Does selection occur at the intermediate zone of two insufficiently isolated populations? A whole-genome analysis along an altitudinal gradient

Naofumi Yoshida¹ · Shin-Ichi Morinaga² · Takeshi Wakamiya³ · Yuu Ishii¹ · Shosei Kubota⁴ · Kouki Hikosaka¹

Received: 3 July 2022 / Accepted: 1 December 2022 / Published online: 22 December 2022 © The Author(s) under exclusive licence to The Botanical Society of Japan 2022

Abstract
Adaptive divergence occurs even between insufficiently isolated populations when there is a great difference in environments between their habitats. Individuals present in an intermediate zone of the two divergent populations are expected to have an admixed genetic structure due to gene flow. A selective pressure that acts on the genetically admixed individuals may limit the gene flow and maintain the adaptive divergence. Here, we addressed a question whether selection occurs in the genetically admixed individuals between two divergent populations. Arabidopsis halleri is a perennial montane plant, which has clear phenotypic dimorphisms between highland and lowland habitats in Mt. Ibuki, central Japan. We obtained the whole-genome sequences of Arabidopsis halleri plants along an altitudinal gradient of 359–1,317 m with a high spatial resolution (mean altitudinal interval of 20 m). We found a zone where the highland and lowland genes were mixing (intermediate subpopulation). In the intermediate subpopulation, we identified 5 and 13 genome regions, which included 3 and 8 genes, that had a high frequency of alleles that are accumulated in highland and lowland subpopulations, respectively. In addition, we also found that the frequency of highland alleles of these selected genome regions was smaller in the lowland subpopulation compared with that of the non-selected regions. These results suggest that the selection in the intermediate subpopulation might limit the gene flow and contribute to the adaptive divergence between altitudes. We also identified 7 genome regions that had low heterozygote frequencies in the intermediate subpopulation. We conclude that different types of selection in addition to gene flow occur at the intermediate altitude and shape the genetic structure across altitudes.

Keywords Gene flow · Homozygote · Local adaptation · Phenotypic divergence · Selection · Whole-genome sequences

Introduction
Species having broad distributions face different selections in each habitat and often obtain population-specific polymorphisms on various traits and genes as a result of their local adaptation to distinct environments (Campbell-Staton et al. 2018; Pruisscher et al. 2018). Such adaptive divergence is more likely to occur between geographically isolated populations because geographical distance can provide a strong reproductive barrier and shape large environmental differences (Galloway and Fenster 2000; Kubota et al. 2015). Even with a short distance between the populations, adaptive divergence may occur if there is a great difference of environments between them (Antonovics 2006; Hämmälä and Savolainen 2019; Skelly 2004). The classical expectation has indicated a negative role of gene flow in adaptive divergence; if two populations are not sufficiently isolated and experience strong gene flow, maladapted genetic variation would be introduced from one population to the other and their local adaptation would be impeded (Lenormand 2002). However, if selective pressure in each environment is strong enough, those maladapted genetic variations would...
be removed (Bisschop et al. 2020). Therefore, on a small scale, a genetic structure would be established on a balance between the two competing evolutionary drivers, gene flow and selective pressure, in each habitat (Slatkin 1987).

When phenotypically and/or genetically diverged populations have not been isolated sufficiently from each other, a genetic admixture mainly occurs in an intermediate zone between the two populations (Le Moan et al. 2016; Lipshutz et al. 2017; Ohtani et al. 2013; Puckett et al. 2016; Richardson and Urban 2013). A genetic structure of individuals in the intermediate zone (intermediate subpopulation) would also be established on a balance between gene flow and selective pressure. If there is no selective pressure in the intermediate zone, the genetic structure of the intermediate subpopulation would be a simple admixture of the two divergent populations depending on the gene flow. If there is a strong selective pressure in the intermediate zone, genes related to adaptation would be fixed even under strong gene flow (Hämälä and Savolainen 2019). Presence of specific selection in the intermediate zone may have an important role in population divergence. If a gene is not selected in the intermediate zone, flow of the gene between two diverged populations may occur and impede the adaptive divergence. If a gene is strongly selected in the intermediate zone, its gene flow may be limited due to the selection and the population divergence may be maintained. Fitness of homo/heterozygosity in the intermediate subpopulation may also have an important role in the maintenance of the divergence. If heterozygote individuals have higher fitness than homozygote individuals, the admixture in the intermediate subpopulation is promoted and the adaptive divergence would be disturbed (Westram et al. 2018). Conversely, if the heterozygote individuals have lower fitness, the divergence would be maintained. Therefore, the influences on the adaptive divergence of the two populations may differ depending on the type of selection. However, the selection in the intermediate subpopulation is poorly understood.

According to the above idea, several selection scenarios in the admixed population can be proposed. We consider a genetic structure in an intermediate subpopulations between two divergent populations X and Y. Scenario 1: If one of the alleles (allele x) of a gene is adaptive in population X and the other (allele y) is adaptive in population Y, but the two alleles have similar influence on the fitness in the intermediate zone between the populations X and Y, an accumulation of the two alleles in admixed individuals is influenced mainly by gene flow. Scenario 2: If the allele x is adaptive not only in the population X but also in the intermediate zone whereas the allele y is adaptive only in the population Y, the allele x accumulates at high frequency in the admixed individuals by directional selection overwhelming gene flow. Scenario 3: If an allele z is adaptive only in the intermediate zone and maladaptive in the populations X and Y, and it accumulates only in the individuals located in the intermediate zone. Focusing on fitness of homo/heterozygote, genes belonging to Scenario 1 may further be divided into three groups: fitness of genotypes xx, xy, and yy are similar to each other (Scenario 1a), fitness of homozygote genotypes xx and yy are similar to each other and higher than that of heterozygote genotype xy (Scenario 1b), and fitness of heterozygote genotype xy is higher than that of homogeneity genotypes (Scenario 1c). Therefore, the heterozygote frequency in the intermediate zone is expected to be different between Scenarios; all types of genotypes xx, xy and yy have been randomly mixed under Scenario 1a, but heterozygote frequency may be extremely lower or higher than that expected from the Hardy–Weinberg equilibrium under Scenario 1b or 1c, respectively. The divergence between populations X and Y would be maintained by directional selection under Scenario 2 but be disturbed by balancing selection under Scenario 1c. Investigating the allele and genotype frequencies in the intermediate zone enables us to assess which of these scenarios would be applicable to genes in the intermediate zone and whether the selection in the intermediate zone influence the divergence between populations X and Y.

