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

Naofumi Yoshida (n.yoshida830@gmail.com)
Tohoku University Graduate School of Life Sciences: Tohoku Daigaku Daigakuin Seimei Kagaku Kenkyuka https://orcid.org/0000-0003-0215-9608
Shin-Ichi Morinaga
Teikyo University of Science: Teikyo Kagaku Daigaku
Takeshi Wakamiya
Hiroshima University: Hiroshima Daigaku
Yuu Ishii
Tohoku University Graduate School of Life Sciences: Tohoku Daigaku Daigakuin Seimei Kagaku Kenkyuka
Shosei Kubota
FASMAC Co. Ltd.
Kouki Hikosaka
Tohoku University Graduate School of Life Sciences: Tohoku Daigaku Daigakuin Seimei Kagaku Kenkyuka

Research Article

Keywords: Local adaptation, selection, gene flow, phenotypic divergence, whole-genome sequences, homozygote

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-447152/v2

Version of Record: A version of this preprint was published at Journal of Plant Research on December 22nd, 2022. See the published version at https://doi.org/10.1007/s10265-022-01429-1.

Abstract

Adaptive divergence would occur even between the insufficiently isolated populations when there is a great difference in the environments between their habitats. The individuals present in the 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 promote the adaptive divergence. Here, we addressed a question whether the selection occurs in the genetically admixed individuals between the divergent populations and assessed its effects on the population divergence. We obtained the whole-genome sequences of a perennial montane plant, *Arabidopsis halleri*, which has clear phenotypic dimorphisms between altitudes, along an altitudinal gradient of 359–1,317 m with a high spatial resolution (mean altitudinal interval of 20 m). We found the zone where the highland and lowland genes were mixing (intermediate subpopulation). We sought genetic variants that had been selected in the intermediate subpopulation. In the intermediate subpopulation, we identified 9 and 22 genetic regions, which included 7 and 19 genes, had a high frequency of alleles that are accumulated in highland and lowland subpopulations, respectively, suggesting that these genes have been selected in the admixed zone. This selection might limit the gene flow and contribute to the adaptive divergence along the altitudes. We also identified a genetic region that contains different alleles from those in both highland and lowland subpopulations, though the variants were not included in encoded genes. We also identified 7 genetic regions had low heterozygote frequencies in the intermediate subpopulation. We could not find any genetic regions whose heterozygote frequency was higher than theoretical expectation. We conclude that different types of selection in addition to gene flow occur at the intermediate altitude and shape the genetic structure across altitudes.

Introduction

Species having broad distributions would 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 (Pruisscher et al. 2018; Campbell-Staton et al. 2018). Such adaptive divergence is more likely to occur between the 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 (Skelly et al. 2004; Antonovics et al. 2006; Hämälä and Savolainen 2019). The classical expectation has indicated a negative role of gene flow in adaptive divergence; if the populations are not sufficiently isolated and experience strong gene flow, maladapted genetic variation would be introduced from one population to another 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, the genetic structure would be established on a balance between the two competing evolutional powers, gene flow and selective pressure, in each habitat (Slatkin 1987).

Genetic exchange between two populations may mainly occur through individuals located in the intermediate zone of the two populations. The selection and gene flow around such admixed individuals in the intermediate zone may play important roles for the divergence of the two populations. Following scenarios may be considered on the genetic structure in the admixed individuals of the two 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, there is no selection on the gene and the accumulation of the two alleles in the 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 in the admixed individuals. Scenario 3: If an allele *z* is adaptive only in the intermediate zone and maladaptive in the populations *X* and *Y*, and it accumulates in the individuals located in the intermediate zone. Focusing on fitness of heterozygote, genes belonging to the Scenario 1 may further be divided into three groups: fitness of genotypes *xx*, *xy*, and *yy* is similar to each other and they are randomly mixed (Scenario 1a), fitnesses of homogeneity genotypes *xx* and *yy* are similar to each other and higher than that of heterogeneity genotype *xy* (Scenario 1b), and fitness of heterogeneity genotype *xy* is higher than those of homogeneity genotypes (Scenario 1c).

Such selections in the intermediate zone may influence gene flow between the two populations. Under the Scenario 1c, the selection favoring heterozygotes would promote the admixture of polymorphism and prevent adaptive divergence of the two populations. Conversely, under the Scenario 1b and 2, the selection eliminating the heterogeneity genotypes or favoring one of the alleles would overwhelm gene flow, which restricts the admixture of polymorphism and acts as a potential driver of population adaptive divergence. Under the Scenario 3, the selection favoring peculiar genes to the environment at the intermediate zone would not relate with adaptive divergence between two edge populations, but would take an important role in the evolution in the intermediate population.

A number of studies have demonstrated that the genetic admixture occurred between phenotypically and/or genetically diverged populations (Ohtani et al. 2013; Richardson and Urban 2013; Le Moan et al. 2016; Puckett et al. 2016; Lipshutz et al. 2017), and individuals in such admixed populations often have intermediate phenotypes of the two populations (Stacy et al. 2016; Hendrick et al. 2016; Linnen et al. 2013). These results would be consistent with Scenario 1a. In contrast, other studies showed that the selection maintains lower recombination rates or genetic diversity on the particular genetic regions in the hybrid populations even under strong gene flow, which would be consistent with Scenario 1b or 2 (Comeault et al. 2015; Hämälä and Savolainen 2019). Heterosis, defined as a vigorous growth in the hybrid offspring of genetically distant individuals relative to their homozygous parents, has been shown in some species, which would be consistent with Scenario 1c (Facon et al. 2005; Li et al. 2018). However, which of these scenarios is applicable to genes in the intermediate population is poorly understood. In particular, the degree of heterozygote frequency and kind of genes that are selected in the intermediate zone have not been studied yet. Furthermore, a high spatial resolution sampling of genetic data is necessary to identify the zone where the admixture of gene occurs, but previous studies have not conducted such sampling.

The altitudinal adaptive divergence is one of the fascinating materials. There are steep environmental gradients along the altitudes to study the fine-scale local adaptation. For instance, temperature and the length of growing season regularly decline with increasing elevation (Körner 2007). Many plant species show intraspecific variations along the altitudinal gradient. 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 the 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*. In Mt. Ibuki, a mountain located in Central Japan, *A. halleri* plants are distributed along the broad altitudinal gradient. Although the horizonal 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 (Wang Q et al. 2016, 2019; Aryal et al. 2018). Analyzing the whole-genome sequences, Kubota et al. (2015) found unidirectional allele frequency shifts along the altitudes in many genes; however, there is a relatively small genetic differentiation between highland hairy ecotype and lowland normal ecotypes (Ikeda et al. 2010; Kubota et al. 2015). The flowering time of lowland population is from the end of April to the end of May, whereas that of 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 individuals has not been studied yet.

