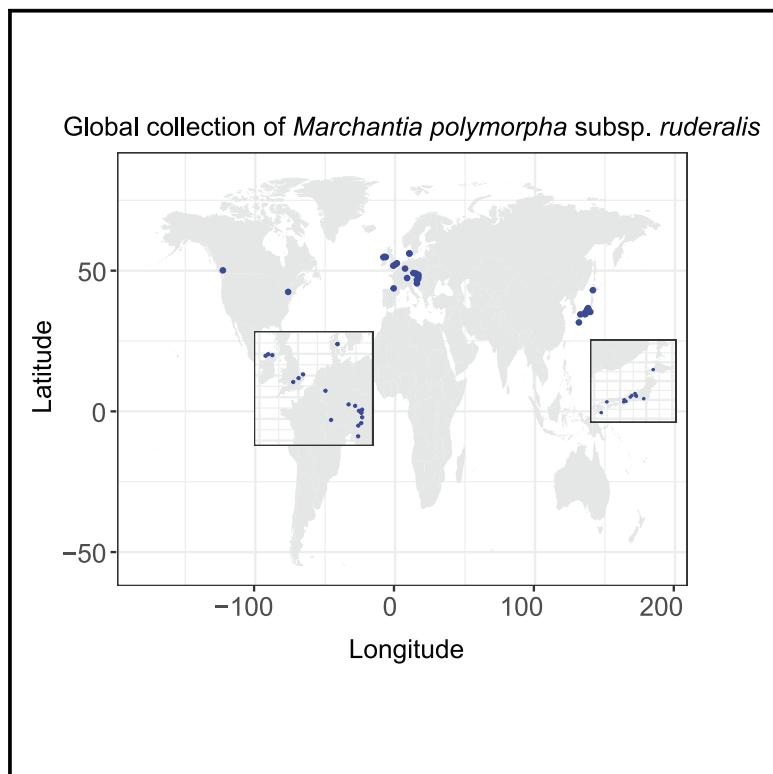


Population genomics of *Marchantia polymorpha* subsp. *ruderaleis* reveals evidence of climate adaptation

Graphical abstract



Authors

Shuangyang Wu, Katharina Jandrasits,
Kelly Swarts, ..., Tetsuya Hisanaga,
Frédéric Berger, Liam Dolan

Correspondence

liam.dolan@gmi.oeaw.ac.at

In brief

Wu et al. report a foundational analysis for population genetics in the bryophyte *Marchantia polymorpha* subsp. *ruderaleis*. They discovered that while Japanese and European populations are distinct, there is no population structure among European accessions. There is also evidence of genetic variation underpinning adaptation to different climates.

Highlights

- A survey of natural genetic variation in *Marchantia polymorpha* subsp. *ruderaleis*
- There is no population structure in European accessions
- SNPs associated with climate adaptation were identified
- This collection forms a foundation for future population genetics research



Article

Population genomics of *Marchantia polymorpha* subsp. *ruderale*s reveals evidence of climate adaptation

Shuangyang Wu,¹ Katharina Jandrasits,¹ Kelly Swarts,^{1,4} Johannes Roetzer,^{1,2} Svetlana Akimcheva,¹ Masaki Shimamura,³ Tetsuya Hisanaga,^{1,5} Frédéric Berger,¹ and Liam Dolan^{1,6,*}

¹Gregor Mendel Institute of Molecular Plant Biology (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna 1030, Austria

²Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna 1030, Austria

³Graduate School of Integrated Sciences for Life, Hiroshima University, 1-3-1 Kagamiyama, Higashi-hiroshima 739-8526, Hiroshima, Japan

⁴Present address: Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå 901 83, Sweden

⁵Present address: Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0192, Japan

⁶Lead contact

*Correspondence: liam.dolan@gmi.oeaw.ac.at

<https://doi.org/10.1016/j.cub.2025.01.008>

SUMMARY

Sexual reproduction results in the development of haploid and diploid cell states during the life cycle. In bryophytes, the dominant multicellular haploid phase produces motile sperm that swim through water to the egg to effect fertilization from which a relatively small diploid phase develops. In angiosperms, the reduced multicellular haploid phase produces non-motile sperm that is delivered to the egg through a pollen tube to effect fertilization from which the dominant diploid phase develops. These different life cycle characteristics are likely to impact the distribution of genetic variation among populations. However, little is known about the distribution of genetic variation among wild populations of bryophytes. To investigate how genetic variation is distributed among populations of a bryophyte and to establish the foundation for population genetics research in bryophytes, we described the genetic diversity of collections of *Marchantia polymorpha* subsp. *ruderale*s, a cosmopolitan ruderal liverwort. We identified 78 genetically unique (non-clonal) from a total of 209 sequenced accessions collected from 37 sites in Europe and Japan. There was no detectable population structure among European populations but significant genetic differentiation between Japanese and European populations. By associating genetic variation across the genome with global climate data, we showed that temperature and precipitation influence the frequency of potentially adaptive alleles. This collection establishes the core of an experimental platform that exploits natural genetic variation to answer diverse questions in biology.

INTRODUCTION

Sexual reproduction results in the development of haploid and diploid cell states during the life cycle. Land plants develop multicellular stages in both the haploid and diploid phases of the life cycle. The haploid phase of the life cycle is dominant over the diploid phase among plants in one of the two monophyletic groups of land plants, the bryophytes. The diploid phase is dominant over the haploid phase in the other monophyletic group, the vascular plants. The relative contributions of the haploid and diploid phases to the life cycle are likely to have consequences on the pattern of genetic diversity in bryophytes and vascular plants.¹

A great deal is known about the distribution of genetic diversity and adaptation among populations of angiosperms, a monophyletic group within the vascular plants (see for example Long et al.,² Huang et al.,³ and 1001 Genomes Consortium⁴). The angiosperm egg is fertilized by a non-motile sperm delivered to

the female by a pollen tube.⁵ By contrast, the bryophyte egg is fertilized by a motile sperm, which requires liquid water to effect fertilization.⁶ *Marchantia polymorpha* subsp. *ruderale*s is a liverwort with a haploid-dominant life cycle, typical of a bryophyte.⁷ It is an experimental model genetic system with a well-annotated genome of approximately 218 Mb.⁸ It is globally distributed in the northern hemisphere and grows in a diversity of climates and environments, often in ruderal communities. Given the diversity of habitats in which this plant is found, there is likely considerable variation in the ways in which these plants adapt to climate and other environmental factors. It may therefore be a system in which we can investigate how genetic variation is distributed among populations both locally and globally.

M. polymorpha subsp. *ruderale*s is a ruderal and grows in disturbed habitats with open ground.^{9–11} This includes disturbed open ground caused by human activity in urban areas, farms, and transport infrastructure. It also colonizes open ground that



forms after natural disturbance, such as the clearings that form after forest or peatland fires.¹² Colonizing populations can expand rapidly through asexual reproduction or a combination of sexual and asexual reproduction.¹³ If populations are derived from a single individual, census number can increase through asexual reproduction, and all individuals in the populations would be genetically identical (clones). Alternatively, populations may be derived from at least one male and one female, and the population would increase through both asexual and sexual reproduction.

The patterns of genetic variation and adaptation may be different in plants with a dominant haploid phase like bryophytes than in plants with a dominant diploid phase like angiosperms. It has been suggested that purging selection may be stronger in the haploid-dominant life cycle than in the diploid-dominant life cycle.¹ Furthermore, the diaspores—dispersal units—are different between angiosperms and bryophytes and may impact the distribution of genetic diversity. In *M. polymorpha*, there are two types of diaspores, spores and gemmae. Spores are produced by meiosis. They are approximately 10 µm in diameter and produced in their millions on individual plants.^{6,14} Their small size allows them to be carried over long distances by air and water currents, enabling long-distance gene flow. Gemmae are larger—approximately 500 µm in diameter and disc shaped—clonal propagules that derive from single cells in the epidermis of adult plants.⁶ While gemmae could be transported by water, they would be too heavy to be carried on air currents. The production of diaspores that could disperse both locally (gemmae) and over very long distances (spores) might be expected to impact genetic differentiation of populations of *M. polymorpha* on a landscape scale. Genome sequences are available for some *M. polymorpha* subsp. *ruderalis* accessions, indicating that there is genetic variation among populations.^{15,16} Based on the sequencing of 15 *M. polymorpha* subsp. *ruderalis* from Southern Ontario (Canada), Sandler et al. have suggested that gene flow can occur over long distances.¹⁶ They also highlighted that both sexual and asexual reproduction contributed to the pattern of genetic diversity in this locality. However, it is unclear how genetic variation is distributed on continental and global scales.

To discover the patterns of natural genetic variation in a species with a dominant haploid life cycle phase, we described the genetic diversity and structure in *M. polymorpha* subsp. *ruderalis*. We collected multiple individuals from subpopulations in Europe to evaluate local genetic diversity. We also collected individuals from single subpopulations in Europe and Japan to evaluate global patterns of genetic diversity. Sequencing 209 of these accessions identified 78 unique individuals, and 131 were genetically identical clones. The sampling of multiple individuals from local patches revealed considerable genetic variation in sexually reproducing local populations. While the European and Japanese accessions are genetically differentiated, we discovered there is no population structure in Europe.