Altitudinal adaptive divergence is a fascinating process. There are steep environmental gradients along altitudes to study a fine-scale local adaptation. For instance, temperature and a length of growing season regularly decline with increasing altitude (Körner 2007). Many plant species show intraspecific variations along altitudinal gradients. For example, with increasing altitude, *Metrosideros polymorpha* increases leaf mass per area, thereby enhancing tolerance to cold (Cordell et al. 1998). *Fallopia japonica* increases their flavonoid contents (Murai et al. 2015) and decreases their optimal temperature of photosynthesis (Machino et al. 2021). Such a small-scale altitudinal divergence has also been reported from a perennial montane plant, *Arabidopsis halleri* subsp. *gemmifera* (Matsum.) O’Kane & Al-Shhebaz. In Mt. Ibuki, a mountain located in central Japan, *A. halleri* plants are distributed along a broad altitudinal gradient. Although the horizontal distance between the top and bottom populations is small (<3 km), there are various phenotypic differences between highland and lowland ecotypes. Highland ecotypes are characterized by dense trichomes on the leaves, whereas lowland ecotypes have glabrous leaves (Fig. 1a, b). Physiological differentiations have also been reported for the tolerance to UV radiation, the response of biomass allocation to soil nutrient, and the water repellency of leaves (Aryal et al. 2018; Wang Q et al. 2016, 2019). Analyzing the whole-genome sequences, Kubota et al. (2015) found that unidirectional allele frequency shifts along the altitudes in many genes; however, there is a relatively small genetic differentiation between highland and lowland
ecotypes (Ikeda et al. 2010; Kubota et al. 2015). The flowering time of the lowland population is from the end of April to the end of May, whereas that of the highland population is from the middle of May to the middle of June. In intermediate altitudes, plants with scarce trichomes on the leaf surface are often observed, suggesting that these plants have an intermediate phenotype between highland and lowland ecotypes. However, the genetic structure of plants inhabiting the intermediate altitudes has not been studied yet.

In this study, we investigated how the selection shapes the genetic structure in the intermediate zone between highland and lowland subpopulations that are evolutionarily diverged from each other. Specifically, we address the following questions: are there genes that are selected in the intermediate zone? We expect that the selected genes match either of Scenarios 1b, 1c, 2 or 3. If so, what types of genes are selected? To answer these questions, we sampled A. halleri individuals with a very high spatial resolution (every 20 m on average from 359 m to 1,317 m above the sea level, Fig. 1c) and analyzed their whole genome. First, we performed a population structure analysis to identify the areas where the highland and lowland ecotypes inhabit and the area where the genetic admixture of two ecotypes mainly occurs. Second, we investigated allele and genotype frequencies to find genes that match the above-mentioned scenarios. If an allele that has been adaptive to the highland environment has been also favored in the intermediate zone, its frequency in the intermediate zone may be similar to that in the highland but higher than that in the lowland (Scenario 2H in Fig. 2). Vice versa if an allele adaptive to the lowland environment has been favored in the intermediate zone (Scenario 2L). If there is an allele that is adaptive only in the intermediate zone, its frequency may be higher in the intermediate zone than that in the highland and lowland (Scenario 3). If the heterozygote of the two alleles is disadvantageous and eliminated, heterozygote frequency may be extremely low (Scenario 1b). If the heterozygote is advantageous in the intermediate zone, heterozygote of the two alleles may be more frequent than that expected from the Hardy–Weinberg equilibrium (Scenario 1c). Finally, we investigated frequencies of altitudinal-dependent alleles to assess whether the selection in the intermediate subpopulation limits gene flow and contributes to genetic divergence between the highland and lowland subpopulations.
in winter, whereas the lowland habitats are characterized by dark forest floor and relatively mild winter weather (Honjo and Kudoh 2019). *A. halleri* is distributed continuously along altitudes and individuals that have scarce trichomes on the leaf surface (intermediate phenotype between the highland and lowland ecotypes) are often observed in the intermediate altitudes. However, since different morphological types of individuals inhabit together around the intermediate altitudes, it is difficult that identifying a range of altitudes where highland, lowland or intermediate ecotype populations exist only by their morphological traits. We harvested leaf samples from an individual plant at 48 positions along the altitude of Mt. Ibuki in 2007 and 2008 (Table S1). We did not measure phenotypes of the sampled individuals. The lowest and highest sampling sites were at 359 m and 1,317 m above sea level, respectively, and their horizontal distance was approximately 2.8 km (Fig. 1c). To avoid sampling from the same clones, the sampling positions were at least 3 m apart from each other. The mean

**Materials and methods**

**Species and study sites**

*A. halleri* subsp. *geminifera* is a diploid (2n = 16), self-incompatible, and perennial montane plant (Al-shehbaz and O’kane 2002; Kolnik and Marhold 2006). In Japan, this plant is distributed in a wide range of altitudinal and latitudinal gradients. Its leaves are glabrous or sparsely hairy, but the ecotypes in the highland areas in Mt. Ibuki and Mt. Fujiwara in central Japan have dense trichomes on the leaves and stems (Fig. 1a, b). A previous genome-wide association analysis suggested that the two highland ecotypes at Mt. Ibuki and Mt. Fujiwara evolved independently from each other (Kubota et al. 2015), though they have similar morphological characteristics.

We used plants growing in Mt. Ibuki, where the highland habitats are characterized by relatively low vegetation heights, bright environment near the ground, and heavy snow

![Fig. 2 The outline of detection of candidate genome windows matching each selection scenario. Each circle represents genome windows that diverge between the combinations of subpopulations: highland and lowland, highland and intermediate, intermediate and lowland. Under Scenario 2H, genome windows are expected to diverge between highland and lowland and between intermediate and lowland, but not between highland and intermediate. Under Scenario 2L, genome windows are expected to diverge between highland and lowland and between intermediate and intermediate, but not between intermediate and lowland. Genome windows that expected to match Scenario 1 will be contained in the circle of genes that diverge largely between highland and lowland except genes matching Scenarios 2H and 2L. Under Scenario 3, genome windows are expected to diverge between highland and intermediate and between intermediate and lowland, but not between highland and lowland. The area A is the genome windows that significantly diverge between the intermediate and lowland subpopulations but not between other combinations of subpopulations. The area B is the genome windows that significantly diverge between the highland and intermediate subpopulations but not between other combinations of subpopulations.](image)
interval of the altitude and horizontal distance between the sampled plants was 20.4 and 59.6 m, respectively. We also used genome information reported in Kubota et al. (2015), which was obtained from _A. halleri_ plants growing at altitudes of 380, 600, 1,000 and 1,250 m (five plants per site) in 2009 and 2010.