In this study, we addressed a question how the selection and gene flow shape the genetic structure in the admixed individuals between two divergent populations. To answer this question, 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 identified the areas where the genetic admixture of highland and lowland ecotypes mainly occurs. Second, we investigated allele frequency to find genes that have been selected in the admixed zone according to the above-mentioned scenarios. If an allele that has been adaptive to the highland environment has been also favored in the admixed zone, its frequency in the admixed 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 admixed zone (Scenario 2L). If there is an allele that is adaptive only in the admixed zone, its frequency may be higher in the intermediate zone than that in the highland and lowland (Scenario 3). If heterozygote of the two alleles has been disadvantageous and eliminated, heterozygote frequency may be extremely low (Scenario 1b). If heterozygote has been advantageous in admixed zone, heterozygote of the two alleles may be more frequent than the expected from the Hardy–Weinberg equilibrium (Scenario 1c). Finally, we investigated function of genes that have been selected in the admixed zone according to these scenarios.

Materials And Methods

Species and Study Sites

Arabidopsis halleri subsp. *gemmifera* 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 in winter, whereas the lowland habitats are characterized by dark forest floor and relatively mild winter weather (Honjo and Kudo 2019). We harvested the leaf samples from an individual plant at 48 positions along the altitude of Mt. Ibuki in 2007 and 2008 (Table S1). The lowest and highest sampling sites were at 359 m and 1,317 m above the sea level, respectively, and their horizontal distance was approximately 2.8 km (Fig. 1c). To avoid the sampling of same clones, the sampling positions were at least 3 m apart from each other. The mean interval of the altitude and horizontal distance between the sampled plants was 20.4 m 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 altitude 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 the genomic DNA from the dried leaf samples of collected 48 individuals using the DNeasy Plant Kit (QIAGEN). Thereafter, we prepared the DNA libraries using the TruSeq Nano DNA Low Throughput Library Prep Kit (Illumina). We generated reads using the 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 the 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 total 68 individuals to the reference genome of *A. halleri* (Briskine et al. 2017) for each individual by the alignment algorism, BWA-MEM v0.6.2 (Li 2013) with the 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 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). For the population structure analysis, we removed the 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 the genomic regions accumulated in the intermediate subpopulation in relation to the Scenario 2 and 3 using an F_{ST} approach. For whole-genome, we calculated the window-averaged F_{ST} values within 10 kilobase pair (kbp) windows that included one or more SNP sites. Using vcftools –weir-fst-pop, we calculated the window-averaged F_{ST} between three combinations of subpopulations, highland and lowland, highland and intermediate, intermediate and lowland (F_{ST_HL}, F_{ST_HI} and F_{ST_IL}, respectively) in each 10 kbp window. To find the windows that were differentiated depending on the altitude, we identified windows that had higher F_{ST_HL} values (included in the higher 1% quantile) as the altitude-dependent windows (ADW). Then, among ADWs, windows that had extremely high F_{ST_IL} (included in the higher 1%) and low F_{ST_HI} (included in the lower 70%) were identified, which were considered to match the Scenario 2H (Table 1). Similarly, among ADWs, windows that had extremely high F_{ST_HI} and low F_{ST_IL} were also identified, which were considered to match the Scenario 2L (Table 1). We also selected windows that had lower F_{ST_HL} values. Among them, windows that had higher F_{ST_IL} and F_{ST_HI} were identified, which were considered to match the Scenario 3 (Table 1). We did not identify windows that are consistent with the Scenario 1 and that are included in the regions A and B in Fig. 2, because we were interested in the genes under the selective pressure in the intermediate subpopulation.

Heterozygote Frequency Analysis for Altitude-Dependent windows

We sought genome regions showing extremely high or low heterozygote frequency in the intermediate subpopulation in relation to the Scenario 1b and 1c. We assumed that the 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 the average of H_{EXP} and H_{OBS} (observed heterozygote frequency in the intermediate subpopulation) of all SNPs located in each 10 kbp window (WA_H_{EXP} and WA_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 WA_H_{OBS} and WA_H_{EXP} in the intermediate subpopulation, ΔH, was calculated for each 10 kbp window.

$\Delta H = (WA\_H_{OBS} - WA\_H_{EXP})$

ΔH 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 ΔH 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 window (ADW) were defined as Homozygote- and Heterozygote-accumulated ADWs, respectively. The Homozygote-accumulated ADWs were considered to match the Scenario 1b or 2. The Heterozygote-accumulated ADWs were considered to match the 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 the Scenario 1b rather than Scenario 2, we further assessed the average of the allele frequency of the highland alleles (W_AF_{High}) in the intermediate subpopulation for each 10 kbp window. W_AF_{High} would be large in the Scenario 2H and smaller in the Scenario 2L, whereas in the Scenario 1b, it would be neither extremely large nor small. Therefore, we categorized the Homozygote-accumulated ADWs (HA-ADWs) into three groups and defined as the Highland- ($W\_AF_{High} > 0.66$), the Lowland- ($W\_AF_{High} < 0.33$), and the Coexisting-HA-ADWs ($0.33 < W\_AF_{High} < 0.66$). The Coexisting-HA-ADWs were considered to match the Scenario 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. halleri* (Briskine et al. 2017). We also referred to Briskine et al. (2017) that performed alignment against *Arabidopsis thaliana* coding sequences (TAIR10) and investigated the functions of candidate genes with TAIR database (https://www.arabidopsis.org/index.jsp). Using snpeff-4.5covid19-1, we investigated the number of SNPs included in the identified genome regions and categorized into synonymous, missense, nonsense or intergenic variants (Cingolani et al. 2012).

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 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). There were admixed individuals between their two ecotypes in the intermediate altitudes (Fig. 3b). When K = 3 was adopted, the third group was found mainly in the plants investigated by Kubota et al. (2015) (Fig. S1a). We suspected that the result of K = 3 reflected an artificial error due to a variation in the data size or quality. However, further analysis revealed that the proportion of third group did not necessary correlate with the number of missing loci of each individual reflecting shortage of coverage depth of them (Fig. S1b). Therefore, we considered that this grouping might reflect not from the differences of data sizes but from the minute differences of population structure among sampling years. In following analyses, based on the result of K = 2, we defined 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 1,051~1,317 m), which had highland genotypes. The intermediate subpopulation was considered as the admixed subpopulation.

Genes matching the Scenarios 2 and 3

We obtained total 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 ($\leq$ 5%) or the high missing rate ($\geq$ 10%), we used 995,710 SNPs in the following analysis. We obtained total 19,872 genome windows that included 10,000 loci and one or more SNP sites and used these windows in following analyses.

We identified 198 windows whose F_{ST_HL} was extremely high (included in the higher 1% quantile of total 19,872 windows) as the altitude-dependent window (ADW). We detected 11 out of 198 ADWs whose F_{ST_HL} and F_{ST_IL} were high and F_{ST_HI} was low, which were expected to match the Scenario 2H (hereafter S2H windows) (Fig. 4, Table 1). In S2H windows, the allele frequencies of highland alleles were high in the intermediate and highland subpopulations but low in the lowland subpopulation (Fig. 5a). Seven S2H windows included the coding regions of 6 genes (S2H genes) (Table 2). Total 198 SNPs in the S2H windows located in the coding regions of S2H genes and 54 of the 198 SNPs were missense variants (Table S2).