RESULTS

Geographical distribution of 209 *M. polymorpha* subsp. *ruderalis* accessions

To capture the genetic diversity of natural populations of *M. polymorpha* subsp. *ruderalis* accessions, we collected a total

of 209 individuals from 37 geographical locations (Figure 1A; Data S1A) in Europe (183), Japan (23), and North America (3). The genomes of all 209 accessions were sequenced to an average coverage of 23.45vr with a 94.08% mapping rate for the nuclear genome (Data S1B). A total of 3,775,179 autosome SNPs were identified with joint genotyping followed by stringent filtering.¹⁷

Multiple individuals were collected from the same subpopulation in twelve sites in Europe (Data S1A). A single accession was collected at all other sites. *M. polymorpha* reproduces both vegetatively (through the production of propagules called gemmae) and sexually (through the production of spores produced by meiosis). Therefore, we hypothesized clonally related individuals may be present in the twelve subpopulations where multiple individuals were collected. To identify clonal individuals, we calculated the Hamming distance as the pairwise genome distance (GD) among the 209 sequenced accessions, using all autosomal SNPs with Plink (Data S1C). We identified two distinct peaks in the GD distribution (Figure 1B), with peak A corresponding to a GD of less than 0.01 and peak B corresponding to a GD of 0.24 (see also Figure S1). We conclude that pairs of accessions with GD values less than 0.01 (peak A) were clones. Using this criterion, two subpopulations—Sopron (HUN, three letter country code of Hungary) and MaG (AUT, three letter country code of Austria) with GD values at 0.008 and 0.0091, respectively—were shown to be genetically identical and therefore clonal (Figure S1). Peak B with higher GD (0.24) included individuals from each of the other ten subpopulations. The contribution of individuals from the ten subpopulations to peaks A and B indicates that some individuals in these populations are derived from sexual reproduction (GD greater than 0.24) while others are derived from vegetative reproduction (GD less than or equal to 0.01). This conclusion was further confirmed by genetic diversity analysis. We measured significantly lower genetic diversity in asexually reproducing (clonal) subpopulations than in sexually reproducing subpopulations (Figure S2; Data S1D). Taken together, these data demonstrate that there were 78 genetically unique accessions originating from 37 diverse geographical locations in the collection (Data S1E). Furthermore, some subpopulations are entirely clonal (Figures S1 and S2), while others are derived from a combination of sexual and asexual reproduction. This refined dataset now serves as a robust foundation for elucidating the genetic diversity present within *M. polymorpha* subsp. *ruderalis* across global landscapes.

Population structure of 78 accessions reveals that Japanese and European populations are genetically distinct

Population structure results from systematic differences in allele frequencies between subpopulations of a population.¹⁸ To explore the broad-scale population structure among accessions, we performed a principal-component analysis (PCA) of the 3,479,055 pure autosome SNPs that were identified from the 78 samples. Three genetically distinct clusters were identified: Japan (left bottom), Cambridge (right top), and a mixed origin cluster (right bottom) (Figure 2A). PC1, which accounts for 8.0% of the variation in the collection distinguishes two groups: Japan and Europe (including the 2 isolates from North America). PC2, which accounts for only 3.0% of the variation,

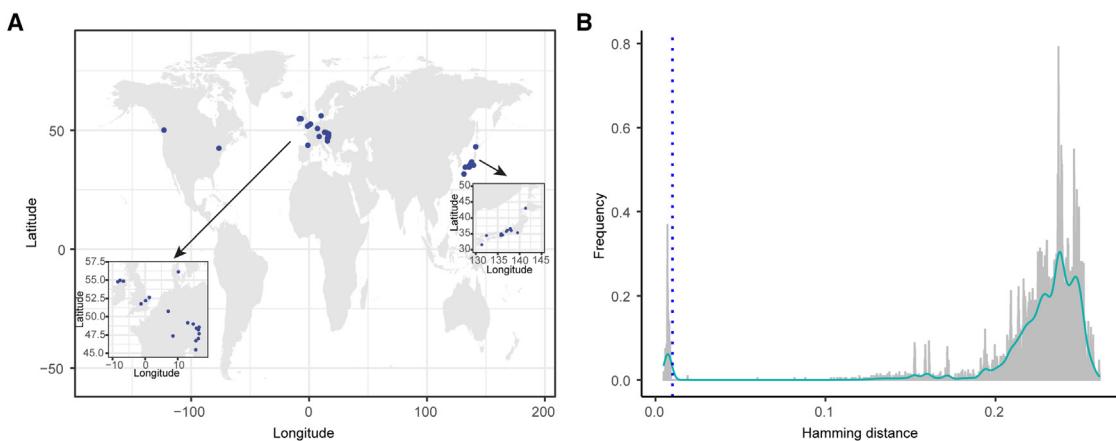


Figure 1. Geographical distribution of 209 *M. polymorpha* subsp. *ruderaleis* accessions and pairwise GD

(A) Geographical distribution of 209 *M. polymorpha* subsp. *ruderaleis* accessions. The blue dots represent the 37 collection sites. The small box on the left shows a zoomed-in view of the European sites, while the small box on the right shows a zoomed-in view of the Japanese sites. See also Data S1A.

(B) Pairwise GD distribution of 209 accessions, and blue dashed line is the cutoff of clonal populations.

See also Figures S1 and S2 and Data S1B and S1C.

separates two European accessions—Cam-1 and Cam-2 from the UK from the other European accessions. Further examination of the European cluster revealed accessions grouped, independent of location (country of origin) (Figure 2B), suggesting that there is no population structure among European accessions. To independently verify the conclusion of the PCA analysis, we performed admixture analyses. Cross-validation (CV) error analysis from admixture indicated that the optimal grouping was achieved at $K = 2$ (Figure 2C). Two distinct groups were identified based on the genetic profile of the complete collection of accessions and their geographical origins (Japan versus Europe and the North American continent) (Figure 2D). When we focused the admixture analysis on the European subpopulations, we found evidence for mixed genetic ancestries of European populations from K3 to K9 (Figure S3A), consistent with a lack of distinct population structure within the European accessions. In two European subpopulations sampled—Schubert (7 individuals) and Ardara (9 individuals)—we found the same number of ancestries as in the broader European populations (from K7 to K9). These data are consistent with the hypothesis that levels of genetic diversity are high in sexually reproducing subpopulations (Figure S3B). The absence of population structure among the European populations, as indicated by the admixture analysis, is consistent with the findings from the PCA analysis (Figures 2A and 2B). In summary, these data demonstrate that the Japanese and European populations are genetically different. However, we detected no evidence of population structure in Europe.

There is no substantial genetic differentiation among the European *Marchantia polymorpha* subsp. *ruderaleis* populations

To test for genetic diversity differences between the European subpopulations, we measured the nucleotide diversity (π) for the seven subpopulations, where three or more individuals were collected (single representatives of clones were included in each respective subpopulation for the analysis to remove

bias introduced by including multiple genetically identical individuals). Autosome SNPs were identified, and the nucleotide diversity for each of these subpopulations was close to 0.0036; Bonn_DEU ($n = 3$), Ardara_IRL ($n = 9$), Letterkenny_IRL ($n = 3$), Oxford_GBR ($n = 4$), Norwich_GBR ($n = 3$), Schubert_AUT ($n = 7$), and Zurich_CHE ($n = 3$) (Figure 3A; Data S1F). This suggests that similar levels of genetic diversity are present in these subpopulations. We then measured genetic divergence (D_{xy})—a measure of the genetic differences between the subpopulations. D_{xy} values among the seven subpopulations were approximately 0.0042 (Figure 3B), which is close to the π value of each subpopulation. This indicates that there are no significant differences in genetic diversity between these European subpopulations.

Fixation index (F_{st}) quantifies the proportion of genetic variation attributable to differences between subpopulations relative to the total genetic variation in a population.¹⁹ To test if genetic differentiation exists between these subpopulations across Europe, an F_{st} analysis was performed. It measured the degree of differentiation between pairs of the 7 subpopulations. This resulted in 21 values from all possible pairwise comparisons. The highest F_{st} value observed was 0.00517, indicating that there was no significant differentiation between any of the European subpopulations (Figure 3C). By contrast, in *Arabidopsis thaliana*, genetic differentiation tends to increase with geographic distance. For example, the F_{st} between populations in North Sweden and South Sweden is ~ 0.2 .² However, our findings suggest that this pattern of genetic differentiation does not hold for *Marchantia polymorpha* subsp. *ruderaleis*.

To test the hypothesis that genetic distance does not increase with geographic separation, an isolation by distance (IBD) analysis was performed between the seven subpopulations. The results revealed no significant positive correlation of IBD between the European subpopulations, as determined by the Mantel test (Figure 3D). We also carried out an isolation by environment (IBE) analysis. This assesses the correlation between genetic distance and environmental distance. Again, there was no positive

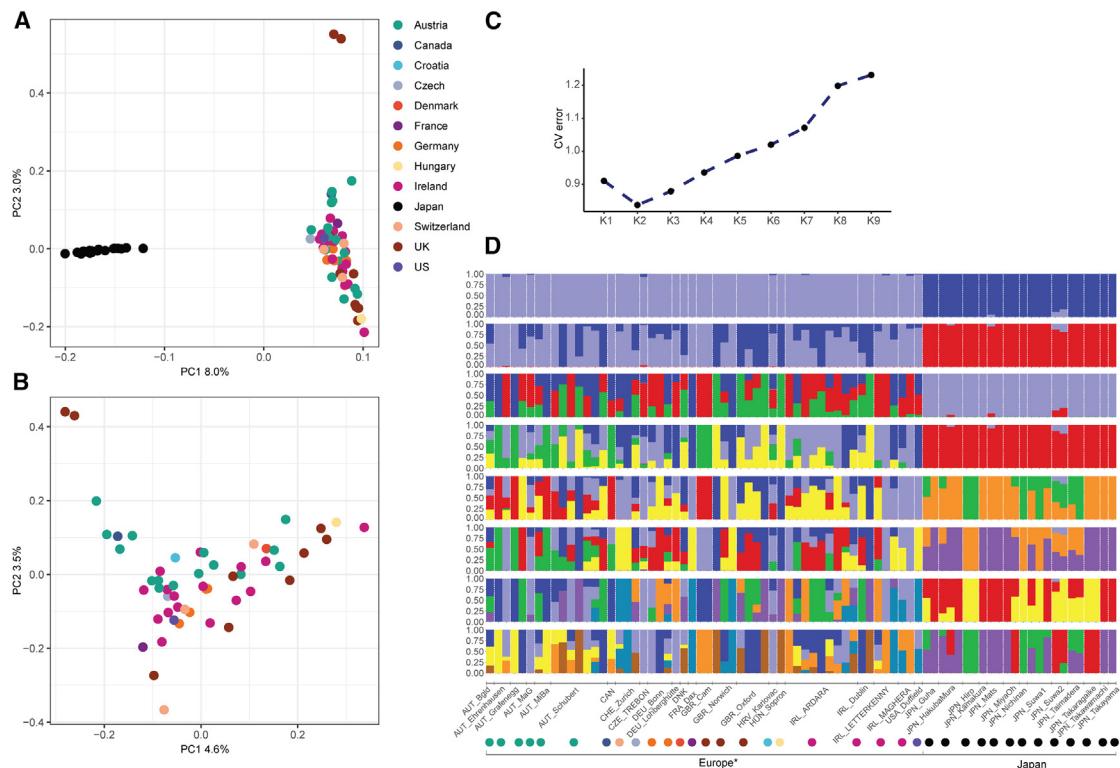


Figure 2. Genetic relationship of collected accessions

(A) Principal-component analysis (PCA) of all SNPs revealed 3 groups, including Japanese (left bottom), Cambridge (right top), and a mixed origin cluster (right bottom). See also [Data S1E](#) and [Figure S10](#).