**DNA extraction, individual-based sequencing, and data processing**

We extracted genomic DNA from the dried leaf samples of the collected 48 individuals using DNaseasy Plant Kit (QIA-GEN). Then, we prepared DNA libraries using TruSeq Nano DNA Low Throughput Library Prep Kit (Illumina). We generated reads using Illumina HiSeq X Ten, and obtained 270 GB of data from the 48 samples. These raw read sequences are available in the DNA Data Bank of Japan Sequenced Read Archive under the accession number DRA010696. Furthermore, we added previously posed sequence reads of 20 individuals collected at the altitudes of 380, 600, 1,000 and 1,250 m on Mt. Ibuki (Kubota et al. 2015). We trimmed low-quality reads (more than half of the nucleotides with quality score less than 30) using FASTX-toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit). After the trimming, we mapped the reads of the total 68 individuals to the reference genome of _A. halleri_ (Briskine et al. 2017) for each individual by alignment algorism, BWA-MEM v0.6.2 (Li 2013) with its default parameters. We removed data of 6 individuals due to their low coverage depth and eventually used 62 individuals. We removed PCR duplication by samtools v0.1.8 (Li et al. 2009) rmdup. Using bcftools mpileup (v0.1.8) (Li 2011), we employed SNP calling for the total 62 individuals at once. Using bcftools filter, we treated genotypes of individuals whose coverage depth (DP) were lower than 6 or whose genotype quality (GQ) were lower than 20 as missing data. After SNP calling, we removed INDEL variant loci and trimmed loci whose minor allele frequencies (MAF) were lower than 5% or whose missing rates were higher than 10% by vcftools v0.1.15 (Danecek et al. 2011). Biallelic SNPs were used for our analysis.

**Population structure analysis**

To estimate the population structure of 62 individuals that were distributed continuously along the altitudes at Mt. Ibuki, we employed genetic clustering analysis with ADMIXTURE ver. 1.3.0 (Alexander et al. 2009). We removed loci that were in the linkage disequilibrium with each other by plink –indep-pairwise 50 10 0.1 (plink v1.90b4) (Chang et al. 2015). For each value of _K_ (the number of subpopulations) ranging 1 to 5, we performed independent runs. To determine the optimal number of subpopulations for the 62 individuals, we calculated the cross-validation (CV) error for each _K_ value. The CV error would be minimized when the number of _K_ was best or appropriate for the data. Based on this result, we divided the individuals into highland, intermediate, and lowland subpopulations.

**Detecting altitude-dependent genomic region**

We sought genomic regions accumulated in the intermediate subpopulation as expected in Scenario 2 and 3. We assumed that adaptive divergence along altitudinal gradient causes genetic hitchhiking—genetic variants linked with a selected variant tend to show similar signal of selection. To identify such regions, a window-based approach is appropriate rather than a SNP-based approach. We sought s genome region wherein many genetic variants similarly showed higher _F_ values between populations (_F_ approach). For the whole-genome, we calculated a window-averaged _F_ value within 10 kilobase pair (kbp) windows that included one or more SNP sites. Using vcftools --weir-fst-pop, we calculated the window-averaged _F_ value between three combinations of subpopulations, highland and lowland, highland and intermediate, intermediate and lowland (_F_ and _F_ respectively) in each 10 kbp window. To find windows that were differentiated depending on the altitude, we identified windows that had higher _F_ values (included in the higher 1% quantile) as altitude-dependent windows (ADW). From the ADWs, we sought windows whose _F_ value was higher than 0.343, which was the minimum _F_ value at the higher 1% quantile, and whose _F_ value was included in the lower 70% quantile of the _F_ value, which were considered to match Scenario 2H (Table 1). Similarly, we also sought windows that had high _F_ and low _F_ as well, which were considered to match Scenario 2L (Table 1). For Scenario 3, we selected windows that had lower _F_ values (lower 70% quantile) with higher _F_ and _F_ values than 0.343 (Table 1). We did not identify windows that are consistent with Scenario 1 and that are included in the regions A and B in Fig. 2, because we were interested in genes under the selective pressure in the intermediate subpopulation.

**Heterozygote frequency analysis for altitude-dependent windows**

We sought genome regions showing extremely high and low heterozygote frequency in the intermediate subpopulation as expected in Scenario 1b and 1c, respectively. We assumed that a genotype frequency of a neutral SNP is in the Hardy–Weinberg equilibrium and that of SNP under the selective pressure is deviated from the equilibrium. Therefore, for each bi-allelic SNP, an expected heterozygote frequency
(H_{EXP}) was calculated by a minor allele frequency (p) and the Hardy–Weinberg equilibrium.

\[ H_{EXP} = 2 \times p \times (1 - p) \]

We calculated an average of H_{EXP} and H_{OBS} (observed heterozygote frequency in the intermediate subpopulation) of all SNPs located in each 10 kbp window (W-H_{EXP} and W-H_{OBS}, respectively). In this calculation, we used the same genome windows as those used in the above-mentioned \( F_{ST} \) approach. Then, the differences between the W-HOBS and W-HEXP in the intermediate subpopulation, \( \Delta H \), was calculated for each 10 kbp window.

\[ \Delta H = (W - H_{OBS} - W - H_{EXP}) \]

\( \Delta H \) changes between -1.0 and +1.0 and is higher if the heterozygote frequency is large. We defined the Homozygote- and Heterozygote-accumulated windows that have \( \Delta H \) value smaller and larger than 1% of total windows in the intermediate subpopulation, respectively. Then, the Homozygote- and Heterozygote-accumulated windows that are also the altitude-dependent windows (ADW) were defined as Homozygote- and Heterozygote-accumulated ADWs, respectively. The Homozygote-accumulated ADWs were considered to match Scenario 1b or 2. The Heterozygote-accumulated ADWs were considered to match Scenario 1c.

Homozygote frequency in the intermediate subpopulation is expected to be high in both Scenarios 1b and 2. To find genes that are specifically consistent with Scenario 1b rather than Scenario 2, we further assessed an average of frequencies of highland alleles (W-AF_{High}) in the intermediate subpopulation > 0.66 × HA-ADWs.

| Name | Condition | Number | Scenario |
|------|-----------|--------|----------|
| Altitude-Dependent window | Top 1% \( F_{ST,HL} \) | 198 windows | 1, 2H, 2L |
| S1a window | Altitude-Dependent window — (S2H window, S2L window and HA-ADW) | 173 windows | 1a |
| S2H window | Top 1% \( F_{ST,HL} \) × High \( F_{ST,HL} \spazio{>}{0.343} \) × Bottom 70% \( F_{ST,HL} \) | 1 window | 2H |
| S2L window | Top 1% \( F_{ST,HL} \) × High \( F_{ST,HL} \spazio{>}{0.343} \) × Bottom 70% \( F_{ST,HL} \) | 3 windows | 2L |
| S3 window | High \( F_{ST,HL} \spazio{>}{0.343} \) × High \( F_{ST,HL} \spazio{>}{0.343} \) × Bottom 70% \( F_{ST,HL} \) | 0 window | 3 |
| Homozygote-accumulated Window | Bottom 1% \( \Delta H \) | 198 windows | |
| Heterozygote-accumulated Window | Top 1% \( \Delta H \) | 198 windows | |
| Homozygote-accumulated Altitude-Dependent Window (HA-ADW) | Homozygote-accumulated Window × Altitude-Dependent window | 21 windows | 1b, 2H, 2L |
| Highland-HA-ADW | Window-averaged frequency of highland alleles (W-AF_{High}) in the intermediate subpopulation > 0.66 × HA-ADWs | 4 windows | 2H |
| Lowland-HA-ADW | W-AF_{High} < 0.33 × HA-ADWs | 10 windows | 2L |
| Coexisting-HA-ADW | 0.33 ≤ W-AF_{High} ≤ 0.66 × HA-ADWs | 7 windows | 1b |
| S2H gene | Included in S2H window or Highland-HA-ADWs | 3 genes | 2H |
| S2L gene | Included in S2L windows or Lowland-HA-ADWs | 8 genes | 2L |
| S1b gene | Included in Coexisting-HA-ADWs | 3 genes | 1b |

