium bromide method (Milligan 1992). DNA

quality, quantity and integrity were assessed with a Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and a Nanodrop spectrophotometer (ThermoFisher Scientific, Darmstadt, Germany). Library preparations and 100-bp paired-end sequencing was performed using a BGISEQ-500 next-generation sequencing platform (BGI, Shenzhen, China).

To obtain whole-genome resequencing data of 16 European individuals, living *A. halleri* plants collected in the field (Stein et al. 2017) were maintained in a greenhouse and propagated vegetatively. We sampled 50–70 mg fresh biomass of young leaves per individual in a 1.5 ml polypropylene tube and shock-froze the tissues in liquid nitrogen. Leaf tissues were homogenized using a single metal bead (5 mm stainless steel ball) per tube in a Retsch mixer mill (Type MM 300, Retsch, Haan, Germany) for 1 min at 30 Hz, using adapters pre-cooled to -80°C . The homogenate was used to isolate DNA using the NucleoMag[®] Plant kit (Macherey Nagel, Düren, Germany) on an epMotion 5075 robot according to the manufacturers' instructions (Eppendorf, Hamburg, Germany). DNA quality, quantity and integrity were assessed on 0.8% (w/v) TAE-agarose gels, with a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, Darmstadt, Germany) and a Qubit 2.0 Fluorometer (Invitrogen, Darmstadt, Germany) using the dsDNA BR Assay Kit. Per sample, 1 μg (50 μl) genomic DNA was used as an input for producing libraries using the Illumina TruSeq DNA PCR-Free Library Prep according to the manufacturer's instructions. Libraries passing QC were diluted, pooled and dispatched to Novogene Europe (Cambridge, UK) for whole-genome sequencing with 150-bp paired-end reads at a targeted minimum depth of 20 \times genome coverage using an Illumina NovaSeq 6,000 instrument (Illumina, San Diego, CA, USA).

SNP identification

Whole-genome short reads were trimmed using Trimmomatic 0.39 with the options of LEADING:5, TRAILING:5, SLIDINGWINDOW:4:15 and MINLEN for 30% of the read length (Bolger et al. 2014). To be consistent throughout the analysis, pair-end reads were concatenated and treated as single-end reads. We then mapped read files with BWA-MEM 0.7.17-r1188 (Li 2013) to *A. halleri* v2.03 assembly (DOE-JGI, <http://phytozome.jgi.doe.gov/>), then sorted and indexed the bam files with samtools v.1.17 (Danecek et al. 2021). The average depths were calculated for each bam file using the samtools depth command. We used bcftools v.1.17 mpileup and call pipeline for variant calling (Danecek et al. 2021). SNPs with quality ≤ 20 , minor allele frequency ≤ 0.05 , total depth $\leq 1,500$ or total depth $\geq 4,300$, and a fraction of missing individuals > 0.1 were filtered out using bcftools. We also conducted SNP identifications without European individuals using the same pipeline and filtering options except for depth (SNPs with total depth $\leq 1,200$ or total depth $\geq 3,600$ were filtered out). This set was used for population structure analysis specific to Japanese individuals, climate association mapping and selection scans. We have identified 1,013,450 SNPs in this Japanese dataset, and additional filtering for tri-allelic SNPs were conducted in each of the further analysis.

Population structure and network analysis

Population structure analysis was conducted using non-negative matrix factorization (sNMF) using the snmf function ($K = 1-12$; 50 repetitions for each K) implemented in the R package LEA v.3.10.2 (Frichot et al. 2014, Frichot and François 2015). We employed the cross-entropy criterion to determine the most supported number of K . A population structure analysis without European individuals was also performed ($K = 1-12$; 10 repetitions for each K), and the individuals were classified into four Japanese subpopulations according to the maximum factor of $K = 4$. Selection scans in Japanese subpopulations were based on this classification.

To analyze the genetic relationship between individuals, we generated a nexus file from the vcf file using vcf2phylip (Ortiz 2019), which was used as an input for SplitsTree4 v.4.19.2 (Huson and Bryant 2006).