(B) PCA of European SNPs revealed no population structure within European populations.

(C) Cross-validation (CV) error analysis of the admixture result, demonstrating the optimal number of groups. The optimal K is 2.

(D) Unsorted genetic profiles of the 78 accessions are shown for K2 to K9. The vertical bars are color coded, with blue, light blue, red, green, yellow, orange, purple, navy blue, and brown representing the genetic profile identified through k-fold CVs. In K2, there are two colors (blue and light blue); in K3, three colors (blue, light blue, and red); in K4, four colors (blue, light blue, red, and green); in K5, five colors (blue, light blue, red, green, and yellow); in K6, six colors (blue, light blue, red, green, yellow, and orange); in K7, seven colors (blue, light blue, red, green, yellow, orange, and purple); in K8, eight colors (blue, light blue, red, green, yellow, orange, purple, and navy blue); and in K9, nine colors (blue, light blue, red, green, yellow, orange, purple, navy blue, and brown). Dots below are color coded by country, consistent with the color scheme in (A). *Two accessions from Canada and USA.

See also Figure S3.

correlation between genetic distance and environmental distance (Figure 3E). Together, these data support the hypothesis that there is no substantial genetic differentiation among the European subpopulations sampled.

There is substantial genetic differentiation between the European and Japanese *M. polymorpha* subsp. *ruderalis* populations

Using the dataset from 54 European and 22 Japanese accessions, we found that these two groups of accessions were genetically distinct from each other. Genetic diversity (π) was higher among the European accessions than among the Japanese accessions across all autosomes (Figures 4A and S4A). The π calculated from the autosomes of the European accessions was 0.00375, and 0.00246 for the Japanese accessions (Figure S4B). To test if π might be correlated with sample size, we compared samples that were similar in size—ranging from 10 to 21 individuals—from Japan and Europe. In Figure S4C, we show the genetic diversity (π) of these European accessions was the same irrespective of sample size—it ranged from 0.00377 to 0.00379. The genetic

diversity (P_i) of the Japanese accessions was also the same irrespective of sample size— P_i ranges from 0.00244 to 0.00247. This analysis demonstrates that the difference in P_i values between the European and Japanese populations was not due to the larger sample size of the European group. Having shown that P_i is different between European and Japanese accessions, we then tested if there is genetic diversity divergence between these two populations by calculating D_{xy} . D_{xy} between the two populations was 0.00403, which is much larger than the P_i value of Japanese accessions (0.00246). This suggests that the two groups of accessions are genetically differentiated. To further test genetic differentiation between European and Japanese groups of accessions, we calculated the F_{ST} value. The genome-wide F_{ST} between Japanese and European populations was 0.137. This was substantially higher than the within-region population differentiation values (with a mean of 0.0081 for Europe) and is consistent with the differentiation revealed by the PCA and admixture analysis (Figures 4B and S4D). This indicates that there was substantial genetic differentiation between the two—European and Japanese—groups of accessions.

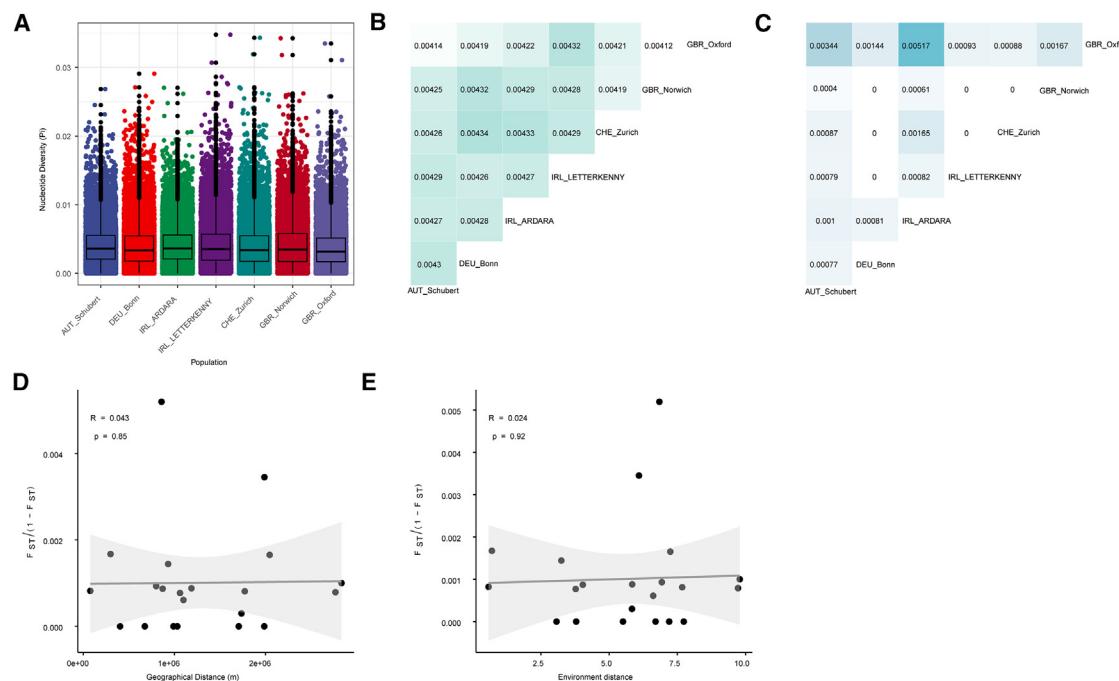


Figure 3. Genetic diversity and differentiation among the seven European populations

- (A) Genetic diversity among population groups was assessed by calculating nucleotide diversity in 20 kb sliding windows across the autosomes, with the median value indicated by the horizontal line. See also Figure S2 and Data S1F.
- (B) Pairwise median Dxy values between populations are represented by color-coded values, reflecting the genetic diversity between them.
- (C) Pairwise mean Fst values between populations are depicted using a color scale, illustrating the extent of genetic differentiation between population pairs.
- (D) Isolation by distance analyses conducted across the seven European populations. The shadow of the linear regression line represents the 95% confidence interval. The *p*-value corresponds to the statistical significance of the slope in the linear regression model.
- (E) Isolation by environment analyses performed within the seven European populations. The shadow of the linear regression line signifies the 95% confidence interval. The *p*-value corresponds to the statistical significance of the slope in the liner regression model.

To identify which chromosomes contribute to the genetic differentiation between European and Japanese accessions, the Fst values for each chromosome were calculated. Chromosomes 5 and 8 contribute more to genetic differentiation than other autosomes (Figures 4B and S4E). These higher Fst values along chromosomes 5 and 8 suggested that pericentromeric regions enriched in non-protein-coding sequences contributed more to the genetic differentiation than the arms (Figure 4A). Furthermore, chromosomes 5 and 8 had distinct centromeric haplotypes. To verify that sequences at pericentric regions of these two chromosomes contributed more to genetic differentiation than sequences in arms, we partitioned each autosome into pericentromere, long arm, and short arm. The analysis demonstrated significantly higher differentiation in the pericentromeric regions of chromosomes 5 and 8 compared with the arm regions of these chromosomes (Figure 4C). These data are consistent with the hypothesis that sequences in the vicinity of the centromeres of chromosomes 5 and 8 contribute disproportionately to the differentiation between the Japanese and European accessions. In conclusion, Japanese and European populations are genetically distinct, and sequence variation in the vicinity of the centromeres of chromosome 5 and chromosome 8 contributes to the high levels of genetic differentiation (higher Fst) (Data S1G).

LD and population demography of Japanese and European populations

Linkage disequilibrium (LD) is the tendency of genetic variants located near each other on a chromosome to be inherited together more frequently than expected. LD is an essential measure to test if there are sufficient genetic markers in the populations for association mapping. To measure LD, we first measured LD decay using allele frequency correlation (r^2) between linked SNP markers. LD was then measured as the pairwise SNP distance where r^2 is half the maximum value.^{20,21} LD decay rate was higher in European groups than in Japanese groups (Figure 5A). From the decay plot (Figure 5A), LD of the European population was estimated to be 1.6 kb ($r^2 = 0.1694$), while the LD of the Japanese population was estimated to be 4.3 kb ($r^2 = 0.3262$). LD for the entire collection of 78 individuals was estimated to be 1.5 kb. This estimation of LD is lower than in *Arabidopsis* (LD: 10 kb) and rice (LD: 150 kb).^{3,22} The minimum number of genetic markers required for genome-wide association (GWA) can be estimated by dividing the genome size by the LD. The autosome size of *M. polymorpha* subsp. *ruderaria* is ~218 Mb. Therefore, the minimum number of SNPs is 218 Mb/1.5 kb = ~145,334 (n). Since we identified 3,479,055 SNPs in the collection, we conclude that there are sufficient genetic markers in this *M. polymorpha* subsp. *ruderaria* accession set to perform GWA analysis.