We also detected 24 windows whose window-averaged F_{ST_HL} and F_{ST_HI} were high and F_{ST_IL} was low, which were expected to match the Scenario 2L (S2L windows) (Fig. 4, Table 1). In S2L windows, the allele frequencies of highland alleles were only high in the highland subpopulations but low in the intermediate and lowland subpopulation (Fig. 5b). Fourteen S2L windows included the coding regions of 14 genes (S2L genes) (Table 2). Total 326 SNPs in S2L windows located in the coding regions of S2L genes and 57 of the 326 SNPs were missense variants (Table S2).

We also detected only one window whose F_{ST_LI} and F_{ST_HI} were high and F_{ST_HL} was low, which were expected to match the Scenario 3 (hereafter S3 window) (Fig. 4, Table 1). One gene was included in the S3 window (S3 gene) (Table S2), but this gene did not include any missense variant.

Genes matching the Scenarios 1b and 1c

We calculated the difference between the observed and expected heterozygosity (ΔH) 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. 6a). In total, 21 windows of Homozygote-accumulated windows that overlapped with Altitude-Dependent windows (ADW) were defined as the Homozygote-accumulated Altitude-Dependent windows (HA-ADWs) (Fig. 6b). There were no Heterozygote-accumulated windows that overlapped with ADW, suggesting that no genes match the Scenario 1c.

According to the index W_AF_{High}, which represents the frequency of highland alleles of each 10kb window in the intermediate subpopulation, 21 HA-ADWs which defined above 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 the 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 (WA_H_{OBS}) were much lower than the expected frequencies from the Hardy–Weinberg equilibrium (WA_H_{EST}) (Table S4). The Highland- and Lowland-HA-ADWs showed high and low frequency of highland alleles in the intermediate subpopulation, respectively, as well as S2H and S2L windows (Table S3). Therefore, the Highland- and Lowland-HA-ADWs were considered to match the Scenario 2H and 2L, respectively. Above-mentioned two S2H windows and two S2L windows also included in the Highland- and Lowland-HA-ADWs, respectively (Tables 2, S2, S3).

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, 2). Seven Lowland-HA-ADWs included the coding regions and 7 genes in the Lowland-HA-ADWs were classified into S2L genes.

Function of the pickup genes

We employed the functional annotation of candidate genes detected in the above-mentioned analyses with GO terms, which describe the 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 lowland allele was observed only under altitude of 800 m, whereas the homozygote of highland alleles existed in the whole altitudes (Fig. 7a). *TPPH* included 49 SNPs in its coding region and 6 of the 49 SNPs were missense variants (Table S3). The S2L genes included *VPS15* (AT4G29380) assigned to “Pollen development” (Xu et al. 2011) (Table

2). The genetic variants in *VPS15* showed that the homozygote of lowland allele existed in the whole altitudes, whereas the homozygote of the highland allele existed only above altitude of 900m (Fig. 7b). Only one out of 30 SNPs in *VPS15* was missense variant (Table S3).

XPO2 (AT2G46520), only one gene found as the S3 gene, was implicated in “cellular protein localization” (TAIR) (Table 2). This gene included only one genetic variant, and one allele was frequent only in the intermediate subpopulation (Fig. 7c).

MDIS1 (AT5G45840), one of three genes found as the S1b genes, included 21 SNPs in its coding region whereas other two genes, *OTP43* (AT1G74900) and *GLTP* (AT4G39670) did not include any SNPs (Table S3). *MDIS1* was implicated in “Pollen tube guidance” (Wang T et al. 2016) (Table 2). Four out of 21 SNPs in *MDIS1* were missense variants (Table S3). The genetic variants in *MDIS1* showed that there was no low heterozygote individual along the altitude between 800 m and 1,200 m (Fig. 7d).

The variants represented in Fig. 7a, b, d were missense variants whereas the variant in Fig. 7c was synonymous variant. In present study, we could not find any nonsense variant in the focused genomic regions. GO terms that were assigned to S1b, S2H, S2L and S3 genes differed from each other in most cases. Exceptionally, the GO term “Pollen development” was assigned to both S2H and S2L genes (Table 2).

Discussion

Genetic structure in the intermediate subpopulation

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

Our results suggest that there are different types of selection in the intermediate subpopulation of *Arabidopsis halleri*. We found total 9 genome windows (S2H windows and Highland-HA-ADWs) whose allele frequency in the intermediate subpopulation is similar to that in highland but different from that in lowland subpopulation, which were consistent with the Scenario 2H (a gene that is adaptive to highland is also adaptive to the intermediate zone) (Fig. 5a). Similarly, we also found total 22 windows (S2L windows and Lowland-HA-ADWs), which were consistent with the Scenario 2L (a gene that is adaptive to lowland is also adaptive to the intermediate zone) (Fig. 5b). We also found one window (S3 window) in which one of alleles had high frequency only in the intermediate subpopulation (Fig. 4, Table 1). This

result was consistent with the Scenario 3 that individuals with peculiar alleles to the intermediate zone are adaptive. We also investigated Altitude-Dependent windows whose heterozygote had relatively low frequency in the intermediate subpopulation and found that the 7 windows had similar frequencies of H and L alleles in the intermediate subpopulation (Coexisting-HA-AD Ws), which were consistent with the Scenario 1b. These alleles were expected to be neutral in the fitness in the intermediate subpopulation but their heterozygotes were negatively selected. In contrast, we did not find any Altitude-Dependent windows that had higher heterozygote frequency than the theoretical expectation (Fig. 6b). This result rejects the Scenario 1c that heterozygous individuals are advantageous (heterosis).

We found total 7 genes that were contained in S2H windows or Highland-HA-ADWs. We considered these genes to be related with adaptation in the intermediate subpopulation. *TPPH*, one of the S2H genes, is implicated in trehalose metabolism in plants (Krasensky et al. 2014). Many experimental studies supported that trehalose metabolism in plants would contribute to environmental stress tolerance (e.g. cold, drought and salinity stresses) (Avonce et al. 2004, Ge et al. 2008, Krasensky et al. 2014, Kosar et al. 2019). As cold stress is known to be greater in higher altitude (Körner 2007), the variants in *TPPH* might relate with adaptation to the environmental stress. Our observations in *TPPH* may be consistent with the fact that the environment of *A. halleri* habitat changes with altitude; the lowland subpopulation was covered by forest canopies, whereas the highland and intermediate subpopulations were exposed to direct light and wind (hemisphere photographs are presented in Wang et al. 2019). Individuals in the highland and intermediate subpopulations are expected to face cold and drought stresses seriously than those in the lowland subpopulation.

We also found total 19 genes that were contained in S2L windows or Lowland-HA-ADWs. *VPS15* is considered to relate with the pollen development (Xu et al. 2011). AT1G64405 is considered to relate with the floral organ abscission process (González-Carranza et al. 2012). As the flowering times of highland and lowland ecotypes of *Arabidopsis halleri* differ from each other, these S2L genes might relate with the phenological variation along altitudes. Although we do not have information on the phenology of the intermediate subpopulation, our results suggest that the variation related in reproductive process may be similar between the lowland and intermediate zones, whereas that in highland differs from others.