Inference of demographic history and population subdivisions

We employed the ABC analysis using the DIYABC-RF v.1.2.1 software (Collin et al. 2021). For the input for ABC analysis, 20,000 SNP sites with no missing individuals were randomly extracted from the filtered SNPs and converted using the vcf2diyabc.py script (<https://github.com/loire/vcf2DIYABC.py>). Based on population structure and network analysis, we assumed an evolutionary scenario in which all the subpopulations diverged simultaneously and three other bifurcating scenarios (Fig. 2; Supplementary Fig. S2). In the latter three scenarios, KR and CJ subpopulations were assumed to form a single clade, as they were suggested to be closely related in the population structure analysis. The recent split between KR and CJ subpopulations are also supported from the PSMC analysis, in which those subpopulations showed similar population size dynamics until about 20,000 years ago, whereas the WJ and NJ subpopulations showed distinct dynamics since early times. Because our primary focus was to obtain the approximate timescale of divergence time between subpopulations, we examined these relatively simple scenarios without migration and hybridization. The scenarios considered were consistent with the patterns of hierarchical clustering of population structure (Fig. 1B). The prior distributions were set to 10–200,000 for all the population size parameters, 20,000–500,000 for the split time of the EU population, and 10–50,000 for other timings. After running 80,600 simulations, the best scenario was chosen based on the number of classification vote, and parameters were estimated for each scenario with 1,000 out-of-bag testing samples and 1,000 trees for random forest analysis.

Historical effective population sizes of four Japanese subpopulations were inferred using the PSMC model (Li and Durbin 2011), implemented in psmc v.0.6.5-r67. We first selected representative individuals with the highest portion of the subpopulation-specific elements in the clustering results of sNMF with an average depth > 28 , calculated by the SNP identification pipeline. We mapped reads of these individuals to hardmasked *Arabidopsis halleri* v2.03 assembly (DOE-JGI, <http://phytozome.jgi.doe.gov/>) using BWA-MEM and identified SNPs using bcftools mpileup and call pipeline. With the average depth calculated from these bam files using the samtools depth command, we converted them into a fastq file using vcfutils.pl with the option -d set to one-third and -D set to twice the average depth, which was then converted using fq2psmcfa with the -q20 option. We conducted the psmc analysis with 100 bootstrap runs using the options of -N25, -t15, -r5 and -p '4 + 25*2 + 4 + 6'. The results were plotted assuming a mutation rate of 7.1×10^{-9} per site per generation (Ossowski et al. 2010) and a generation time of 2 years.

Genome-wide scans for climate associations

We conducted a genome-wide association mapping to identify loci associated with environmental gradients. We used elevation and bioclimatic variables of the places of origin based on the WorldClim v1.4 historical climate data (Hijmans et al. 2005) obtained using the raster package (Hijmans 2024) in R. Details about bioclimatic variables are available in Supplementary Table S4. These variables were at the 30 arc-second resolution data. As these variables were correlated with each other, we performed principal component analysis (PCA) for 19 bioclimatic variables of the individuals using the prcomp function in R. We used seven principal components (PC1–PC7) for further analysis, which cumulatively explained 99.27% of the variance (Supplementary Fig. S15). We then filtered out tri-allelic SNPs and imputed missing genotypes using the impute function in the R package LEA (Gain and François 2021) based on the result of population structure analysis within Japan. Genome-wide association mapping was performed using LFMM implemented in the R package LEA lfmm2 function (Caye et al. 2019, Gain and François 2021), specifying the number latent factor $K = 7$. We then identified 4-kb windows containing SNPs with $P < 0.05$ using Benjamini–Hochberg FDR correction. This window size was based on the genome-wide pattern of linkage disequilibrium decay within 4 kb, estimated by PopLDdecay (Zhang et al. 2019) (Supplementary Fig. S16).

Genome-wide selection scan

We performed genome-wide selection scans based on the cross-population extended haplotype homozygosity (XP-EHH) to detect subpopulation-specific selection. To calculate XP-EHH, we used the R package *rehh* v.3.2.2 (Sabeti et al. 2007, Gautier and Vitalis 2012, Gautier et al. 2017, Klassmann and Gautier 2022). Positive and negative values of XP-EHH reflect the direction of selection based on the difference in haplotype homozygosity between subpopulations. We first calculated iES statistics for bi-allelic SNPs on chromosome 1–8 using the *scan_hh* function in the pairwise combination of four Japanese subpopulations. We then calculated XP-EHH using the *ies2xpehh* function and extracted maximum and minimum scores in the 4-kb windows. We also calculated weighted F_{ST} in the 4-kb window using VCFtools v. 0.1.16 —weir-fst-pop option (Danecek et al. 2011).