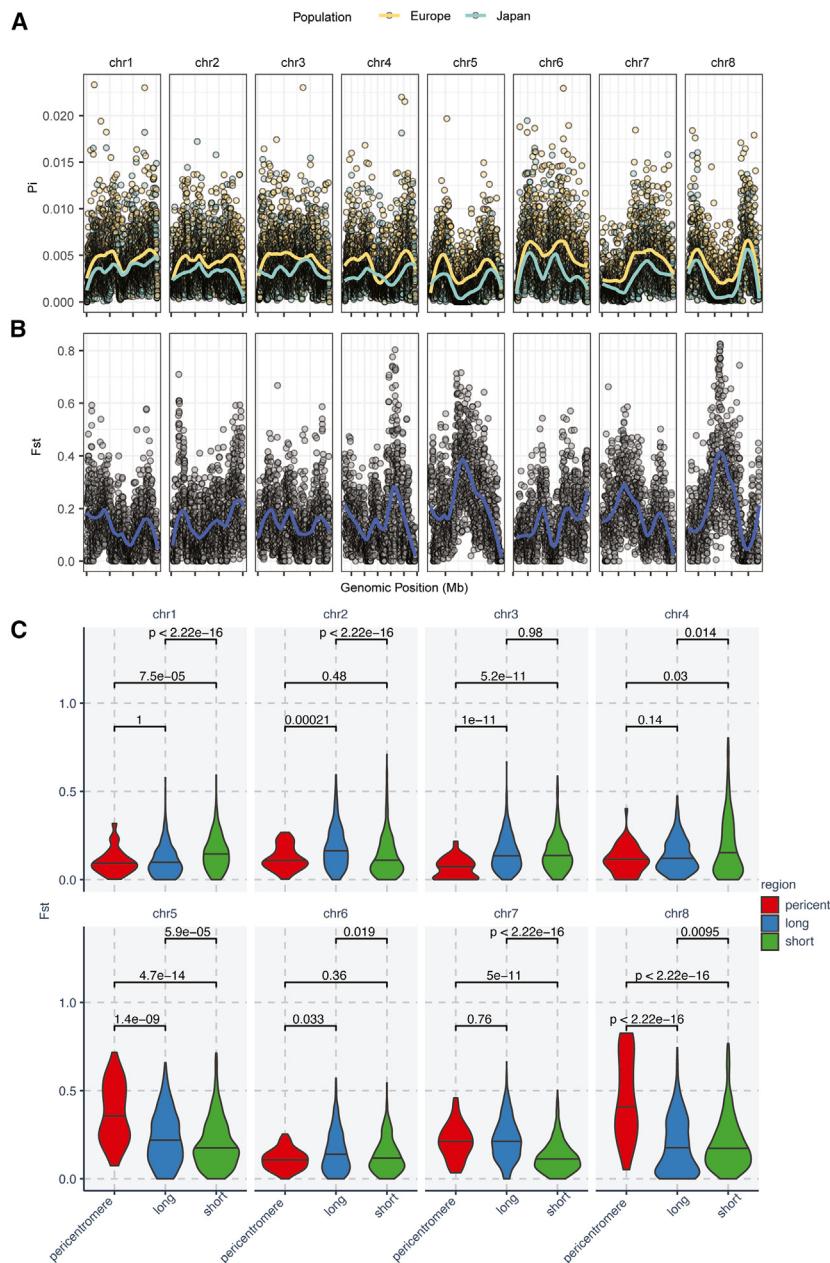


Figure 4. Chromosomal patterns of *Marchantia polymorpha* subsp. *ruderalis* genetic diversity and differentiation

(A) Genome-wide nucleotide diversity for European (yellow) and Japanese (green) populations. Each dot represents the Pi value for a sliding window of 20 kb. See also Figures S4A and S4B.

(B) Genome-wide Fst distribution between European and Japanese populations. Solid lines depict LOESS (locally estimated scatterplot smoothing) functions that have been fitted to the data, providing the representation of trends and patterns. See also Figure S4E.

(C) The Fst distribution was analyzed for three segments of each autosome, the pericentromere, the long arm, and the short arm of chromosomes. The p-values were obtained using Wilcoxon test. See also Data S1G.

Summer temperature and precipitation affect the frequency of adaptive alleles in European and Japanese accessions

To find evidence of genetic adaptation to environmental factors, we searched for the association of markers with local climate. WorldClim2 is a global climate database that provides high-resolution data for climatic variables across the world, including 11 temperature factors and 8 precipitation factors.²³ We used GWA in combination with global climate data to test if there is genetic adaptation to specific climate factors. Integrating allele frequencies and climatic data using the gradient forest (GF) method,²⁴ we identified environmental variables correlated with genetic variation and assessed how allele frequencies changed along environmental gradients. A total of 2,191 climate-associated genetic variants (SNPs) were identified using latent factor mixed models (LFMMs).²⁵ We calculated the allele frequencies from the 2,191 SNPs in our European and Japanese accessions.

Relatively low LD indicates larger effective population sizes, while higher LD in a population indicates a lower effective population size. The lower LD in the European population compared with the Japanese population (Figure 5A) suggests that the effective population size is smaller in Japan than in Europe. To test this hypothesis, we characterized the effective population size of European and Japanese populations (Figure 5B). Our results revealed that current *M. polymorpha* subsp. *ruderalis* effective population sizes (N_e) are significantly smaller than ancestral populations (10^6 years ago). We estimate that the European N_e declined dramatically approximately 13,000 years ago, while the N_e declined approximately 5,000 years ago in Japan. Furthermore, the European current effective population size of ($N_e = \sim 3,000$) is larger than the current Japanese effective population size ($N_e = \sim 700$).

Following GF ranking, significant correlations with five climate variables were identified: temperature seasonality (BIO4), precipitation of warmest quarter (BIO18), maximum temperature of warmest month (BIO5), mean temperature of warmest quarter (BIO10), and mean temperature of wettest quarter (BIO8) (Figure 6A). Maximum temperature of warmest month (BIO5), mean temperature of the warmest quarter (BIO10), and mean temperature of the wettest quarter (BIO8) were ranked equally important. These data indicate that warm temperature and precipitation are the key climate factors influencing allele frequencies.

According to GF analysis, four temperature factors and one precipitation factor are among the climate variables that most affect allele frequencies. To test if these climate variables are

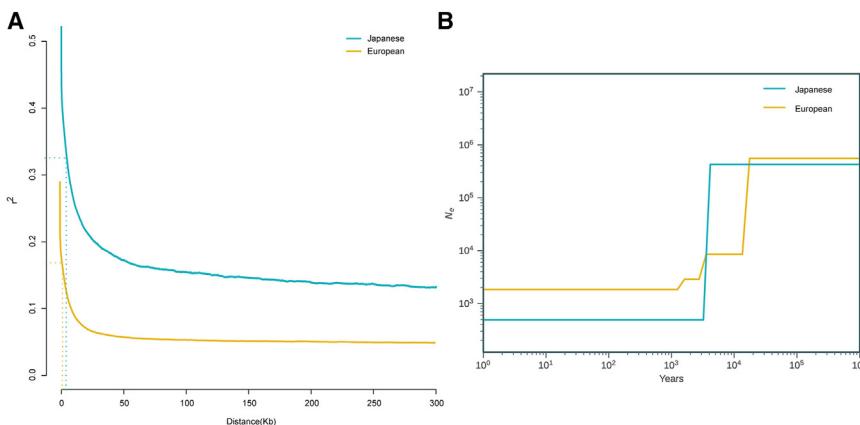


Figure 5. The LD decay and population demographic history of each population

(A) LD decay in Japanese and European populations. Blue solid line is LD decays plot for the Japanese groups, and yellow solid line is LD decay plot for the European groups. The LD decay rate of the European group is greater than in the Japanese group. The blue dashed line marks the X-intercept at half-maximal r^2 for the Japanese group (LD). The yellow dashed line marks the X-intercept at half-maximal r^2 for the European group (LD).

(B) Effective population sizes (N_e) distribution in Japanese groups (solid blue line) and European groups (solid yellow line).

seasonal, we performed pairwise Spearman correlation analysis among the 19 climate factors. The pairwise Spearman correlation coefficients of BIO4, BIO5, BIO10, and BIO8 exceeded 0.85, indicating a significant pairwise positive correlation, suggesting these four temperature factors are associated with warm season temperature (Figure S5). Furthermore, these four factors were significantly negatively correlated with BIO6 (minimum temperature of coldest month) and BIO11 (mean temperature of coldest quarter), consistent with the positive association with warm season temperatures. Furthermore, all four temperature variables (BIO4, BIO5, BIO10, and BIO8) were positively correlated (**) with BIO18 (maximum temperature of the warmest month), suggesting synergistic effects of precipitation of warmest quarter with the warm season temperature. These data suggest that summer temperature and precipitation may be the most influential climate factors affecting the frequency of adaptive alleles.

To test if geographical groups were adapted to specific local climates, we integrated 2,191 SNP allele frequencies and the 5 most important climate variables and performed allelic turnover analyses with 22 geographical locations. The allelic turnover function predicts how frequencies of genetic variants in different geographical groups change with associated environmental variables. The overall allelic frequency turnover function for adaptive SNPs (cumulative importance of allele frequencies, y axis) projected along these 5 factors (BIO4, BIO18, BIO5, BIO10, and BIO8, x axis) was sigmoidal (Figures 6B–6F). Geographical group projections of cumulative importance of allele frequencies indicated distinct patterns between European and Japanese groups. The Japanese groups were in the “upper” region of the sigmoid (higher cumulative importance of allele frequencies) while the European groups were in the “lower” region of the sigmoid (lower cumulative importance of allele frequencies). The sigmoid of the allelic turnover function suggested that the Japanese groups are adapted to warmer temperatures and higher precipitation, and European accessions are adapted to cooler temperatures and lower precipitation (Figure S6). Furthermore, the analysis suggests that European Atlantic accessions are adapted to cooler wetter climate conditions than those in central Europe.