We found only one S3 gene, *XPO2*, which were consistent with the Scenario 3. *XPO2* is implicated in cellular protein localization (TAIR). *XPO2* included one SNP whose allele frequencies were low in highland and lowland subpopulations but increased only in the intermediate subpopulation (Fig. 7c). This phenomenon suggests that this allele is adaptive only in the intermediate altitude.

We found three S1b genes which were consistent with the Scenario 1b. *MDIS1* (AT5G45840), one of S1b genes, is implicated in “Pollen tube guidance” (Wang T et al. 2016) (Table 2). *MDIS1* included 21 SNPs whose heterozygote frequencies were relatively small in the intermediate subpopulation (Fig. 7d). This phenomenon suggests that heterozygote individual is maladaptive in the intermediate altitude.

In this study, 58 out of 247 SNPs in S2H genes, 125 out of 450 SNPs in S2L genes, one SNP in the S3 gene and 7 out of 21 SNPs in S1b genes were synonymous variants (Tables S2 and S3), which may be

neutral in terms of selection (Moutinho et al. 2020). We found that 60 missense variants located in 7 S2H genes and 77 missense variants located in 11 S2L genes. We also found that 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 (Mayr 2017, Guan et al. 2017). Therefore, S2H, S2L and S3 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 total 201 intergenic variants near the coding regions. The mutation in the intergenic regions would also change the expression levels of genes that locate near the mutation sites (Ochiai et al. 2014). Also, undiscovered small coding genes might be hided in the intergenic regions. Although much 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. We need to further assess whether the candidate genes actually relate with altitudinal adaptation by improving the amount of data and/or using reverse genetics.

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

A part of the results supported our hypothesis that the selective pressure at the intermediate zone could constrain the gene flow between the two populations and act as a potential driver of population adaptive divergence. In genetic variant in *TPPH*, one of the S2H genes, the homozygotes of L alleles were not observed below an altitude of 800 m, whereas the homozygotes of H alleles existed in the whole altitudes (Fig. 7a). 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 in the whole altitudes. In contrast, 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 in the whole altitudes (Fig. 7b). This pattern would 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 in the whole altitudes. Theoretical studies have suggested that the divergent selection due to environmental differences between the habitats can reduce the effect of gene flow and lead to the population adaptive divergence in particular environments during the speciation process, especially ecological speciation (reviewed in Nosil et al. 2005). Although supporting examples for this hypothesis were reported in the previous empirical studies that focused on the populations experiencing migration from neighboring populations (Nosil et al. 2005), there seems no empirical study that focused on an intermediate zone located between the two different environments. Our results of *TPPH* and *VPS15* would suggest that 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 these would also suggest that constraint for gene flow by divergent selection might be weak.

We also inspected the other scenario (Scenario 1b) that the selection eliminating the heterogeneity genotypes would restrict the admixture of polymorphism and act as a potential driver of population adaptive divergence. In genetic variants in *MDIS1*, one of the S1b genes, the heterozygotes were rarely found in the intermediate subpopulation but found frequently in the highland and lowland subpopulations (Fig. 7d). 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 S4). These patterns may suggest that the heterozygotes are eliminated only in the intermediate altitudes and the selective pressure has not prevented gene flow between altitudes. Therefore, our result did not support that the selection eliminating the heterozygotes contributes the adaptive divergence between altitudes.

Finally, we found one gene, *XPO2*, which was considered to match the Scenario 3. In the genetic variant in *XPO2*, one allele was frequent only in the intermediate subpopulation (Fig. 7c). This pattern may suggest the existence of selective pressure only in the intermediate altitude. The selection favoring peculiar genes to the environment at the intermediate zone would take an important role in the evolution in the intermediate population, but would not relate with adaptive divergence between altitudes.

Conclusions

In this study, we investigated a fine-scale local adaptation along the altitudinal gradient using the whole-genome sequence of individuals sampled with a very high spatial resolution. We identified the genetically admixed zone between highland and lowland ecotypes in the intermediate altitudes. We found genomic regions that suggest the existence of different types of selective pressure in the admixed zone locating in the intermediate altitudes. In 26 genes included in the candidate genomic regions, the distribution of genotypes was separated above and below the intermediate zone, suggesting that the selection might prevent the admixture of highland and lowland genotypes by the gene flow over the intermediate altitudes. This selection might limit the gene flow and contribute to the adaptive divergence along the altitudes. We also found 7 genomic regions whose heterozygote frequencies in the intermediate subpopulation were much lower than 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. On the other hand, we could not find genome windows wherein the heterozygotes were positively selected. We also identified a genetic region that contains different alleles from those in both highland and lowland subpopulations, though the variants were not included in encoded genes. This result suggests that the selection favoring peculiar genes to the environment at the intermediate zone, but this selection might not contribute to the adaptive divergence along the altitudes. We suggest that selective pressures at the intermediate zone would partly contribute to the divergence between the highland and lowland populations.

Declarations

Acknowledgments

We would like to thank Masakado Kawata, Wataru Iwasaki, Takashi Makino and Kousuke Hanada for support in this research project. This study was partly supported by JST CREST Grant Number JPMJCR11B3, JST SPRING Grant Number JPMJSP2114 and KAKENHI (No. 25291095, 17H03727, 20H03317).

Funding:

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

Conflicts of interest / Competing interests

Not applicable

Ethics approval

Not applicable

Consent to participate

approve

Consent for publication

approve

Availability of data and material

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

Code availability

Not applicable

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.

References

1. Al-Shehbaz I A, O'Kane SL Jr (2002) Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). The Arabidopsis Book 1
2. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655–1664
3. Antonovics J (2006) Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. Heredity 97: 33–37
4. Aryal B, Shinohara W, Honjo MN, Kudoh 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
5. 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
6. Bisschop G, Setter D, Rafajlović M, Baird SJE, Lohse K (2020) The impact of global selection on local adaptation and reproductive isolation. Phil Trans R Soc B 375: 20190531
7. Briskine RV, Paape T, Shimizu-Inatsugi 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
8. 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
9. 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
10. 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 w^{1118}; *iso-2; iso-3*. Fly 6(2):80-92
11. Comeault AA, Flaxman SM, Riesch R, Curran E, Soria-Carrasco V, Gompert Z, Farkas TE, Muschick M, Parchman TL, Schwander T, Slate J, Nosil P (2015) Selection on a Genetic Polymorphism Counteracts Ecological Speciation in a Stick Insect.Curr Biol25: 1975–1981
12. Cordell S, Goldstein G, Mueller-Dombois D, Webb D, Vitousek PM (1998) Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. Oecologia 113:188–196
13. Danecek P, Auton A, Abecasis G, Albers C A, Banks E, DePristo M A, 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
14. Facon B, Jarne P, Pointier JP, David P (2005) Hybridization and invasiveness in the freshwater snail *Melanoides tuberculata*: hybrid vigour is more important than increase in genetic variance. J Evol Biol18: 524–535