We then calculated the enrichment of environmentally associated windows in the top or bottom extreme tail of the XP-EHH and F_{ST} values. The threshold for the XP-EHH tail was set to 2.5% for both sides. Windows containing climatic-associated sites with $P < 0.05$ (FDR correction) were considered. For the F_{ST} values, we set the threshold to the top 5%.

We conducted permutation tests to determine the statistical significance of fold enrichment. We used the so-called ‘genome rotation’ scheme which randomly rotates one of the two genome-wide window sets. For each of the permutations, we calculated fold enrichment between climate-associated window sets and randomly shifted XP-EHH window sets, preserving the relative position of the windows to take linkage disequilibrium between loci into account (Nordborg et al. 2005, Atwell et al. 2010, Horton et al. 2012, Tsuchimatsu et al. 2020, Sasaki et al. 2022). We conducted 1,000 times of permutation for each combination of bioclimatic variables and populations.

We also analyzed genes overlapping with the significant windows in climatic associations and selection scans. We especially focused on the significant windows in more than two pairwise combinations of subpopulations. The gene annotation was based on *Arabidopsis halleri* v.2.1.0 annotation data (DOE-JGI, <http://phytozome.jgi.doe.gov/>), and information on genes linked to each SNP was obtained by SnpEff v.5.1d (Cingolani et al. 2012). We plotted the geographic and bioclimatic distribution of imputed alleles on the loci of interest using the *vcfR* package (Knaus and Grünwald 2017), with bioclimatic variables converted using the raster package (Hijmans 2024) in R.

Ecological niche modeling

We conducted ecological niche modeling using MaxEnt v.3.4.4 (Phillips et al. 2004, 2024) based on the 2.5 arc-minute resolution bioclimatic variables obtained from WorldClim v1.4 (Hijmans et al. 2005). These variables were cropped into the geographical range of 125–146°E, 30–46°N using a raster package (Hijmans 2024) in R. As the bioclimatic variables are correlated with each other, we first conducted PCA for historical bioclimatic data using *prcomp* function in R to prevent overfitting. We then used PC1–PC6, which cumulatively explained 99.08% of variance as environmental layers (Supplementary Fig. S17). These components were projected to paleo climate variables at Mid-Holocene (about 6,000 years ago) and LGM (about 22,000 years ago) on WorldClim v1.4, which were then used as projection layers. Analysis was performed based on sample locations with default parameters.

Supplementary Data

Supplementary data are available at PCP online.

Data Availability

The sequence data underlying this article are available in European Nucleotide Archive (accession number: PRJEB72840). The source code and SNP data used in this study are available on Dryad (<https://doi.org/10.5061/dryad.1jwstqk3s>).

Funding

JSPS KAKENHI (grant numbers: 22K21352 and 23H02537 to T.T., 16K18623 and 19K06835 to S.K.), Environment Research and Technology Development Fund (S9) and JST CREST (JPMJCR11B3) to S.I.M., and European Union ERC-AdG LEAP-EXTREME (788380) to U.K.

Acknowledgments

The authors thank Yu Okamura for helpful discussions, Naoko Ishikawa, Takaya Iwasaki, Masayuki Maki, Shuichi Nemoto, Shota Sakaguchi, Mahoro Suzuki, Akitomo Uchida, Yoichi Watanabe and Yoshiharu Yamamoto for collecting samples, Atsushi Ebihara, Noriaki Murakami, Jin Murata, Hidetoshi Nagemasu, Hiroyoshi Ohashi, Takashi Shiga and Hideki Takahashi for providing herbarium specimens and Motomi Ito, Yutaka Suzuki, and Sumio Sugano for supporting genome sequencing.