Analysis of sigmoid patterns indicated that cumulative allele frequency importance correlated with five climate variables ($r = 1$, $p < 2.26e-16$, Figure S7) across 22 geographical

locations. To test if longitude and/or latitude influence adaptation, we projected cumulated importance of allelic frequencies (y axis) onto longitude and latitude of geographic groups (x axis). Projection of cumulative allele frequency importance onto longitude and latitude highlights associations with maximum temperature of warmest month (BIO5) and mean temperature of warmest quarter (BIO10), in European and Japanese groups (Figures S8A and S8B). In both groups, cumulative importance increased with decreasing latitude (north to south) for BIO5 and BIO10, while cumulative importance increased with increasing longitude (west to east) in Europe (Figures S8A and S8B, Data S1H). Conversely, cumulative importance in Japanese groups increased slightly with decreasing longitude (East to West) (Figures S8C and S8D; Data S1H). The correlation extended to BIO4, reflecting seasonal climate variation, where cumulative importance increased with decreasing latitude in European groups (Figure S8E) and with increasing latitude and longitude in Japanese groups (Figures S8E and S8F). However, no positive correlation was found in Japan for BIO8 and BIO18 (Figures S8G–S8J). The cumulative importance increased with increasing longitude for BIO8 (Figure S8J), while cumulative importance tended to increase with decreasing latitude for BIO8 and BIO18 in European groups. We conclude that the cumulated importance of allelic frequencies correlated with climate variables (BIO4, BIO5, and BIO10) that varied with latitude and longitude in Europe. Similarly, in Japan, the cumulated importance of allelic frequencies was correlated with climate variables (BIO4, BIO5, BIO10, BIO8, and BIO18) that varied with latitude and longitude. These data are consistent with the hypothesis that the frequencies of genetic variants in different geographic groups shift with different environmental variables.

Association mapping of the five most important climate variables

Genotype-environment association studies (GEAS) can discover genetic variants correlated with local climate factors.²⁶ We searched for association between genetic variants and climate variables. Before we performed GEAS, we calculated the heritability of 19 climate factors because traits characterized by high heritability may be mapped by genetic association. Heritability (H^2) values exceeded 0.99 for BIO4, BIO5, BIO8, and BIO10 and were 0.89 for BIO18 (Data S1I). These data indicated that it should be possible to test for genetic association with BIO4,

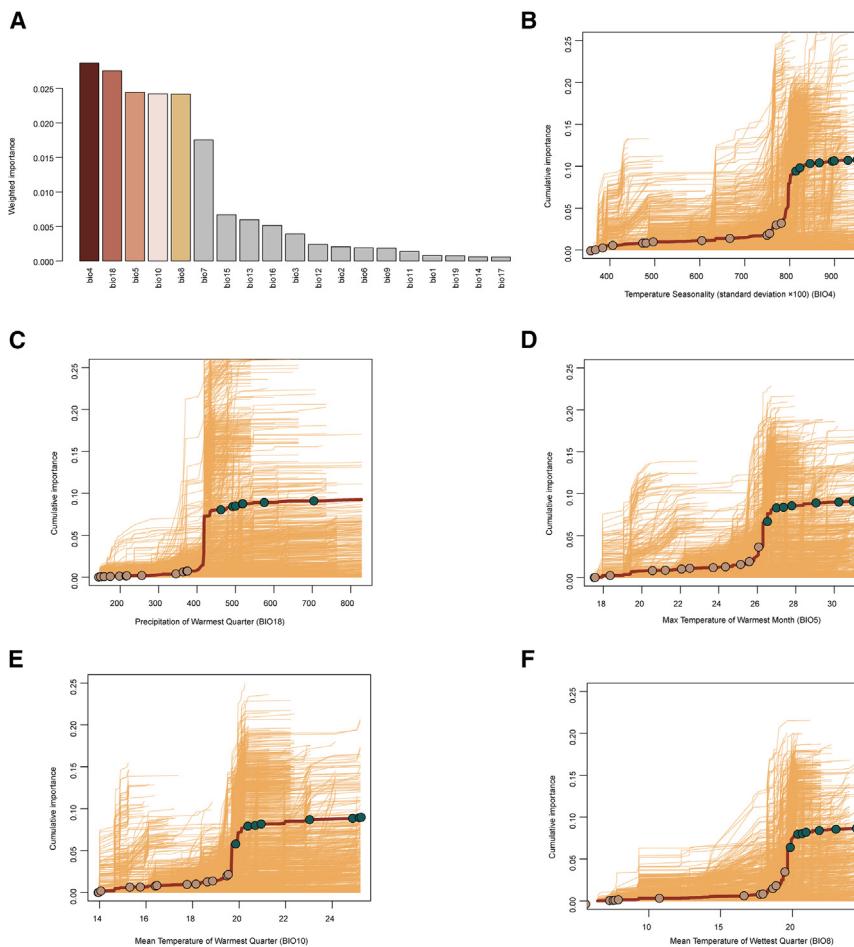


Figure 6. GF and turnover function in climate factors

(A) Ranked importance of 19 climate factors based on GF analysis using 2,191 candidate SNPs. The top five bioclimatic factors in the color gradient demonstrate the strongest inferred correlation. See also [Figure S5](#).

(B–F) Allelic turnover functions in relation to bioclimatic factors BIO4 (B), BIO18 (C), BIO5 (D), BIO10 (E), and BIO8 (F) on the x axis. The y axis represents cumulated importance of allele frequencies, indicating the significance of SNPs in the GF models and reflecting the total turnover in allele frequency across the seasonality gradient. Thin yellow lines depict cumulative turnover for adaptive SNPs. The solid red line indicates turnover across all candidate SNPs, with circles along the line representing geographical groups organized based on different bioclimatic factors. Colors denote their inclusion in either the Japanese (green) or the European geographical groups (brown).

See also [Figures S6–S8](#) and [Data S1H](#).

BIO5, BIO8, BIO10, and BIO18. GEAS was performed after employing a stringent Bonferroni-based threshold and quantile-quantile (Q-Q) plots method on the five most influential climate factors ([Figure S9](#)). We found six unlinked, significant SNPs associated with BIO4, ten unlinked, significant SNPs associated with BIO5, four unlinked, significant SNPs associated with BIO10, and nine unlinked, significant SNPs associated with BIO8 ([Figure S9; Data S1J](#)). However, we did not find genetic variants significantly associated with BIO18 ([Figure S9](#)). These data demonstrate that specific genetic variants putatively associated with roles in adaptation to local climate factors can be identified.

DISCUSSION

We report an initial analysis of natural variation of accessions of *M. polymorpha* subsp. *ruderale*s from Europe and Japan to establish a framework for population genetics in this species. We found considerable genetic diversity within local subpopulations. Despite the high genetic diversity of some subpopulations, others comprise individuals that are genetically identical. These patches are derived from a single founder and expand through asexual reproduction. Among the genetically diverse patches, some individuals are genetically identical. This demonstrates that both sexual and asexual reproduction occurs in these

subpopulations. Despite substantial local genetic diversity, we found no evidence of population structure across Europe. The reasons for the lack of genetic structure are unknown. However, it is likely that this lack of structure results, in part, from high levels of gene flow across the continent. Despite the lack of population structure on a continental scale across Europe, European and Japanese populations are genetically distinct. The genetic differentiation likely results from limited gene flow between European and Japanese populations.

The lack of population structure in Europe may be due, in part at least, to long-distance dispersal and gene flow throughout the continent. A similar lack of population structure has been identified in the conifer *Pinus sylvestris*.²⁷ The lack of structure may result from the production of large amounts of haploid pollen (30–100 µm in diameter²⁸) that travel large distances on air currents. Such potential for long-distance gene flow could contribute to a lack of genetic differentiation on a continental scale. We speculate that, like the haploid pollen of *P. sylvestris*, the potential for long-distance transport of the *M. polymorpha* subsp. *ruderale*s haploid spores⁷ may account for the lack of population structure in Europe. This lack of genetic differentiation on a continental scale contrasts with what has been observed in the ruderal angiosperm *A. thaliana*. The 1001 Genomes Consortium confirmed that the genetic distance between individuals reflects geographic distance in Europe.⁴ Furthermore, the native Eurasian range of *A. thaliana* exhibits continuous IBD at every geographic scale.^{2,29} By contrast, we found no evidence that isolation by geographical distance contributes to genetic distance in European *M. polymorpha* subsp. *ruderale*s. The differences in genetic structure between *M. polymorpha* subsp. *ruderale*s and *A. thaliana* across Europe may be accounted for by the different reproductive and dispersal strategies

of these species: *M. polymorpha* is an outcrossing species and produces large number of small spores and fewer relatively heavy gemmae, while *A. thaliana* is largely inbreeding and produces fewer relatively heavy seeds.

Mapping genetic diversity from plants growing in different environments allowed us to use genetic association to identify alleles that may underpin local adaptation. We report significant genetic adaptation to summer temperature and precipitation, particularly in the Japanese population. Our demonstration that there may be a genetic basis for adaptation to climate factors using GWA is consistent with previous findings in other plant species. For example, GWA studies in *A. thaliana*³⁰ and *Populus koreana*³¹ identified genomic regions associated with adaptation to temperature and precipitation.

The discovery that summer temperature and precipitation play a significant role in shaping the genetic landscape of populations in *M. polymorpha* may reflect the requirement for water in the bryophyte life cycle at this time of the year. Bryophytes produce motile spermatozoids that swim through water to effect fertilization. It is therefore likely that the summer precipitation is a constraint on the completion of the *M. polymorpha* sexual life cycle. Such a constraint would impose considerable selection pressure that would be reflected in the presence of adaptive alleles that are correlated with temperature and precipitation during the time of year that *M. polymorpha* sexually reproduces. These alleles would be strongly selected in ruderal environments, which are generally open and subject to drying because of direct incident sunlight on the soil surface and exposure to the desiccating effect of wind.