**Annotation of genes including candidate genomic regions**

We investigated functional genes included in the identified genomic regions by annotation to General Feature Format (GFF) file of the reference genome of *A. thaliana* (Briskine et al. 2017). We also referred to Briskine et al. (2017) that performed alignment against *Arabidopsis thaliana* coding sequences (TAIR10) and investigated functions of the candidate genes with the TAIR database (https://www.
inhabiting altitude below 630 m were nearly occupied by one group (blue in Fig. 3b), whereas those above 1,000 m were nearly occupied by the other (red). We defined individuals whose estimated genetic structure were composed more than 95% by one cluster (red or blue) as highland or lowland individuals and other individuals as admixed individuals between two ecotypes. There were admixed individuals mainly in the intermediate altitudes (Fig. 3b). In the following analyses, based on the result of \( K = 2 \), we defined the following three subpopulations: lowland subpopulation (23 individuals at 359~687 m), which had lowland genotypes; intermediate subpopulation (15 individuals; one at 600 m and 14 at 724~1,000 m), which had mixed genotypes; and highland subpopulation (24 individuals at 1,000~1,317 m), which had highland genotypes. We assigned two admixed individuals locating in lowland altitudes (359 and 423 m) into the lowland subpopulation. The intermediate subpopulation was considered as the admixed subpopulation.

**Assessment of gene flow**

To assess whether the selection in the intermediate subpopulations limits gene flow and contributes to genetic divergence between the highland and lowland subpopulations, we further investigated \( W_{AF_{High}} \) (the averages of frequencies of highland alleles) in the lowland subpopulation and \( W_{AF_{Low}} \) (the averages of frequencies of lowland alleles) in the highland subpopulation for each ADW. If the selection in the intermediate subpopulation limits gene flow, we expect that \( W_{AF_{High}} \) in the lowland subpopulation or \( W_{AF_{Low}} \) in the highland subpopulation of the selected genes is smaller than those of the non-selected genes. We chose ADWs that match each scenario (Scenario 1a, 1b, 1c, 2H, 2L) and calculated their \( W_{AF_{High}} \) in the lowland subpopulation and \( W_{AF_{Low}} \) in the highland subpopulation. We assessed differences in the mean of \( W_{AF_{High}} \) and \( W_{AF_{Low}} \) between ADWs matching each scenario by Wilcoxon rank sum test (R package exactRankTests) (Hothorn and Hornik 2022).

**Results**

**Population structure**

The ADMIXTURE analysis showed that the CV error was minimum when \( K = 2 \) (Fig. 3a). However, the CV errors were similarly low when \( K = 1 \) or 3, suggesting that the population was weakly differentiated. When \( K = 2 \) was adopted, the divergence was clearly found along the altitude; individuals estimated membership in a particular genetic cluster, whereas the horizontal axis represents altitude (m) where each individual was sampled. The upper side of figure represents the subpopulations (white, lowland; hatched, intermediate; black, highland subpopulation).

**arabidopsis.org/index.jsp** Using snpeff-4.5covid19-1, we investigated the number of SNPs included in the identified genome regions and categorized SNPs into synonymous, missense, nonsense or intergenic variants (Cingolani et al. 2012).

**Genes matching scenarios 2 and 3**

We obtained a total of 5,882,976 SNP loci by the whole-genome resequence. Selecting bi-allelic SNPs only, and eliminating SNPs with the low MAF in the 62 individuals (≤5%) or the high missing rate (≥10%), we used 995,710 SNPs in the following analysis. We obtained a total of 19,872 genome windows that included 10,000 loci and one or more SNP sites, and used these windows in the following analyses.

We identified 198 windows whose \( F_{ST_{HL}} \) were extremely high (included in the higher 1% quantile of total 19,872 windows) as the altitude-dependent window (ADW). ADWs included the coding regions of a total of 273 genes (Table S2). The minimum \( F_{ST_{HL}} \) of the ADWs was 0.343, which was used to seek windows that match Scenario 2. There were windows whose \( F_{ST_{HL}} \) were higher than 0.343.
high and $F_{ST_{HL}}$ was low, which were expected to match Scenario 2H (hereafter S2H window) (Fig. 5; Table 1). In the S2H window, the frequencies of highland alleles were diverged clearly between intermediate and lowland (or highland) subpopulations. We detected 1 out of 198 ADWs whose $F_{ST_{HL}}$ and $F_{ST_{IL}}$ were

**Fig. 4** The histograms of the window-averaged $F_{ST}$ values calculated between three combinations of subpopulations, (a) highland and lowland ($F_{ST_{HL}}$), (b) highland and intermediate ($F_{ST_{HI}}$), (c) intermediate and lowland ($F_{ST_{IL}}$) in each 10 kbp genome window. The vertical axis is the number of genome windows.
high in the intermediate and highland subpopulations but low in the lowland subpopulation (Fig. 6a). This S2H window included the coding region of one gene (S2H gene) (Table 2). A total of 32 SNPs in the S2H window located in the coding region of S2H gene and 3 of the 32 SNPs were missense variants (Table S3).

We also detected 3 windows whose window-averaged $F_{ST,HL}$ and $F_{ST,HI}$ were high and $F_{ST,IL}$ was low, which were expected to match Scenario 2L (S2L windows) (Fig. 5; Table 1). In the S2L windows, the frequencies of highland alleles were only high in the highland subpopulations but low in the intermediate and lowland subpopulations (Fig. 6b). One of three S2L windows included the coding region of one gene (S2L gene) (Table 2). A total of 27 SNPs in S2L windows located in the coding region of S2L gene and 3 of the 27 SNPs were missense variants (Table S3).

We could not find a window whose $F_{ST,LI}$ and $F_{ST,HI}$ were high and $F_{ST,IL}$ was low, suggesting that no genes match Scenario 3 (Fig. 5; Table 1). However, we found a SNP whose allele frequency in the intermediate subpopulation was different from that in the highland and lowland subpopulations, though the $F_{ST}$ value of the window that includes the SNP was not statically significant (Fig. S1).