15. Galloway LF, Fenster CB (2000) Population differentiation in an annual legume: local adaptation. Evolution 54: 1173–1181
16. 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
17. González-Carranza ZH, Shahid AA, Zhang L, Liu Y, Ninsuwan U, Roberts JA (2012) A Novel Approach to Dissect the Abscission Process in Arabidopsis. Plant Physiology 160(3): 1342–1356
18. Guan H, Dong Y, Liu C et al. (2017) A splice site mutation in *shrunken1-m* causes the *shrunken 1* mutant phenotype in maize. Plant Growth Regul 83 429–439
19. Hämälä T, Savolainen O (2019) Genomic Patterns of Local Adaptation under Gene Flow in *Arabidopsis lyrate*. Mol Biol Evol 36: 2557–2571
20. Hanada K, Higuchi-Takeuchi M, Okamoto M, Yoshizumi T, Shimizu M, Nakaminami K, Nishi R, Ohashi C, Iida K, Tanaka M, Horii Y, Kawashima M, Matsui K, Toyoda T, Shinozaki K, Seki M, Matsui M (2013) Small open reading frames associated with morphogenesis are hidden in plant genomes. PNAS 110: 2395-2400
21. Hendrick MF, Finseth FR, Mathiasson ME, Palmer KA, Broder EM, Breigenzer P, Fishman L (2016) The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*. Mol Ecol 25: 5647–5662
22. Honjo MN, Kudoh H (2019) *Arabidopsis halleri*: a perennial model system for studying population differentiation and local adaptation. AoB Plants 11, plz076
23. Ikeda H, Setoguchi H, Morinaga S (2010) Genomic Structure of Lowland and Highland Ecotypes of *Arabidopsis halleri* subsp. *gemmifera* (Brassicaceae) on Mt. Ibuki. Acta Phytotax Geobot 61: 21–26
24. Kolnik M, Marhold K (2006). Distribution, chromosome numbers and nomenclature conspect of *Arabidopsis halleri* (Brassicaceae) in the Carpathians. Biologia 61: 41–50
25. Körner C (2007) The use of 'altitude' in ecological research. Trends Ecol Evol22: 569–574
26. 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
27. 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
28. Kubota S, Iwasaki T, Hanada K, Nagano AJ, Fujiyama A, Toyoda A, et al. (2015) A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild *Arabidopsis* Species. PLoS Genet11: e1005361
29. Le Moan A, Gagnaire PA, Bonhomme F. (2016) Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Mol Ecol 25: 3187–3202
30. Lenormand T (2002) Gene flow and the limits to natural selection. Trends Ecol Evol 17: 183–189

31. 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
32. 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
33. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv: 1303.3997
34. Li Y, Stift M, van Kleunen M (2018) Admixture increases performance of an invasive plant beyond first-generation heterosis. J Ecol 106: 1595–1606
35. Linnen CR, Poh YP, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE (2013) Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science 339: 1312–1316
36. 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 leucophrys*). Mol Ecol 26: 3011–3027
37. 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. doi.org/10.1016/j.envexpbot.2020.104248
38. Mayr C (2017) Regulation by 3'-Untranslated Regions. Annu Rev Genet 51: 171-194
39. Moutinho AF, Bataillon T, Dutheil JY (2020) Variation of the adaptive substitution rate between species and within genomes. Evol Ecol 34: 315–338
40. Murai 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
41. Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59: 705–719
42. Ochiai H, Miyamoto T, Kanai A, Hosoba K, Sakuma T, Kudo Y, Asami K, Ogawa A, Watanabe A, Kajii T, Yamamoto T, Matsuura S (2014) TALEN-mediated single-base-pair editing identification of an intergenic mutation upstream of *BUB1B* as causative of PCS (MVA) syndrome. PNAS 111 (4): 1461-1466
43. 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
44. Pruisscher P, Nylin S, Gotthard K, Wheat CW (2018) Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. Mol Ecol 27: 3613–3626
45. 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

46. Richardson JL, Urban MC (2013) Strong selection barriers explain microgeographic adaptation in wild salamander populations. Evolution 67: 1729–1740
47. Skelly DK (2004) Microgeographic countergradient variation in the wood frog, *Rana sylvatica*. Evolution 58: 160–165
48. Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 15: 787–792
49. Stacy E, Johansen J, Sakishima T et al. (2016) Genetic analysis of an ephemeral intraspecific hybrid zone in the hypervariable tree, *Metrosideros polymorpha*, on Hawai'i Island. Heredity, 117: 173–183
50. 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 51: 88-95
51. Wang Q, Nagano S, Ozaki H, Morinaga S, Hidema 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
52. 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
53. 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
54. 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

Tables

Table 1. The definition for searching genomic regions matching each scenario of selection and the numbers of detected genes. The values of F_{ST} are calculated for three combinations of two subpopulations; between highland and lowland, between intermediate and lowland, and between highland and intermediate (F_{ST_HL}, F_{ST_IL} and F_{ST_HI}, respectively).

Name	Condition	Number	Scenario
Altitude-Dependent window	Top 1% F_{ST_HL}	198 windows	1, 2H, 2L
S2H window	Top 1% F_{ST_HL} × Top 1% F_{ST_IL} × Bottom 70% F_{ST_HI}	11 windows	2H
S2L window	Top 1% F_{ST_HL} × Top 1% F_{ST_HI} × Bottom 70% F_{ST_IL}	24 windows	2L
S3 window	Top 1% F_{ST_HI} × Top 1% F_{ST_IL} × Bottom 70% F_{ST_HL}	1 window	3
Homozygote-accumulated Window	Bottom 1% ΔH	198 windows	
Heterozygote-accumulated Window	Top 1% ΔH	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}) > 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 windows or Highland-HA-ADWs	7 genes	2H
S2L gene	Included in S2L windows or Lowland-HA-ADWs	19 genes	2L
S1b gene	Included in Coexisting-HA-ADWs	3 genes	1b

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		both of S2H and Highland-HA-ADW	biological process
	AT1G65470	FAS1	S2H window	nucleosome
	AT3G23130	SUP	S2H window	specification of floral organ identity
	AT4G03540	CASPL1C1	S2H window	tissue development
	AT4G05330	AGD13	S2H window	pollen development
	AT4G12710		S2H window	hormone-mediated signaling pathway
	AT4G39770	TPPH	Highland-HA-ADW	trehalose biosynthetic process
S2L genes	AT1G05940	CAT9	S2L window	amino acid transport
	AT1G07550		S2L window	
	AT1G07570	APK1A	S2L window	protein phosphorylation
	AT1G64405		both of S2L and Lowland-HA-ADW	floral organ abscission
	AT2G13840		Lowland-HA-ADW	
	AT2G18328	RL4	both of S2L and Lowland-HA-ADW	biological process
	AT2G23370		Lowland-HA-ADW	biological process
	AT2G25170	PKL	S2L window	root development
	AT2G26230	UOX	S2L window	urate catabolic process
	AT2G40230		Lowland-HA-ADW	root morphogenesis
	AT2G42370		S2L window	
	AT3G29390	RIK	S2L window	negative regulation of gene expression
	AT3G29575	AFP3	S2L window	response to abscisic acid
	AT3G60020	SK5	S2L window	protein ubiquitination