Authors' Contributions

R.A.S., S.K., S.-I.M., and T.T. conceived and designed the study. S.K., V.K., U.K. and S.-I.M. generated the sequence data. R.A.S. analyzed the data with help from V.C. and T.T. R.A.S. and T.T. wrote the paper with inputs from all authors.

Disclosures

The authors have no conflicts of interest to declare.

References

- Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Bamba, M., Kawaguchi, Y.W. and Tsuchimatsu, T. (2019) Plant adaptation and speciation studied by population genomic approaches. *Dev. Growth Differ* 61: 12–24.
- Barghi, N., Hermisson, J. and Schlötterer, C. (2020) Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.* 21: 769–781.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bourgeois, Y.X.C. and Warren, B.H. (2021) An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Mol. Ecol.* 30: 6036–6071.
- Capblancq, T., Fitzpatrick, M.C., Bay, R.A., Exposito-Alonso, M. and Keller, S.R. (2020) Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annu. Rev. Ecol. Evol. Syst.* 51: 245–269.
- Castric, V. and Vekemans, X. (2004) INVITED REVIEW: plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol. Ecol.* 13: 2873–2889.
- Caye, K., Jumentier, B., Lepeule, J. and François, O. (2019) LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* 36: 852–860.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6: 80–92.

- Collin, F., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., et al. (2021) Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Mol. Ecol. Resour.* 21: 2598–2613.
- Comes, H.P. and Kadereit, J.W. (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends Plant Sci.* 3: 432–438.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008.
- Davis, M.B. and Shaw, R.G. (2001) Range shifts and adaptive responses to quaternary climate change. *Science* 292: 673–679.
- De Lafontaine, G., Napier, J.D., Petit, R.J. and Hu, F.S. (2018) Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology* 99: 1530–1546.
- DOE-JGI. *Arabidopsis halleri* v2.03. 2022. <http://phytozome.jgi.doe.gov/> (February 22, 2023, date last accessed).
- Durand, E., Chantreau, M., Le Veve, A., Stetsenko, R., Dubin, M., Genete, M., et al. (2020) Evolution of self-incompatibility in the Brassicaceae: lessons from a textbook example of natural selection. *Evol. Appl.* 13: 1279–1297.
- Durvasula, A., Fulgione, A., Gutaker, R.M., Alacakaptan, S.I., Flood, P.J., Neto, C., et al. (2017) African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 114: 5213–5218.
- Elith, J. and Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- EPICA community members. (2004) Eight glacial cycles from an Antarctic ice core. *Nature* 429: 623–628.
- Excoffier, L., Foll, M. and Petit, R.J. (2009) Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* 40: 481–501.
- Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H.A., Bosse, O., Nielsen, R. and Weigel, D. (2019) Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* 573: 126–129.
- Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H.A. and Weigel, D. (2017) Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat. Ecol. Evol.* 2: 352–358.
- Frichot, E. and François, O. (2015) LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* 6: 925–929.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. and François, O. (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196: 973–983.
- Gain, C. and François, O. (2021) LEA 3: factor models in population genetics and ecological genomics with R. *Mol. Ecol. Resour.* 21: 2738–2748.
- Gautier, M., Klassmann, A. and Vitalis, R. (2017) REHH 2.0: a reimplementation of the R package REHH to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17: 78–90.