The accessions were collected from different environments and could potentially be used to identify the genetic basis for a range of environmental adaptations. For example, the European accessions include plants growing in a North Atlantic maritime climate (e.g., Ardara_IRL) and continental climate (e.g., Schubert). The Ardara subpopulation grows in full light without shade, while the Schubert subpopulation grows in shade and never receives direct sunlight. Furthermore, the Schubert plants generally shrivel during the hot summer in Vienna and revive in the autumn. This occasionally happens in the Ardara environment but much less frequently. It is likely that plants in Ardara shrivel less during the summer because of the lower mean temperature and higher precipitation along the Atlantic coastline than in central Europe at this time of the year. Genetic variation associated with adaptation to different light, temperature, and precipitation regimes may be found by comparing maritime Ardara and continental Schubert accessions. Given the diversity of environments from which the 78 accessions were collected, it is likely that the genes associated with other diverse environmental adaptations will be identified among these lines in the future.

Conclusions and perspectives

We present a population genomic analysis of *M. polymorpha* subsp. *ruderalis* natural selection and adaptation based on resequencing of unique 78 genomes of wild accessions mainly from Europe and Japan. There was considerable genetic variation within subpopulations that reproduced both sexually and asexually. Other subpopulations reproduce exclusively through asexual reproduction and are genetically monomorphic. Furthermore, we observed little differentiation among European

subpopulations. By contrast, Japanese and European populations are genetically distinct. The absence of genetic structure across Europe indicates that genetic diversity is distributed differently from other ruderal plants like *A. thaliana*, where there is considerable genetic differentiation across the continent. Summer temperature and precipitation appear to be the most important climate factors affecting the frequency of adaptation alleles among these populations. Further sequencing of individuals in populations from all continents and isolated islands will allow higher-resolution mapping of alleles associated with local adaptation.

This collection will form the core around which a platform for population genetics can be constructed. Such a platform will allow exploration of the population genetics of an organism with both multicellular haploid and diploid phases in its life cycle and in which the haploid phase is dominant to the diploid phase. The platform will constitute a model system in which the population genetics of sex chromosomes—which are not present in the many seed plant genomic model species—can be investigated. This foundational genetic resource can be exploited by a diversity of biologists to investigate almost any question where genetic variation can provide a way to understanding the process in more detail.

RESOURCE AVAILABILITY

Lead contact

Requests for resources and further information should be directed to Liam Dolan (liam.dolan@gmi.oeaw.ac.at).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- High-throughput sequencing data generated in this study have been deposited at NCBI SRA with code SRA: PRJNA1127023.
- All scripts and codes performed in this study are available on GitHub: <https://github.com/wushyer/Population-genomics-of-Marchantia-polymorpha-sub.-ruderalis>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank Lab Support of GMI/IMBA/IMP and the VBCF Plant Sciences unit for their support and the Next Generation Sequencing Facility for generating the DNA sequencing data (VBCF). We thank Jonathan Terhorst for his suggestions of using SMC++ on *M. polymorpha* sub. *ruderalis* population history analysis. We thank Masanobu Yoshidomi and Takahiro Mochizuki for help in collecting plant materials in Japan. We are grateful to two anonymous reviewers whose detailed and helpful comments helped us to improve the manuscript. This work was funded by a grant from the Austrian Academy of Sciences to the Gregor Mendel Institute and a European Research Council advanced grant DE-NOVO-P (project no. 787613) to L.D. and the Marie Skłodowska-Curie Grant (agreement no. 847548, VIP2) to S.W.

AUTHOR CONTRIBUTIONS

S.W., F.B., and L.D. designed the project. S.W. carried out the analysis with help from K.S. K.J. and S.A. cultivated the isolated accession. Accessions were collected by J.R., M.S., T.H., F.B., and L.D. K.J. isolated DNA and performed genome sequencing and maintained the accessions. S.W. analyzed the DNA sequence data. S.W. and L.D. wrote the manuscript. The manuscript

draft was extensively revised with considerable input from F.B., J.R., and K.S. All authors contributed to revising the manuscript.

DECLARATION OF INTERESTS

L.D. is a co-founder, shareholder, and board member of MoA Technology.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Genomic DNA sequencing
 - Read mapping and variant calling
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Removal of clones produced by asexual reproduction
 - Population genetic analysis
 - Isolation by distance (IBD) and isolation by environment (IBE) analysis
 - Demographic history analysis
 - Genotype-environment associations

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2025.01.008>.

Received: November 5, 2024

Revised: December 16, 2024

Accepted: January 7, 2025

Published: February 10, 2025

REFERENCES

1. Immler, S., and Otto, S.P. (2018). The Evolutionary Consequences of Selection at the Haploid Gametic Stage. *Am. Nat.* 192, 241–249. <https://doi.org/10.1086/698483>.
2. Long, Q., Rabanal, F.A., Meng, D., Huber, C.D., Farlow, A., Platzer, A., Zhang, Q., Vihjálmsson, B.J., Korte, A., Nizhynska, V., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884–890. <https://doi.org/10.1038/ng.2678>.
3. Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. <https://doi.org/10.1038/ng.695>.
4. 1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oew.ac.at; 1001 Genomes Consortium (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481–491. <https://doi.org/10.1016/j.cell.2016.05.063>.
5. Kenrick, P., and Crane, P.R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33–39. <https://doi.org/10.1038/37918>.
6. Shimamura, M. (2016). *Marchantia polymorpha*: Taxonomy, Phylogeny and Morphology of a Model System. *Plant Cell Physiol.* 57, 230–256. <https://doi.org/10.1093/pcp/pcv192>.
7. Bowman, J.L., Arteaga-Vazquez, M., Berger, F., Briginshaw, L.N., Carella, P., Aguilar-Cruz, A., Davies, K.M., Dierschke, T., Dolan, L., Dorantes-Acosta, A.E., et al. (2022). The renaissance and enlightenment of *Marchantia* as a model system. *Plant Cell* 34, 3512–3542. <https://doi.org/10.1093/plcell/koac219>.
8. Montgomery, S.A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J.L., Cognat, V., Maréchal-Drouard, L., et al. (2020). Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr. Biol.* 30, 573–588.e7. <https://doi.org/10.1016/j.cub.2019.12.015>.
9. Brinkman, A.H. (1929). Hepatics and Sites: A Short Study in the Ecology of Hepatics. *Bryologist* 32, 29–30. <https://doi.org/10.2307/3238614>.
10. Atherton, I., Bosanquet, S., and Lawley, M. (2010). *Mosses and Liverworts of Britain and Ireland - a Field Guide* (British Bryological Society).
11. Linde, A.M., Sawangproh, W., Cronberg, N., Szővényi, P., and Lagercrantz, U. (2020). Evolutionary History of the *Marchantia polymorpha* complex. *Front. Plant Sci.* 11, 829. <https://doi.org/10.3389/fpls.2020.00829>.
12. Graff, P.W. (1936). Invasion by *Marchantia polymorpha* Following Forest Fires. *Bull. Torrey Bot. Club* 63, 67–74. <https://doi.org/10.2307/2481055>.
13. Stieha, C.R., Middleton, A.R., Stieha, J.K., Trott, S.H., and McLetchie, D.N. (2014). The dispersal process of asexual propagules and the contribution to population persistence in *Marchantia* (*Marchantiaceae*). *Am. J. Bot.* 101, 348–356. <https://doi.org/10.3732/ajb.1300339>.
14. Attrill, S.T., Mulvey, H., Champion, C., and Dolan, L. (2024). Microtubules and actin filaments direct nuclear movement during the polarisation of *Marchantia* spore cells. *Development* 151, dev202823. <https://doi.org/10.1242/dev.202823>.
15. Beaulieu, C., Libourel, C., Mbadinga Zamar, D.L., Mahboubi, K.E., Hoey, D.J., Keller, J., Girou, C., Clemente, H.S., Diop, I., Amblard, E., et al. (2023). The *Marchantia* pangenome reveals ancient mechanisms of plant adaptation to the environment. Preprint at bioRxiv. <https://doi.org/10.1101/2023.10.27.564390>.
16. Sandler, G., Agrawal, A.F., and Wright, S.I. (2023). Population Genomics of the Facultatively Sexual Liverwort *Marchantia polymorpha*. *Genome Biol. Evol.* 15, evad196. <https://doi.org/10.1093/gbe/evad196>.
17. Doyle, S.R., Søe, M.J., Nejsum, P., Betson, M., Cooper, P.J., Peng, L., Zhu, X.Q., Sanchez, A., Matamoros, G., Sandoval, G.A.F., et al. (2022). Population genomics of ancient and modern *Trichuris trichiura*. *Nat. Commun.* 13, 3888. <https://doi.org/10.1038/s41467-022-31487-x>.
18. Li, M., Reilly, M.P., Rader, D.J., and Wang, L.S. (2010). Correcting population stratification in genetic association studies using a phylogenetic approach. *Bioinformatics* 26, 798–806. <https://doi.org/10.1093/bioinformatics/btq025>.
19. Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat. Rev. Genet.* 10, 639–650. <https://doi.org/10.1038/nrg2611>.
20. Yang, T., Liu, R., Luo, Y., Hu, S., Wang, D., Wang, C., Pandey, M.K., Ge, S., Xu, Q., Li, N., et al. (2022). Improved pea reference genome and pangenome highlight genomic features and evolutionary characteristics. *Nat. Genet.* 54, 1553–1563. <https://doi.org/10.1038/s41588-022-01172-2>.
21. Chen, R., Chen, K., Yao, X., Zhang, X., Yang, Y., Su, X., Lyu, M., Wang, Q., Zhang, G., Wang, M., et al. (2024). Genomic analyses reveal the stepwise domestication and genetic mechanism of curd biogenesis in cauliflower. *Nat. Genet.* 56, 1235–1244. <https://doi.org/10.1038/s41588-024-01744-4>.
22. Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D., and Nordborg, M. (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39, 1151–1155. <https://doi.org/10.1038/ng.2115>.
23. Cerasoli, F., D'Alessandro, P., and Biondi, M. (2022). Worldclim 2.1 versus Worldclim 1.4: Climatic niche and grid resolution affect between-version mismatches in Habitat Suitability Models predictions across Europe. *Ecol. Evol.* 12, e8430. <https://doi.org/10.1002/ece3.8430>.
24. Ellis, N., Smith, S.J., and Pitcher, C.R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93, 156–168. <https://doi.org/10.1890/11-0252.1>.
25. 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. <https://doi.org/10.1093/molbev/msz008>.