	AT4G28250	EXPB3	S2L window	sexual reproduction
	AT4G29410		Lowland-HA-ADW	
	AT4G29380	VPS15	Lowland-HA-ADW	pollen development
	AT5G10170	MIPS3	S2L window	inositol biosynthetic process
	AT5G38220		S2L window	biological process
S3 gene	AT2G46520	XPO2	S3 window	cellular protein localization

Figures

(a)

(b)

(c)

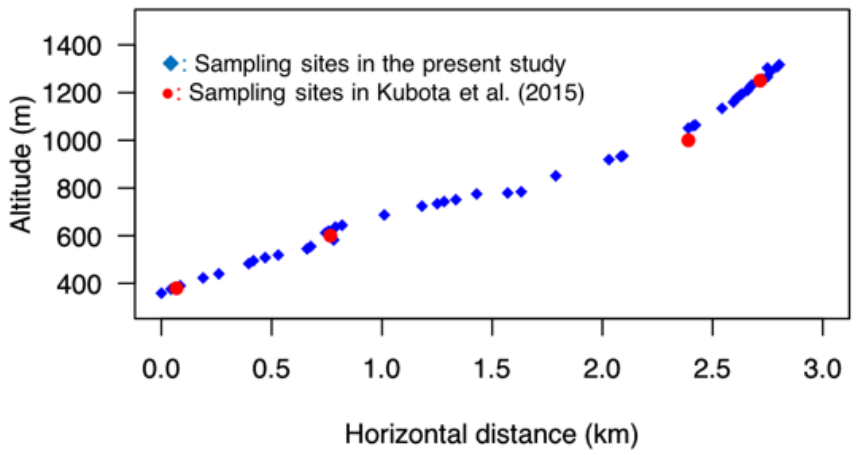

Figure 1

The leaves of (a) highland (hairy) and (b) lowland ecotypes (glabrous) of *Arabidopsis halleri* that were grown from the seeds in the same growth chamber. (c) The sampling point of individuals analyzed in this study. The horizontal axis represents the horizonal distance (km) from the bottom sampling point (359 m above sea level) to each point.

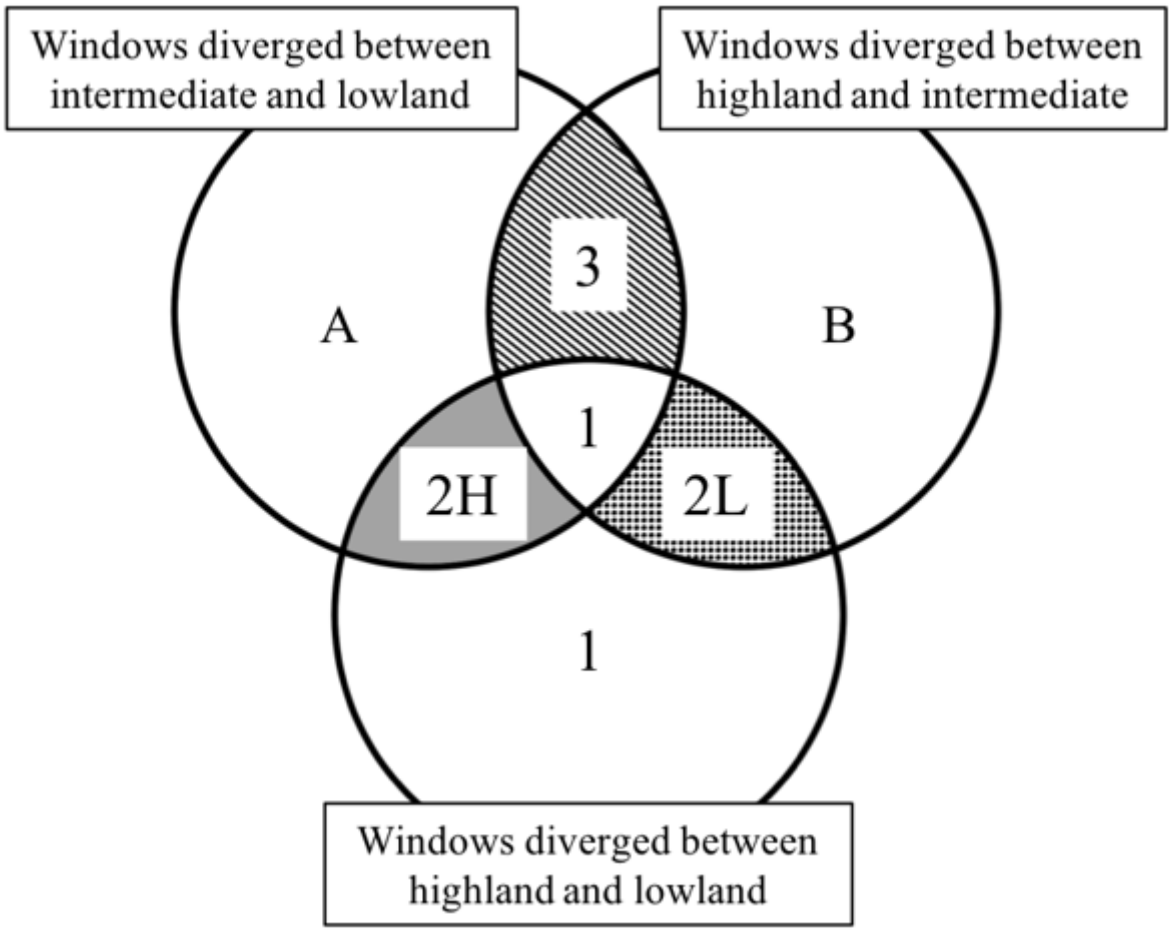

Figure 2

the outline of detection of candidate genome windows matching each selective scenario. Each circle represents genome windows that diverge between the combinations of subpopulations; Highland and Lowland, Highland and Intermediate, Intermediate and Lowland. Under the Scenario 2H, genomic windows are expected to diverge between Highland and Lowland and between Intermediate and Lowland, but not between Highland and Intermediate. Under the Scenario 2L, genome windows are expected to diverge between Highland and Lowland and between Highland and Intermediate, but not between Intermediate and Lowland. Genome windows that expected to match the Scenario 1 will be contained in the circle of genes that diverge largely between Highland and Lowland except genes matching the Scenarios 2H and 2L. Under the 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.