- Gautier, M. and Vitalis, R. (2012) *rehh*: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28: 1176–1177.
- Gavin, D.G., Fitzpatrick, M.C., Gugger, P.F., Heath, K.D., Rodríguez-Sánchez, F., Dobrowski, S.Z., et al. (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytol.* 204: 37–54.
- GBIF Backbone Taxonomy. (2023) *Arabidopsis halleri* subsp. *gemmaifera* (Matsum.) O’Kane & Al-Shehbaz in GBIF Secretariat.
- Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., et al. (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.
- Healy, T.M., Brennan, R.S., Whitehead, A. and Schulte, P.M. (2018) Tolerance traits related to climate change resilience are independent and polygenic. *Glob. Change Biol.* 24: 5348–5360.
- Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the quaternary. *Philos. Trans. R Soc. London, Ser. B* 359: 183–195.
- Hijmans, R.J. (2024) Raster: geographic data analysis and modeling.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Honjo, M.N. and Kudoh, H. (2019) *Arabidopsis halleri*: a perennial model system for studying population differentiation and local adaptation. *AoB PLANTS* 11: 1z076.
- Horton, M.W., Hancock, A.M., Huang, Y.S., Toomajian, C., Atwell, S., Auton, A., et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Genet.* 44: 212–216.
- Huson, D.H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254–267.
- Iwasaki, T., Aoki, K., Seo, A. and Murakami, N. (2012) Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation. *J. Plant Res.* 125: 207–221.
- Keller, S.R., Soolanayakanahally, R.Y., Guy, R.D., Silim, S.N., Olson, M.S. and Tiffin, P. (2011) Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). *Am. J. Bot.* 98: 99–108.
- Klassmann, A. and Gautier, M. (2022) Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. *PLoS ONE* 17: e0262024.
- Knaus, B.J. and Grünwald, N.J. (2017) VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17: 44–53.
- Koch, M.A. (2018) The plant model system *Arabidopsis* set into an evolutionary, systematic and spatio-temporal context. *J. Exp. Bot.* 70: 55–67.
- Krämer, U. (2010) Metal hyperaccumulation in plants. *Annu. Rev. Plant Biol.* 61: 517–534.
- Kubota, S., Iwasaki, T., Hanada, K., Nagano, A.J., Fujiyama, A., Toyoda, A., et al. (2015) A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLoS Genet.* 11: e1005361.
- Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., et al. (2017) On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat. Commun.* 8: 14458.
- Lee, S., Lee, M.H., Kim, J.-I. and Kim, S.Y. (2015) *Arabidopsis* putative MAP kinase kinases Raf10 and Raf11 are positive regulators of seed dormancy and ABA response. *Plant Cell Physiol.* 56: 84–97.
- Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Li, H. and Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475: 493–496.
- Luqman, H., Wegmann, D., Fior, S. and Widmer, A. (2023) Climate-induced range shifts drive adaptive response via spatio-temporal sieving of alleles. *Nat. Commun.* 14: 1080.
- Ma, L.-J., Cao, L.-J., Chen, J.-C., Tang, M.-Q., Song, W., Yang, F.-Y., et al. (2024) Rapid and repeated climate adaptation involving chromosome inversions following invasion of an insect. *Mol. Biol. Evol.* 41: msae044.
- Magota, K., Sakaguchi, S., Lee, J.-S., Yamamoto, M., Takahashi, D., Nagano, A.J., et al. (2021) Phylogeographic analysis of *Saxifraga fortunei* complex (Saxifragaceae) reveals multiple origins of morphological and ecological variations in the Japanese Archipelago. *Mol. Phylogenet. Evol.* 163: 107230.
- Mahrez, W., Shin, J., Muñoz-Viana, R., Figueiredo, D.D., Trejo-Arellano, M.S., Exner, V., et al. (2016) BRR2a affects flowering time via FLC splicing. *PLoS Genet.* 12: e1005924.