26. Lasky, J.R., Josephs, E.B., and Morris, G.P. (2023). Genotype-environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell* 35, 125–138. <https://doi.org/10.1093/plcell/koac267>.
27. Tyrmi, J.S., Vuosku, J., Acosta, J.J., Li, Z., Sterck, L., Cervera, M.T., Savolainen, O., and Pyhäjärvi, T. (2020). Genomics of Clinal Local Adaptation in *Pinus sylvestris* Under Continuous Environmental and Spatial Genetic Setting. *G3 (Bethesda)* 10, 2683–2696. <https://doi.org/10.1534/g3.120.401285>.
28. Pawlik, M.M., and Ficek, D. (2022). Validation of measurements of pine pollen grain concentrations in Baltic Sea waters. *Oceanologia* 64, 233–243. <https://doi.org/10.1016/j.oceano.2021.11.001>.
29. Platt, A., Horton, M., Huang, Y.S., Li, Y., Anastasio, A.E., Mulyati, N.W., Agren, J., Bossdorf, O., Byers, D., Donohue, K., et al. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* 6, e1000843. <https://doi.org/10.1371/journal.pgen.1000843>.
30. Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr., Loudet, O., Colot, V., and Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biol.* 22, 138. <https://doi.org/10.1186/s13059-021-02348-5>.
31. Sang, Y., Long, Z., Dan, X., Feng, J., Shi, T., Jia, C., Zhang, X., Lai, Q., Yang, G., Zhang, H., et al. (2022). Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. *Nat. Commun.* 13, 6541. <https://doi.org/10.1038/s41467-022-34206-8>.
32. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
33. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
34. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
35. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185. <https://doi.org/10.1093/bioinformatics/bts356>.
36. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
37. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
38. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
39. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
40. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
41. Korunes, K.L., and Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21, 1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
42. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>.
43. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* 49, 303–309. <https://doi.org/10.1038/ng.3748>.
44. Montgomery, S.A., Hisanaga, T., Wang, N., Axelsson, E., Akimcheva, S., Sramek, M., Liu, C., and Berger, F. (2022). Polycomb-mediated repression of paternal chromosomes maintains haploid dosage in diploid embryos of *Marchantia*. *eLife* 11, e79258. <https://doi.org/10.7554/eLife.79258>.
45. Casey, C., Köcher, T., Champion, C., Jandrasits, K., Mosiolek, M., Bonnot, C., and Dolan, L. (2023). Reduced coenzyme Q synthesis confers non-target site resistance to the herbicide thaxtomin A. *PLoS Genet.* 19, e1010423. <https://doi.org/10.1371/journal.pgen.1010423>.
46. Delmans, M., Pollak, B., and Haseloff, J. (2017). MarpoDB: An Open Registry for *Marchantia Polymorpha* Genetic Parts. *Plant Cell Physiol.* 58, e5. <https://doi.org/10.1093/pcp/pcw201>.
47. Ryan, W.H., Aida, J., and Krueger-Hadfield, S.A. (2021). The Contribution of Clonality to Population Genetic Structure in the Sea Anemone, *Diadumene lineata*. *J. Hered.* 112, 122–139. <https://doi.org/10.1093/jhered/esaa050>.
48. Shriner, D. (2013). Overview of admixture mapping. *Curr. Protoc. Hum. Genet. Chapter 1, Unit1.23*. <https://doi.org/10.1002/0471142905.hg0123s76>.
49. Shaw, A.J., Piatkowski, B., Duffy, A.M., Aguero, B., Imwattana, K., Nieto-Lugilde, M., Healey, A., Weston, D.J., Patel, M.N., Schmutz, J., et al. (2022). Phylogenomic structure and speciation in an emerging model: the *Sphagnum magellanicum* complex (Bryophyta). *New Phytol.* 236, 1497–1511. <https://doi.org/10.1111/nph.18429>.
50. Sotiropoulos, A.G., Arango-Isaza, E., Ban, T., Barbieri, C., Bourras, S., Cowger, C., Czembor, P.C., Ben-David, R., Dinoor, A., Ellwood, S.R., et al. (2022). Global genomic analyses of wheat powdery mildew reveal association of pathogen spread with historical human migration and trade. *Nat. Commun.* 13, 4315. <https://doi.org/10.1038/s41467-022-31975-0>.
51. Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145, 1219–1228. <https://doi.org/10.1093/genetics/145.4.1219>.
52. Linde, A.M., Eklund, D.M., Cronberg, N., Bowman, J.L., and Lagercrantz, U. (2021). Rates and patterns of molecular evolution in bryophyte genomes, with focus on complex thalloid liverworts, *Marchantiopsida*. *Mol. Phylogenet. Evol.* 165, 107295. <https://doi.org/10.1016/j.ympev.2021.107295>.
53. Aguirre-Liguori, J.A., Ramírez-Barahona, S., and Gaut, B.S. (2021). The evolutionary genomics of species' responses to climate change. *Nat. Ecol. Evol.* 5, 1350–1360. <https://doi.org/10.1038/s41559-021-01526-9>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Samples were used for genome sequencing from 10 European countries, Japan, US and Canada	This study	N/A
Chemicals, peptides, and recombinant proteins		
Sodium dichloroisocyanurate (NaDCC)	Sigma-Aldrich	Cat#218928
Gamborg B5 medium	Duchefa	Cat#G0209
Cefotaxime sodium	Duchefa	Cat#C0111
Sodium Hypochlorite Solution 13% w/v technical grade	PanReac AppliChem	Cat#213322
MES monohydrate	Duchefa	Cat#M1503
Triton X-100	PanReac AppliChem	Cat#A1388
Ethanol absolute	Merck	Cat#1.00983
Plant Agar	Duchefa	Cat#P1001
Critical commercial assays		
Qiagen DNeasy Plant Pro Kit	Qiagen	Cat#69204
NEBNEXT® Ultra™ II FS DNA Library Prep Kit for Illumina	New England Biolabs	Cat#E7805
Deposited data		
WorldClim (version 2.1)	Cerasoli et al. ²³	https://www.worldclim.org/data/worldclim21.html
78 Illumina sequencing reads for whole genome	This study	PRJNA1127023
Software and algorithms		
Trimmomatic (version 0.39)	Bolger et al. ³²	http://www.usadellab.org/cms/?page=trimmomatic
BWA (version 0.7.17-r1188)	Li and Durbin ³³	https://github.com/lh3/bwa
SAMtools (version 1.9)	Li et al. ³⁴	https://github.com/samtools/samtools
RSeQC (version 2.6.4)	Wang et al. ³⁵	https://rseqc.sourceforge.net/
Genome Analysis Toolkit Haplotype Caller (version 4.2.6.1)	McKenna et al. ³⁶	https://gatk.broadinstitute.org/hc/en-us
VCFTools (version 0.1.16)	Danecek et al. ³⁷	https://vcftools.github.io/index.html
Plink (version 2.0)	Chang et al. ³⁸	https://www.cog-genomics.org/plink/2.0/
GCTA (version 1.93.2 beta)	Yang et al. ³⁹	https://yanglab.westlake.edu.cn/software/gcta/
ADMIXTURE (version 1.3)	Alexander et al. ⁴⁰	https://dalexander.github.io/admixture/
Pixy (version 1.2.7)	Korunes et al. ⁴¹	https://pixy.readthedocs.io/en/latest/
Geosphere	Hijmans	https://github.com/rspatial/geosphere
PopLDdecay (version 3.42)	Zhang et al. ⁴²	https://github.com/BGI-shenzhen/PopLDdecay
SMC++ (version 1.15.5)	Terhorst et al. ⁴³	https://github.com/popgenmethods/smcpp
Gradient forest	Eis et al. ²⁴	https://github.com/r-forge/gradientforest
Other		
Microbox containers round	SacO2	O95/114+OD95 #40 NG/NP
Neuhaus Huminsubstrat N3	Klasmann-Deilmann GmbH	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

205 *Marchantia polymorpha* subsp. *ruderaria* accessions were collected from 10 European countries, Japan, US and Canada from August 2021 to June 2023 (Data S1A). Plants were collected from a diversity of environments, including pavements in urban setting some of which were subject to human influence. Plants were also collected from natural riverbanks, waste ground and other natural

settings. Plants were identified to subspecies using a field guide.⁵ Cam-1, Cam-2, Tak-1, and Tak-2 were described previously.^{7,44–46} Tissue samples were precleaned individually with tap water then grown on soil (2:1 ratio of Neuhaus Huminsubstrat N3 from Klasmann-Deilmann GmbH and fine vermiculite) in Sac O2 Microboxes (OV80+OVD80 #40 NG/NP) under standard growth conditions (23°C, continuous white light 50 - 60 µmol m⁻² s⁻¹). The plants were allowed to grow until they developed gemmae. Gemmae were surface sterilized in 0.1 - 1% sodium hypochlorite containing 0.01% Triton-X for 5 minutes, washed in sterile dH2O for 1 minute and finally rinsed three times in sterile dH2O. Gemmae were transferred to sterile plates containing solid 0.5x Gamborg medium (1.5 g/L B5 Gamborg, 0.5 g/L MES hydrate, 1% plant agar, pH adjusted to 5.5) supplemented with Cefotaxime (100mg/l) and grown under standard growth conditions.

Thallus sterilization was performed for *M. polymorpha* subsp. *ruderalis* samples which didn't develop gemmae cups. Thallus explants of 0.4 - 0.8 cm² were sterilized in 1% NaDCC (Sodium dichloroisocyanurate) with 0.01% Triton-X for 1 minute, washed in sterile dH2O for another minute and rinsed three times with sterile dH2O before drying on filter paper. Following this surface sterilization process the thallus explants were moved to solid 0.5x Gamborg medium with Cefotaxime (100mg/l) and grown under standard growth conditions. For samples which didn't remain free of visible contamination, the thallus sterilization procedure was repeated after cleaning the explants in 70% EtOH for about 5 seconds.