**Genes matching scenarios 1b and 1c**

We calculated the difference between the observed and expected heterozygosity ($\Delta H$) for the whole-genome 19,872 windows. We defined the top 1% (198 windows) and bottom 1% windows (198 windows) as Heterozygote- and Homozygote-accumulated windows, respectively (Fig. 7a). In total, 21 windows of Homozygote-accumulated windows that overlapped with the Altitude-Dependent windows (ADW) were defined as the Homozygote-accumulated Altitude-Dependent windows (HA-ADWs) (Fig. 7b). There were no Heterozygote-accumulated windows that overlapped with the ADWs, suggesting that no genes match Scenario 1c.
Although the heterozygotes existed frequently in the lowland and/or highland subpopulations, the heterozygote frequencies in the intermediate subpopulation (W-H$_{OBS}$) were much lower than the expected frequencies from the Hardy–Weinberg equilibrium (W-H$_{EST}$) (Table S5). The Highland- and Lowland-HA-ADWs showed high and low frequency of highland alleles in the intermediate subpopulation, according to the frequency of highland alleles of each 10 kbp window (W-AF$_{High}$) in the intermediate subpopulation, 21 HA-ADWs were classified into the Highland- (W-AF$_{High} > 0.66$; 4 windows), the Lowland- (W-AF$_{High} < 0.33$; 10 windows), and the Coexisting-HA-ADWs ($-0.33 < W$-AF$_{High} < 0.66$; 7 windows). The Coexisting-HA-ADWs were considered to match Scenario 1b. In the Coexisting-HA-ADWs, although the heterozygotes existed frequently in the lowland and/or highland subpopulations, the heterozygote frequencies in the intermediate subpopulation (W-H$_{OBS}$) were much lower than the expected frequencies from the Hardy–Weinberg equilibrium (W-H$_{EST}$) (Table S5). The Highland- and Lowland-HA-ADWs showed high and low frequency of highland alleles in the intermediate subpopulation,
Function of the pickup genes

We employed functional annotation of the candidate genes detected in the above-mentioned analyses with GO terms, which describe functions of gene products. The S2H genes included, for instance, TP×PH (AT4G39770) implicated in “Trehalose biosynthetic process” (Krasensky et al. 2014) (Table 2). The genetic variants in this gene showed that the homozygote of the lowland allele was observed only under an altitude of 800 m, whereas the homozygote of the highland alleles existed at all altitudes (Fig. 8a). TP×PH included 49 SNPs in its coding region and 6 of the 49 SNPs were mis-sense variants (Table S4). The S2L genes included VPS15 (AT4G29380) assigned to “Pollen development” (Xu et al. respectively, as well as S2H and S2L windows (Table S4). Therefore, the Highland- and Lowland-HA-ADWs were considered to match Scenario 2H and 2L, respectively. There was no overlap between above-mentioned S2H window and the Highland-HA-ADWs or between S2L windows and Lowland-HA-ADWs, respectively (Tables S3, S4).

Three Coexisting-HA-ADWs included the coding regions of 3 genes that were considered as S1b genes. Three Highland-HA-ADWs included the coding regions and 2 genes in the Highland-HA-ADWs were classified into S2H genes (Tables 1 and 2). Seven Lowland-HA-ADWs included the coding regions and 7 genes in the Lowland-HA-ADWs were classified into S2L genes.

### Table 2 The list of candidate gene that were expected to match with each scenario and their GO terms. The columns of “Group” represent the gene group which each gene was classified into. The column of “Included in” represents genome window including each candidate gene

| Group      | AGI code  | Other name | Included in          | GO term                   |
|------------|-----------|------------|----------------------|---------------------------|
| S1b genes  | AT1G74900 | OTP43      | Coexisting-HA-ADW    | RNA splicing              |
|            | AT4G39670 | GLTP       | Coexisting-HA-ADW    | ceramide transport        |
|            | AT5G45840 | MDIS1      | Coexisting-HA-ADW    | pollen tube guidance      |
| S2H genes  | AT1G32375 |            | Highland-HA-ADW      | biological process        |
|            | AT4G05330 | AGD13      | S2H window           | pollen development        |
|            | AT4G39770 | TPPH       | Highland-HA-ADW      | trehalose biosynthetic    |
| S2L genes  | AT1G07570 | APK1A      | S2L window           | protein phosphorylation   |
|            | AT1G64405 |            | Lowland-HA-ADW       | floral organ abscission   |
|            | AT2G13840 |            | Lowland-HA-ADW       | biological process        |
|            | AT2G18328 | RL4        | Lowland-HA-ADW       | biological process        |
|            | AT2G23370 |            | Lowland-HA-ADW       | root morphogenesis        |
|            | AT2G40230 |            | Lowland-HA-ADW       |                            |
|            | AT4G29410 |            | Lowland-HA-ADW       |                            |
|            | AT4G29380 | VPS15      | Lowland-HA-ADW       | pollen development        |

Fig. 7 (a) The histogram of the differences between the observed and expected heterozygote frequency in the intermediate subpopulation ($\Delta H$). The vertical axis is the number of genome windows. The top 1% (198 windows) and bottom 1% windows are defined as Heterozygote- and Homozygote-accumulated windows, respectively. (b) The relationship among the Altitude-Dependent windows (ADW), Heterozygote- and Homozygote-accumulated windows. The overlap between the ADW and Homozygote-accumulated windows are defined as Homozygote-accumulated and Altitude-Dependent windows (HA-ADWs).

Function of the pickup genes

We employed functional annotation of the candidate genes detected in the above-mentioned analyses with GO terms, which describe functions of gene products. The S2H genes included, for instance, $TPPH$ (AT4G39770) implicated in “Trehalose biosynthetic process” (Krasensky et al. 2014) (Table 2). The genetic variants in this gene showed that the homozygote of the lowland allele was observed only under an altitude of 800 m, whereas the homozygote of the highland alleles existed at all altitudes (Fig. 8a). $TPPH$ included 49 SNPs in its coding region and 6 of the 49 SNPs were mis-sense variants (Table S4). The S2L genes included $VPS15$ (AT4G29380) assigned to “Pollen development” (Xu et al.
MDIS1 (AT5G45840), one of three genes found as the S1b genes, included 21 SNPs in its coding region whereas the other two genes, OTP43 (AT1G74900) and GLTP (AT4G39670) did not include any SNPs (Table S4). MDIS1 was implicated in “Pollen tube guidance” (Wang T et al. 2011) (Table 2). The genetic variants in VPS15 showed that the homozygote of the lowland allele existed at all altitudes, whereas the homozygote of the highland allele existed only above an altitude of 900 m (Fig. 8b). Only one out of 30 SNPs in VPS15 was a missense variant (Table S4).
Four out of 21 SNPs in *MDISI* were missense variants (Table S4). The genetic variants in *MDISI* showed that there was no low heterozygote individual along the altitude between 800 m and 1,200 m (Fig. 8c).

The variants represented in Fig. 8a, b, c were missense variants. In the present study, we could not find any nonsense variant in the focused genomic regions. GO terms that were assigned to S1b, S2H, and S2L genes differed from each other in most cases. Exceptionally, the GO term “Pollen development” was assigned to both S2H and S2L genes (Table 2).