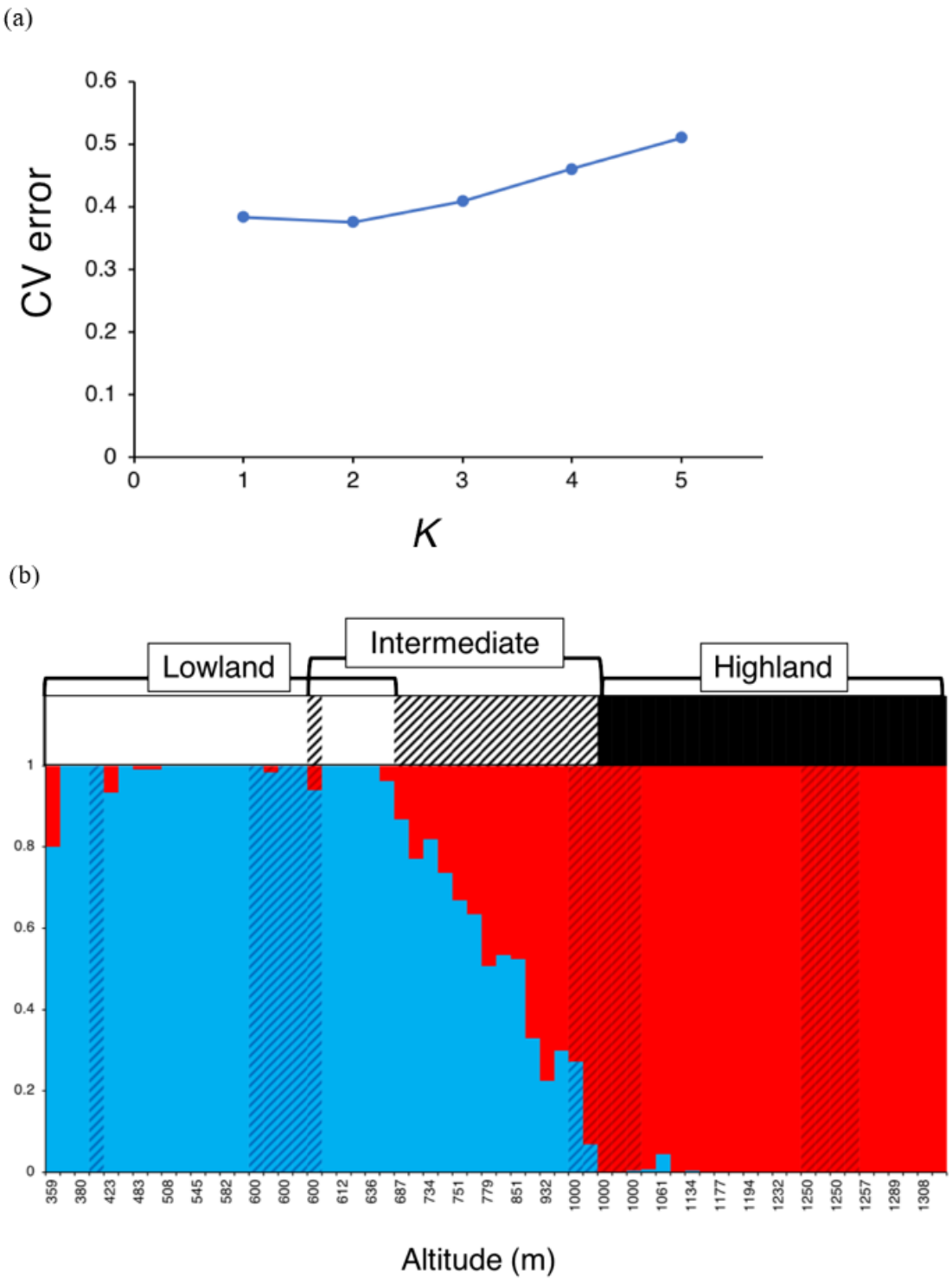

Figure 3

(a) The cross-validation (CV) error as a function of the number of subpopulations (*K*). (b) The population structure when *K*=2. Each bar represents one individual. The shaded portion represent individuals sampled previously in Kubota et al. (2015), and other portions represent individuals sampled in this study. The vertical axis represents the estimated membership in a particular genetic cluster, whereas the

horizonal axis represents altitude (m) where each individual was sampled. The upper side of figure represents the subpopulations (white, lowland; hatched, intermediate; black, highland subpopulation).

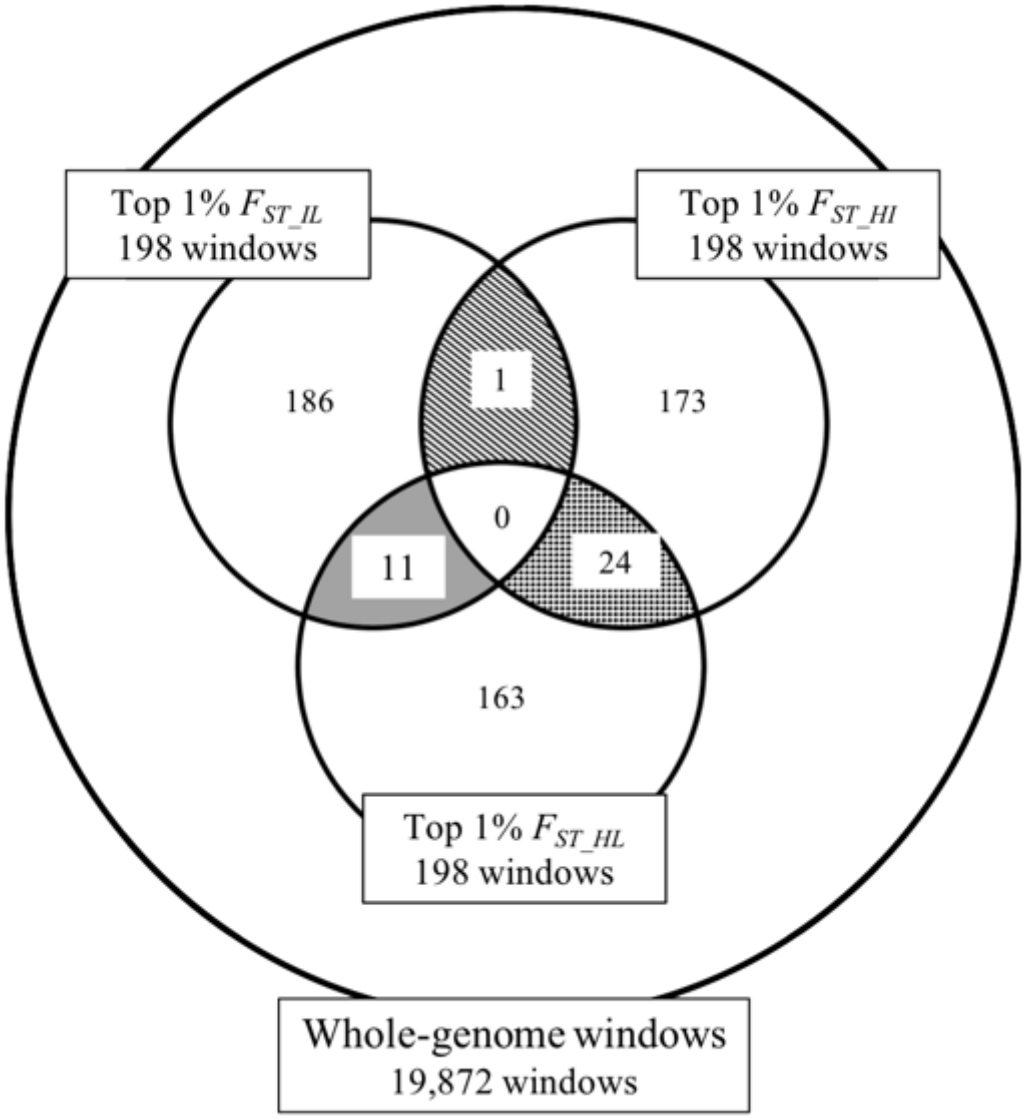

Figure 4

The numbers of genomic windows detected by F_{ST} approach. Each small circle in a large circle represents outliers whose statistic values of F_{ST} was extremely high (the higher 1%) in each calculation.