METHOD DETAILS

Genomic DNA sequencing

Axenic natural accession lines were grown from gemmae on solid 0.5x Gamborg medium under standard growth conditions. After three weeks of growth, 70-80 mg of plant tissue was harvested and flash frozen in liquid nitrogen. Genomic DNA was extracted using the Qiagen DNeasy Plant Pro Kit (Cat. # 69204) according to the kit protocol with an additional incubation (65°C for 10min) prior to the bead-beating step (using Retsch mill at 30 Hz, 4min) to improve the yield. The gDNA was eluted in EB buffer.

The gDNA concentration was measured using a Qubit 4.0 fluorometer according to the instruction manuals. Genomic DNA samples were sent for sequencing to the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria. DNA Libraries were prepared using NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina and fragment size was determined using a BioLabTech Fragment Analyzer. The DNA was sequenced on Illumina NovaSeq SP / NovaSeq S1 / NovaSeq S4 flowcells using 150 bp paired-end reads.

Read mapping and variant calling

Adapters and low-quality reads ($q < 20$) were removed using Trimmomatic (version 0.39).³² Trimmed reads with a length shorter than 36 bp were discarded. The remaining clean reads were then aligned to the Tak-1 V6.1 genome using BWA-MEM (version 0.7.17-r1188) with default parameters.³³ Alignment quality was assessed using SAMtools (MAPQ = 30)³⁴ and RSeQC (version 2.6.4).³⁵

For variant calling, the Genome Analysis Toolkit Haplotype Caller (version 4.2.6.1)³⁶ was employed. Initially, per-sample GVCf files were generated, followed by joint genotyping with GenotypeGVCFs (using -ploidy 1). A stringent filtering approach was applied based on the removal of tails in the variant distributions, as suggested by a previous study.¹⁷ According to this approach, SNPs with the following criteria were excluded: 'QUAL < 43.4 || DP < 68 || DP > 1295 || MQ < 37.4 || SOR > 3.61 || QD < 4.81 || FS > 13.2 || MQRankSum < -3.72 || ReadPosRankSum < -1.96 || ReadPosRankSum > 2.23'. Subsequently, VCFtools³⁷ was utilized to obtain pure SNPs with the following parameters: --min-alleles 2 --max-alleles 2 --hwe 1e-06 --maf 0.02 --max-missing 0.9.

QUANTIFICATION AND STATISTICAL ANALYSIS

Removal of clones produced by asexual reproduction

Hamming distance, calculated by Plink,³⁸ was used to calculate the genome distance between individuals based on clean SNPs. We then computed the hamming distance among the 209 accessions using all autosomal SNPs. The GD distribution revealed two distinct peaks (Figure S1B): peak A corresponding to a GD of less than 0.01, and peak B corresponding to a GD of 0.24. We also plotted the GD distribution within each population and observed a similar pattern (Figure 1B). Therefore, we concluded that pairs of accessions with GD values less than 0.01 (peak A) were considered to be clones. Following the removal of these clones, a total of 78 accessions were retained for downstream analysis.

Previous research highlighted that clonal populations could cluster tightly, skewing the principal components and obscuring meaningful patterns of genetic variation in populations.⁴⁷ Inclusion of clones in the analysis would therefore misrepresent the true ancestral proportions of the broader population.⁴⁷ Given that *M. polymorpha* reproduces asexually, we evaluated if clones should be removed from our analyses because they would skew, erroneously the conclusions. To evaluate if clones impacted the population structure, analysis using SNPs by PCA was carried out with all individuals 209 – including both unique and clonal individuals. This defined four groups (Figure S10A): Japanese, most European accessions, Sopron (Hungary), and MaG (Austria). However, Sopron (24 individuals) and MaG (24 individuals) populations are clonal; pairwise genome distance within Sopron and MaG populations was less than 0.01. When we carried out the analysis but included single individuals from Sopron and MaG populations, the analysis identified two groups – Europe and Japan and Sopron and MaG accessions were classified within the European group (Figure S10D). This demonstrates that the inclusion of clonal individuals biased the analysis. Consequently, and all clones were represented by one individual in subsequent analyses. Furthermore, the inclusion of the clonal subpopulations – Sopron and MaG – also affected the

admixture result. Sopron and MaG subpopulations retained their distinct ancestries from K=2 to K=9 (Figures S10B and S10C). Thus, failing to remove clonal accessions introduces bias in sub structure analysis. All subsequent analyses were carried out using the 78 genetically unique accessions.

Population genetic analysis

Population structure analyses were performed with two methods, PCA and admixture. PCA analysis was performed on the pure SNP set using GCTA (v1.93.2 beta).³⁹ ADMIXTURE (v 1.3)^{48,40} with cross-validation was used to investigate population structure across all individuals, with the number of clusters (K) being set from 1 to 9. The best K is 2 based on the smallest CV value.

Pi, Dxy and Fst were calculated using pixy⁴¹ with a nonoverlapping 20-kb sliding window across the genome suggested by previous study on haploid bryophyte and fungus.^{49,50} The input to pixy was an all-sites VCF containing both variant and non-variant sites generated by GATK ('GenotypeGVCFs -all-sites') suggested by Pixy manual. The median values across all sliding windows were reported for Pi, Dxy, and Fst based on previous study.^{17,31}

Isolation by distance (IBD) and isolation by environment (IBE) analysis

Isolation by distance (IBD) was analyzed based on relationship between genetic distance and geographical distance. Genetic distance was defined with $F_{ST}/(1-F_{ST})$ based on previous study.^{29,31,51} Geographical distance was calculated using pair-wise Latitude and Longitude information with Geosphere (<https://github.com/rspatial/geosphere>). The correlation between genetic distance and geographical distance was tested using spearman test. Isolation by environment (IBE) was analyzed based on relationship between genetic distance and environment distance. Environment distance was evaluated by Euclidean distance using 19 climate factors downloaded from Worldclim (2.1).²³ The correlation between genetic distance and environment distance was tested using spearman test.

Demographic history analysis

Linkage disequilibrium (LD) was assessed using PopLDdecay (version 3.42)⁴² with default parameters based on previous study.^{20,21} The LD decay was visualized by Plot_MutPop.pl function implemented in PopLDdecay. LD was quantified as the pairwise SNP distance at which the allele frequency correlation (r^2) dropped to half of its maximum value.

To elucidate the historical changes in the effective population size of *Marchantia* populations, demographic analyses were performed employing SMC++ (version 1.15.5).⁴³ The effective population size denotes the number of breeding individuals in an idealized population that would result in the observed genetic diversity. For each population, SNPs were called using VCFtools (-max-missing 1), and subsequently, smc++ vcf2smc was executed for each autosome. Estimated population sizes were computed using smc++ estimate, incorporating a mutation rate of 2.5e-9 mutations per site per generation.⁵² The effective population size per generation was scaled based on an estimated generation time of one generation per year.

Genotype-environment associations

Nineteen climate variables were retrieved from the WorldClim v2.1 database²³, encompassing BIO1 (Annual Mean Temperature), BIO2 (Mean Diurnal Range), BIO3 (Isothermality), BIO4 (Temperature Seasonality), BIO5 (Max Temperature of Warmest Month), BIO6 (Min Temperature of Coldest Month), BIO7 (Temperature Annual Range), BIO8 (Mean Temperature of Wettest Quarter), BIO9 (Mean Temperature of Driest Quarter), BIO10 (Mean Temperature of Warmest Quarter), BIO11 (Mean Temperature of Coldest Quarter), BIO12 (Annual Precipitation), BIO13 (Precipitation of Wettest Month), BIO14 (Precipitation of Driest Month), BIO15 (Precipitation Seasonality), BIO16 (Precipitation of Wettest Quarter), BIO17 (Precipitation of Driest Quarter), BIO18 (Precipitation of Warmest Quarter), BIO19 (Precipitation of Coldest Quarter). Climate-associated SNPs were identified using the LFMM method, implementing a Latent Factor Mixed Model. Considering the number of ancestry clusters inferred with ADMIXTURE, LFMM was executed with two latent factors to address population structure in the genotype data. Adaptive SNPs were selected based on the top 0.0005% adjusted P value of all SNPs from LFMM, resulting in 2191 adaptive SNPs across the 19 climate variables. Significant SNPs were identified using a Bonferroni-based threshold and the Quantile-Quantile (Q-Q) method.

Heritability estimation are based on clean SNPs by GCTA-GREML.³⁹ 1) Genomic Relationship Matrix construction: GCTA starts by creating a genomic relationship matrix (GRM), which quantifies the genetic similarity between pairs of individuals based on genome-wide SNP data. This matrix represents the proportion of shared genetic material between individuals. 2) Phenotypic Variance Decomposition: Using the GRM and the bioclimate factors, GCTA applies the REML algorithm to partition the total phenotypic variance (V_p) into two components: Genetic variance (V_g): The proportion of variance due to additive genetic factors. Environmental variance (V_e): The proportion of variance due to non-genetic factors. 3) Heritability Estimation: The narrow-sense heritability (h^2) is calculated as the ratio of genetic variance to total phenotypic variance (V_g/V_p).

Gradient forest (GF), a nonparametric, machine-learning regression tree method, utilized SNP allele frequencies and climatic data to identify environmental gradients associated with genetic variation and determine how allele frequencies change along those gradients.²⁴ Allelic frequencies for 2191 SNPs were calculated using VCFtools across 22 geographical groups. Weighted importance was assessed based on the 19 Bioclimate variables and allelic frequencies using GF. Cumulative importance of allelic frequency was calculated for the five most significant climate variables, projecting onto the geographical groups. This cumulative importance analysis is followed the protocol based on previous study.⁵³