**Assessment of gene flow**

The frequencies of highland alleles of each 10 kbp window (W-AF<sub>High</sub>) in the lowland subpopulation were significantly smaller in windows matching Scenario 2L (3 S2L windows and 10 Lowland-HAAD windows) than those matching Scenario 1a (173 S1a windows) (*p* value < 0.01) (Table S6). On the other hand, there was no significant difference in the frequencies of lowland alleles of each 10 kbp window (W-AF<sub>Low</sub>) in the highland subpopulation between the windows matching Scenario 2H (1 S2H window and 4 Highland-HAAD windows) and S1a windows. Neither W-AF<sub>High</sub> in the lowland subpopulation nor W-AF<sub>Low</sub> in the highland subpopulation differed significantly between the genome regions matching Scenario 1b (7 Coexisting-HAAD windows) and S1a windows.

**Discussion**

**Is selection at the intermediate altitudes potentially a driver of adaptive divergence between the highland and lowland subpopulations?**

We analyzed the individual-based whole-genome resequencing data of *A. halleri* sampled along the altitude between 359 m and 1,317 m with very high spatial resolution. This dataset enabled us to find the zone where the highland and lowland genes were mixing (Fig. 3b) and to investigate how selective pressure shapes the genetic structure in the intermediate subpopulation. The result of ADMIXTURE (Fig. 3b) suggests that the direct crossing between the lowland and highland subpopulations might not occur frequently, probably because of less overlap of the flowering time between the highland and lowland subpopulations. Therefore, the gene flow between the highland and lowland subpopulations is expected to occur through the intermediate subpopulation, which has relatively similar phenology to both highland and lowland subpopulations.

The genotype or allele frequencies in the intermediate subpopulation greatly differed among genome regions, suggesting that there are multiple patterns of selection rather than a single pattern of selection in the intermediate subpopulation of *A. halleri*. We found 5 altitude-dependent windows (ADWs) (1 S2H window and 4 Highland-HA-ADWs) whose allele frequency in the intermediate subpopulation is similar to that in the highland but different from that in the lowland subpopulation, which were consistent with Scenario 2H (a gene that is adaptive to highland is also adaptive to the intermediate zone) (Fig. 6a). Similarly, we also found a total of 13 ADWs (3 S2L windows and 10 Lowland-HA-ADWs), which were consistent with Scenario 2L (a gene that is adaptive to lowland is also adaptive to the intermediate zone) (Fig. 6b). We also found 7 ADWs whose heterozygote had relatively low frequency in the intermediate subpopulation but whose allele frequencies of H and L alleles were similar to each other in the intermediate subpopulation (Coexisting-HA-ADWs), which were consistent with Scenario 1b (both alleles are neutral in the fitness in the intermediate subpopulation, but their heterozygotes were negatively selected).

A part of the results supported our hypothesis that a selective pressure at the intermediate zone could constrain gene flow between two populations and contribute to maintenance of adaptive divergence. In the windows matching Scenario 2L, the frequencies of H alleles (W-AF<sub>High</sub>) in the lowland subpopulation were significantly smaller than those in the windows matching Scenario 1a (Table S6). This result strongly suggests that the introduction of H alleles to lower altitudes has been prevented by the selection on the S2L genes. Such trends were also observed in functional genes. For example, in genetic variant in *VPS15*, one of the S2L genes, the homozygotes of H alleles were rarely observed in the lowland and intermediate subpopulations, whereas the homozygotes of L alleles existed at all altitudes (Fig. 8b). This pattern would also suggest that the introduction of H alleles to lower altitudes is prevented by the selection eliminating the homozygote of H alleles at the intermediate subpopulation, whereas L alleles were not eliminated at all altitudes.

On the other hand, in the windows matching Scenario 2H, the frequencies of L alleles (W-AF<sub>Low</sub>) in the highland subpopulation did not significantly differ from those in S1a windows (Table S6). We consider that the difference of W-AF<sub>Low</sub> between scenarios might be hindered due to the very small sample size of windows matching Scenario 2H (5 windows). In fact, some functional genes showed very low frequencies of L alleles in the highland subpopulation. For example, in *TPPH*, one of the S2H genes, we found that the homozygotes of L alleles were not observed below an altitude of 800 m, whereas the homozygotes of H alleles existed at all altitudes (Fig. 8a). This pattern would
suggest that the introduction of L alleles to higher altitudes is prevented by the selection eliminating the homozygote of L alleles at the intermediate subpopulation, whereas H alleles were not eliminated at all altitudes. These results imply that the selection on the S2H genes constrains the introduction of L alleles to higher altitudes. Our results of genes matching Scenario 2H or 2L suggest that the selection to a particular genome region at the intermediate altitudes potentially acted as a driver of adaptive divergence between the highland and lowland subpopulations, though the constraint for gene flow by divergent selection might be weak. Theoretical studies have suggested that a divergent selection due to environmental differences between habitats can reduce an effect of gene flow and lead to a adaptive divergence in particular environments during speciation process, especially ecological speciation (reviewed in Nosil et al. 2005). Although supporting examples for this hypothesis have been reported in previous empirical studies that focused on the populations experiencing migration from neighboring populations (Nosil et al. 2005), such studies did not consider the intermediate zone located between the two different environments. Our result is the first report suggesting that gene flow is influenced by the selection in the intermediate zone.

We also inspected another scenario (Scenario 1b) that a selection eliminating heterozygote genotypes would restrict an admixture of polymorphism and act as a potential driver of adaptive divergence. However, there was no significant differences in $W$-$AF_{High}$ in the lowland subpopulation and $W$-$AF_{Low}$ in the highland subpopulation between the windows matching Scenario 1b and Scenario 1a (Table S6). In genetic variants in $MDISI$, one of the S1b genes, the heterozygotes were rarely found in the intermediate subpopulation but found frequently in the highland and lowland subpopulations (Fig. 8c). In all the Coexisting-HA-ADWs, the observed heterozygote frequencies in the highland and/or lowland subpopulations were higher than that in the intermediate subpopulations (Table S5). These patterns may suggest that the heterozygotes are eliminated only in the intermediate altitudes and the selective pressure has not prevented gene flow between the lowland and highland subpopulations. Therefore, our result did not support that the selection eliminating the heterozygotes contributes to the adaptive divergence between altitudes.

Altitudinal variation in functional genes

We identified 198 windows, which had extremely high $F_{ST\, HS}$ values, as the altitude-dependent window (ADW). ADWs included the coding regions of a total of 273 genes (Table S2), suggesting that these genes are related to adaptive divergence between the highland and lowland subpopulations. In our previous study, Kubota et al. (2015) also identified 153 genes altitudinal diverged genes for the same population in Mt. Ibuki. However, only 11 genes overlapped between the two studies. This was mainly caused by the differences in the sample sizes and the processes to detect altitudinal-dependent genetic variations. First, sample size was smaller in Kubota et al. (2015), which sampled 5 individuals from each of 4 altitudes (380, 600, 1,000 and 1,250 m). According to the classification into the subpopulations in the present study, 5 and 7 individuals used in Kubota et al. (2015) were included in the lowland and highland subpopulations, respectively. Second, the present study detected alleles that are different between the highland and lowland subpopulations, whereas Kubota et al. (2015) detected alleles whose frequency was correlated with altitude. Third, Kubota et al. (2015) used the genome of lowland ecotypes as reference genome and sought genetic variants that accumulated derived alleles at high altitude to identify genes that are adaptive rather in highland than in lowland, whereas the present study used the genome of outgroup (Tada mine, Hyogo, Japan; Briskine et al. 2017) as the reference genome and sought target genes.