- Milligan, B.G. (1992) Plant DNA Isolation. In *Molecular Genetic Analysis of Populations: A Practical Approach*. Edited by Hoelzel, A.R. pp. 59–88. IRL Press, Oxford.
- Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., et al. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3: e196.
- Novikova, P.Y., Tsuchimatsu, T., Simon, S., Nizhynska, V., Voronin, V., Burns, R., et al. (2017) Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* msw299.
- Ohsawa, T. and Ide, Y. (2011) Phylogeographic patterns of highland and lowland plant species in Japan. *Alp Bot.* 121: 49–61.
- Ortiz, E.M. (2019) vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.
- Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327: 92–94.
- Phillips, S.J., Dudík, M. and Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. In *Twenty-First International Conference on Machine Learning - ICML'04*. Presented at the Twenty-first International Conference, Banff, Alberta, Canada, p. 83. ACM Press.
- Phillips, S.J., Dudík, M. and Schapire, R.E. (2024) Maxent software for modeling species niches and distributions.
- Provan, J. and Bennett, K. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol. Evol.* 23: 564–571.
- Qiu, Y.-X., Fu, C.-X. and Comes, H.P. (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol. Phylogenet. Evol.* 59: 225–244.
- Relstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. and Holderegger, R. (2015) A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24: 4348–4370.
- Roux, C., Castric, V., Pauwels, M., Wright, S.I., Saumitou-Laprade, P. and Vekemans, X. (2011) Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One* 6: e26872.
- Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449: 913–918.
- Sakaguchi, S., Kimura, T., Kyan, R., Maki, M., Nishino, T., Ishikawa, N., et al. (2018) Phylogeographic analysis of the East Asian goldenrod (*Solidago virgaurea* complex, Asteraceae) reveals hidden ecological diversification with recurrent formation of ecotypes. *Ann. Bot.* 121: 489–500.
- Sasaki, E., Gunis, J., Reichardt-Gomez, I., Nizhynska, V. and Nordborg, M. (2022) Conditional GWAS of non-CG transposon methylation in *Arabidopsis thaliana* reveals major polymorphisms in five genes. *PLoS Genet.* 18: e1010345.
- Sato, Y. and Kudoh, H. (2014) Fine-scale genetic differentiation of a temperate herb: relevance of local environments and demographic change. *Arab. J. Bot.* 6.
- Savolainen, O., Lascoux, M. and Merilä, J. (2013) Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14: 807–820.
- Schmitt, T. (2007) Molecular biogeography of Europe: pleistocene cycles and postglacial trends. *Front. Zool.* 4: 11.
- Šrámková, G., Kolář, F., Závěská, E., Lučanová, M., Španiel, S., Kolník, M., et al. (2019) Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Syst. Evol.* 305: 885–898.
- Šrámková-Fuxová, G., Závěská, E., Kolář, F., Lučanová, M., Španiel, S. and Marhold, K. (2017) Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. *Bot. J. Linn. Soc.* 185: 321–342.
- Stein, R.J., Höreth, S., De Melo, J.R.F., Syllwasschy, L., Lee, G., Garbin, M.L., et al. (2017) Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model *Arabidopsis halleri*. *New Phytol.* 213: 1274–1286.
- Tang, K., Thornton, K.R. and Stoneking, M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5: e171.
- The 1001 Genomes Consortium. (2016) 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166: 481–491.
- Tsuchimatsu, T., Kaiser, P., Yew, C.-L., Bachelier, J.B. and Shimizu, K.K. (2012) Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. *PLoS Genet.* 8: e1002838.
- Tsuchimatsu, T., Kakui, H., Yamazaki, M., Marona, C., Tsutsui, H., Hedhly, A., et al. (2020) Adaptive reduction of male gamete number in the selfing plant *Arabidopsis thaliana*. *Nat. Commun.* 11: 2885.
- Tsuchimatsu, T., Suwabe, K., Shimizu-Inatsugi, R., Isokawa, S., Pavlidis, P., Städler, T., et al. (2010) Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* 464: 1342–1346.
- Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Relstab, C., Kofler, R., et al. (2020) Evolutionary genomics can improve prediction of species' responses to climate change. *Evol. Lett.* 4: 4–18.
- Weigel, D. and Nordborg, M. (2015) Population genomics for understanding adaptation in wild plant species. *Annu. Rev. Genet.* 49: 315–338.
- Willi, Y., Lucek, K., Bachmann, O. and Walden, N. (2022) Recent speciation associated with range expansion and a shift to self-fertilization in North American *Arabidopsis*. *Nat. Commun.* 13: 7564.
- Yang, F., Crossley, M.S., Schrader, L., Dubovskiy, I.M., Wei, S. and Zhang, R. (2022) Polygenic adaptation contributes to the invasive success of the Colorado potato beetle. *Mol. Ecol.* 31: 5568–5580.
- Yoshida, N., Morinaga, S.-I., Wakamiya, T., Ishii, Y., Kubota, S. and Hikosaka, K. (2023) Does selection occur at the intermediate zone of two insufficiently isolated populations? A whole-genome analysis along an altitudinal gradient. *J. Plant Res.* 136: 183–199.
- Yumoto, G., Sasaki-Sekimoto, Y., Aryal, B., Ohta, H. and Kudoh, H. (2021) Altitudinal differentiation in the leaf wax-mediated flowering bud protection against frost in a perennial *Arabidopsis*. *Oecologia* 195: 677–687.
- Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. and Yang, T.-L. (2019) PopLD-decay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35: 1786–1788.