We found a total of 3 genes that were contained in the S2H window or Highland-HA-ADWs. $TPPH$, one of the S2H genes, is implicated in the trehalose metabolism (Kraensky et al. 2014), which may contribute to tolerance to several environmental stresses (e.g. cold, drought and salinity stresses) (Avonce et al. 2004, Ge et al. 2008, Kosar et al. 2019). We also found a total of 8 genes that were contained in the S2L windows or Lowland-HA-ADWs. $VPS15$, one of the S2L genes, is considered to relate with the pollen development (Xu et al. 2011). We also found three S1b genes that were consistent with Scenario 1b. $MDISI$ (AT5G45840), one of the S1b genes, is implicated in pollen tube guidance (Wang T et al. 2016) (Table 2). We consider that these genes are related with adaptation in the intermediate subpopulation.

In this study, 29 out of 160 SNPs in S2H genes, 65 out of 153 SNPs in S2L genes, and 7 out of 21 SNPs in S1b genes were synonymous variants (Tables S3, S4), which may be neutral in terms of selection (Moutinho et al. 2020). We found 28 missense variants located in 3 S2H genes and 24 missense variants located in 6 S2L genes. We also found 4 missense variants located in one S1b gene. In genes containing missense variants, such as $TPPH$ and $VPS15$, amino-acids substitution and change in the protein structure might occur. Mutation occurring in UTR or splice region would also regulate the gene expression and relate with the phenotypic divergence (Guan et al. 2017; Mayr 2017). Therefore, S2H and S2L genes containing non-synonymous variants might contribute to the adaptive divergence between the
highland and lowland populations. Although 4 out of 19 S2L genes and 2 out of 3 S1b genes did not include any genetic variant in their coding regions, there were a total of 201 intergenic variants near the coding regions. Mutation occurring in intergenic regions would also change expression levels of genes that locate near the mutation sites (Ochiai et al. 2014). Also, undiscovered small coding genes might be hidden in the intergenic regions. Although many small coding genes (30–100 amino acids) have played functional roles for morphogenesis and environmental changes, current gene annotation methods tend to miss such small genes (Hanada et al. 2013; Takahashi et al. 2019). Therefore, SNPs in the intergenic regions might be the target of selection. However, we did not have phenotypic data of the sampled individuals in the present study. To assess whether the genetic variants in the above-mentioned candidate genes actually relate with the phenotype, fitness and adaptation in the intermediate subpopulation, assembling an empirical evidence based on phenotypic data and improving the amount of genome data will be required in the future.

Conclusion

In this study, we investigated the genetic structure of an Arabidopsis halleri population along an altitudinal gradient and assessed whether the selection occurs in the intermediate subpopulation between insufficiently diverged populations using the whole-genome sequence of individuals sampled with a very high spatial resolution. We identified the genetically admixed zone between the highland and lowland ecotypes in the intermediate altitude. We found genome regions that suggest the existence of different types of selective pressure in the admixed zone. In a total of 25 genome regions, the distribution of alleles was separated above or below the intermediate zone, suggesting that the selection might prevent the admixture of highland and lowland alleles by the gene flow over the intermediate altitudes. This selection might limit the gene flow and contribute to the maintenance of adaptive divergence between altitudes. We also found 7 genome regions whose heterozygote frequencies in the intermediate subpopulation were much lower than the theoretical expectation. These suggest that the heterozygotes were negatively selected in the intermediate altitudes. However, this selection might not limit the gene flow because the heterozygote individuals existed frequently in the lowland and highland altitudes. We suggest that selective pressures at the intermediate zone would partly contribute to the divergence between the highland and lowland populations.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s10265-022-01429-1.

Acknowledgements We thank Masakado Kawata, Wataru Iwasaki, Takashi Makino and Kousuke Hanada for support in this research project. We also thank anonymous reviewers for constructive comments.

Author contributions K. H., S.M. and N.Y. designed the research, S. M. performed material sampling, N. Y. performed the experiments and data analysis, T. W. and Y. I. and S. M. contributed the experiment and analysis, and N.Y. wrote the manuscript with comments from other authors.

Funding This study was partly supported by JST CREST Grant Number JPMJCR11B3 and JST SPRING Grant Number JPMJSP2114 and KAKENHI (No. 25291095, 17H03727, 20H03317).

Data availability The raw read sequences in the present study are available in the DNA Data Bank of Japan Sequenced Read Archive under the accession number DRA010696.

Code availability Not applicable.

Declarations

Conflicts of interest/competing interests Not applicable.

Ethics approval Not applicable.

Consent to participate Approve.

Consent for publication Approve.

References

Al-Shehbaz IA, O’Kane SL Jr (2002) Taxonomy and phylogeny of Arabidopsis (Brassicaceae). The Arabidopsis Book 1
Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664
Antonovics J (2006) Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. Heredity 97:33–37
Aryal B, Shimohara W, Honjo MN, Kadoh H (2018) Genetic differentiation in cauline-leaf-specific wettability of a rosette-forming perennial Arabidopsis from two contrasting montane habitats. Ann Bot 121:1351–1360
Avonce N, Leyman B, Mascorro-Gallardo JO, Van Dijck P, Thevelein JM, Iturriaga G (2004) The Arabidopsis trehalose-6-P synthase AtTPS1 gene is a regulator of glucose, abscisic acid and stress signaling. Plant Physiol 136:3649–3659
Bisschop G, Setter D, Rajafarić M, Baird SJE, Lohse K (2020) The impact of global selection on local adaptation and reproductive isolation. Phil Trans R Soc B 375:20190531
Briskine RV, Paape T, Shimizu-Insuaggi R, Nishiyama T, Akama S, Sese J, Shimizu KK (2017) Genome assembly and annotation of Arabidopsis halleri, a model for heavy metal hyperaccumulation and evolutionary ecology. Mol Ecol Resour 17:1025–1036
Campbell-Staton SC, Bare A, Losos JB, Edwards SV, Cheviron ZA (2018) Physiological and regulatory underpinnings of geographic variation in reptilian cold tolerance across a latitudinal cline. Mol Ecol 27:2243–2255
Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK - rising to the challenge of larger and richer datasets. GigaScience 4:7
Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (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

Cordell S, Goldstein G, Mueller-Dombois D, Webb D, Vitousek PM (1998) Physiological and morphological variation in Metrodorphos polymorpha, a dominant hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. Oecologia 113:188–196

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Galloway LF, Fenster CB (2000) Population differentiation in an annual legume: local adaptation. Evolution 54:1173–1181

Ge LF, Chao DY, Shi M, Zhu MZ, Gao JP, Lin HX (2008) Overexpression of the trehalose-6-phosphate phosphatase gene OsTPP1 confers stress tolerance in rice and results in the activation of stress responsive genes. Planta 228:191–201

Guan H, Dong Y, Liu C et al (2017) A splice site mutation in<br>shrunken1-m causes the shrunken 1 mutant phenotype in maize. Plant Growth Regul 83 429–439

Hämälä T, Savolainen O (2019) Genomic patterns of local adaptation under Gene Flow in Arabidopsis lyrata. Mol Biol Evol 36:2557–2571

Hanada K, Higuchi-Takeuchi M, Okamoto M, Yashizumi T, Shimizu M, Nakaminami K, Nishi R, Ohashi C, Iida K, Tanaka M, Hori Y, Kawashima M, Matsu M, Toyoda T, Shimozaki K, Seki M, Matsu M (2013) Small open reading frames associated with morphogenesis are hidden in plant genomes. Proc Natl Acad Sci USA 110:2395–2400

Honjo MN, Kudoh H (2019) Arabidopsis halleri: a perennial model system for studying population differentiation and local adaptation. AoB Plants 11:plz076

Hothorn T, Hornik K (2022) Exact distributions for rank and permutation tests. CRAN. https://CRAN.R-project.org/package=exactRankTests

Ikeda H, Setoguchi H, Morinaga S (2010) Genomic structure of Lowland and Highland Ecotypes of Arabidopsis halleri subsp. gemmifera (Brassicaceae) on mt. Ibuski Acta Phytotax Geobot 61:21–26

Kolnik M, Marhold K (2006) Distribution, chromosome numbers and nomenclature conspect of Arabidopsis halleri (Brassicaceae) in the Carpathians. Biologia 61:41–50

Körner C (2007) The use of ‘altitude’ in ecological research. Trends Ecol Evol 22:569–574

Kosar F, Akram NA, Sadiq M et al (2019) Trehalose: a key organic osmolyte effectively involved in plant abiotic stress tolerance. J Plant Growth Regul 38:606–618

Krasensky J, Broyart C, Rabanal FA, Jonak C (2014) The redox-sensitive chloroplast trehalose-6-phosphate phosphatase AtTPPD regulates salt stress tolerance. Antioxid Redox Signal 21:1289–1304

Kubota S, Iwasaki T, Hanada K, Nagano AJ, Fujiyama A, Toyoda A et al (2015) A genome scan for genes underlying microgeographic variation under Gene Flow in Arabidopsis lyrata. Mol Ecol 110:2395–2400

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv: 1303.3997

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Lipshutz SE, Overcast IA, Hickerson MJ, Brumfield RT, Derryberry EP (2017) Behavioural response to song and genetic divergence in two subspecies of white-crowned sparrows (Zonotrichia leucocephala). Mol Ecol 26:3011–3027

Machino S, Nagano S, Hikosaka K (2021) The latitudinal and altitudinal variations in the biochemical mechanisms of temperature dependence of photosynthesis within Fallopia japonica. Environ Exp Bot, in press. https://doi.org/10.1016/j.envexpbot.2020.104248

Mayr C (2017) Regulation by 3’-Untranslated regions. Annu Rev Genet 51:171–194

Moutinho AF, Bataillon T, Dutheil JY (2020) Variation of the adaptive substitution rate between species and within genomes. Evol Ecol 34:315–338

Muray Y, Setoguchi H, Kitajima J, Iwashina T (2015) Altitudinal variation of flavonoid content in the leaves of Fallopia japonica and the needles of Larix kaempferi on Mt. Fuji. Nat Prod Commun 10:407–411

Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719

Ochiai H, Miyamoto T, Kanai A, Hosoha K, Sakuma T, Kudo Y, Asami K, Ogawa A, Watanabe A, Kaji T, Yamamoto T, Matsura S (2014) TALEN-mediated single-base-pair editing identification of an intergenic mutation upstream of BUSB as causative of PSC (MVA) syndrome. Proc Natl Acad Sci USA 111:1461–1466

Ohtani M, Kondo T, Tani N, Ueno S, Lee LS, Ng KKS, Muhammad N, Finkeldey R, Na’iem M, Indrioko S, Kamiya K, Harada K, Diway B, Khoo E, Kawamura K, Tsumura Y (2013) Nuclear and chloroplast DNA phylogeography reveals pleistocene divergence and subsequent secondary contact of two genetic lineages of the tropical rainforest tree species Shorea leprosula (Dipterocarpaceae) in South-East Asia. Mol Ecol 22:2264–2279

Pruisscher P, Nylin S, Gotthard K, Wheat CW (2018) Genetic variation underlying local adaptation of diapause induction along acline in a butterfly. Mol Ecol 27:3613–3626

Puckett EE, Park J, Combs M, Blum MJ, Bryant JE, Caccone A, Costa F, Deinum EE, Esther A, Himsworth CG, Keightley PD, Ko A, Lundkvist Å, McElhinney LM, Morand S, Robins J, Russell J, Strand TM, Suarez O, Yon L, Munshi-South J (2016) Global population divergence and admixture of the brown rat (Rattus norvegicus). Proc R Soc B 283:20161762

Richardson JL, Urban MC (2013) Strong selection barriers explain microgeographic adaptation in wild salamander populations. Evolution 67:1729–1740

Skelly DK (2004) Microgeographic countergradient variation in the wood frog, Rana sylvatica. Evolution 58:160–165

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 15:787–792

Takahashi F, Hanada K, Kondo T, Shinozaki K (2019) Hormone-like peptides and small coding genes in plant stress signaling and development. Curr Opin Plant Biol 31:88–95

Wang Q, Nagano S, Ozaki H, Morinaga S, Hitendra J, Hikosaka K (2016) Functional differentiation in UV-B-induced DNA damage and growth inhibition between highland and lowland ecotypes of two Arabidopsis species. Environ Exp Bot 131:110–119

Wang Q, Daumal M, Nagano S et al (2019) Plasticity of functional traits and optimality of biomass allocation in elevational ecotypes of Arabidopsis halleri grown at different soil nutrient availabilities. J Plant Res 132:237–249
Wang T, Liang L, Xue Y et al (2016) A receptor heteromer mediates the male perception of female attractants in plants. Nature 531:241–244

Westram AM, Rafajlović M, Chaube P, Faria R, Larsson T, Panova M, Ravinet M, Blomberg A, Mehlig B, Johannesson K, Butlin R (2018) Clines on the seashore: the genomic architecture underlying rapid divergence in the face of gene flow. Evol Lett 2:297–309

Xu N, Gao XQ, Zhao XY et al (2011) Arabidopsis AtVPS15 is essential for pollen development and germination through modulating phosphatidylinositol 3-phosphate formation. Plant Mol Biol 77:251

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.