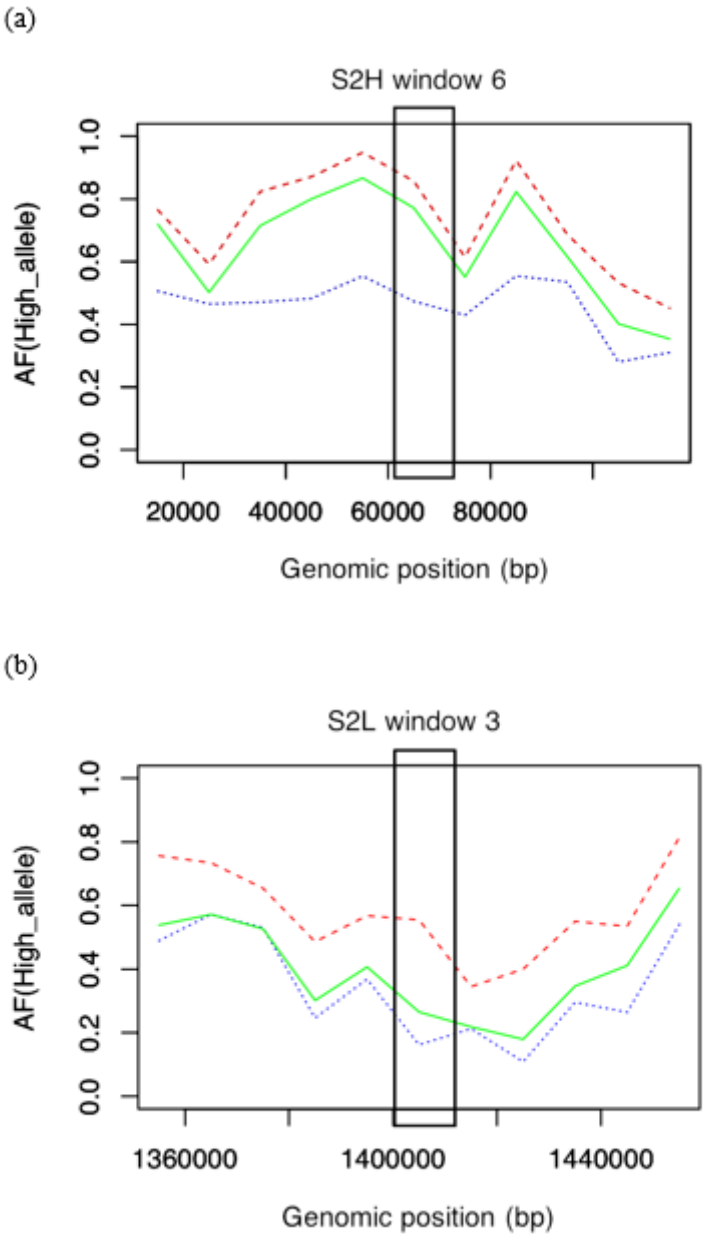

Figure 5

The average of allele frequencies of highland alleles calculated in each genomic window and in each subpopulation as a function of the position of the window in the genome sequence of *Arabidopsis halleri* (Red dashed line; highland, Green solid line; intermediate, Blue dotted line; lowland). The squares in each figure locate on the F_{ST} outlier windows matching (a) the Scenario 2H and (b) 2L.

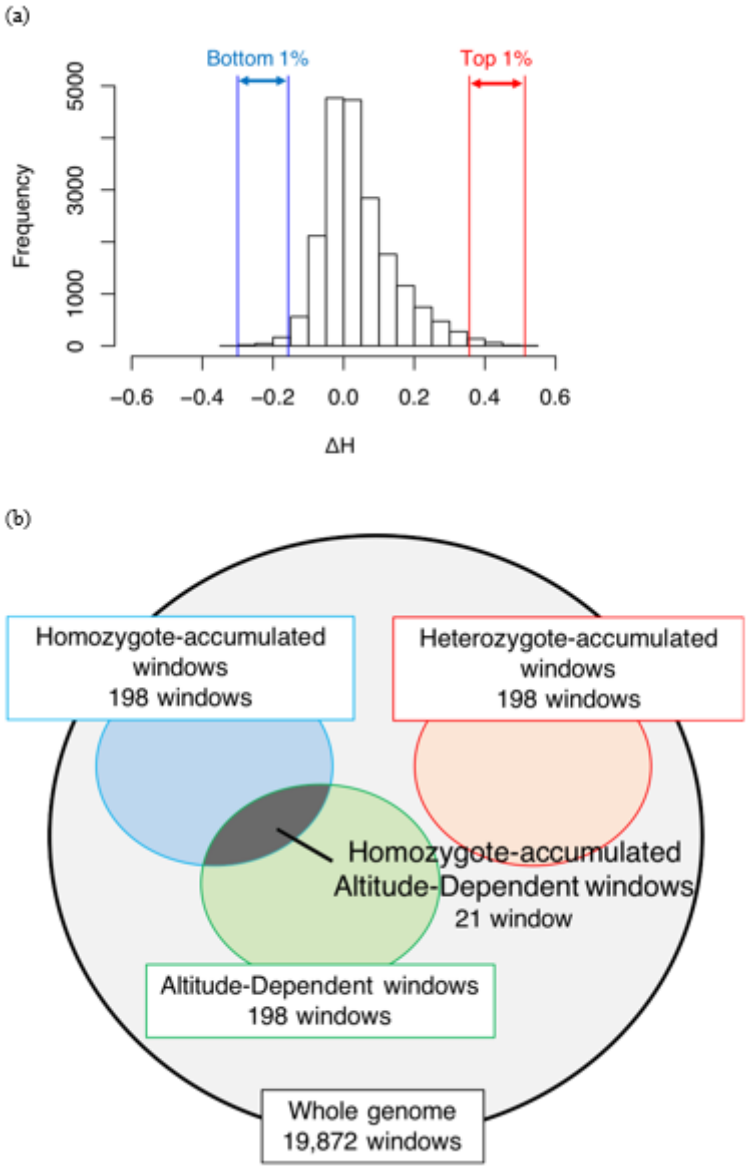

Figure 6

(a) The histogram of the differences between the observed and expected heterozygosity in the intermediate subpopulation (ΔH). The vertical axis is the number of genomic 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).

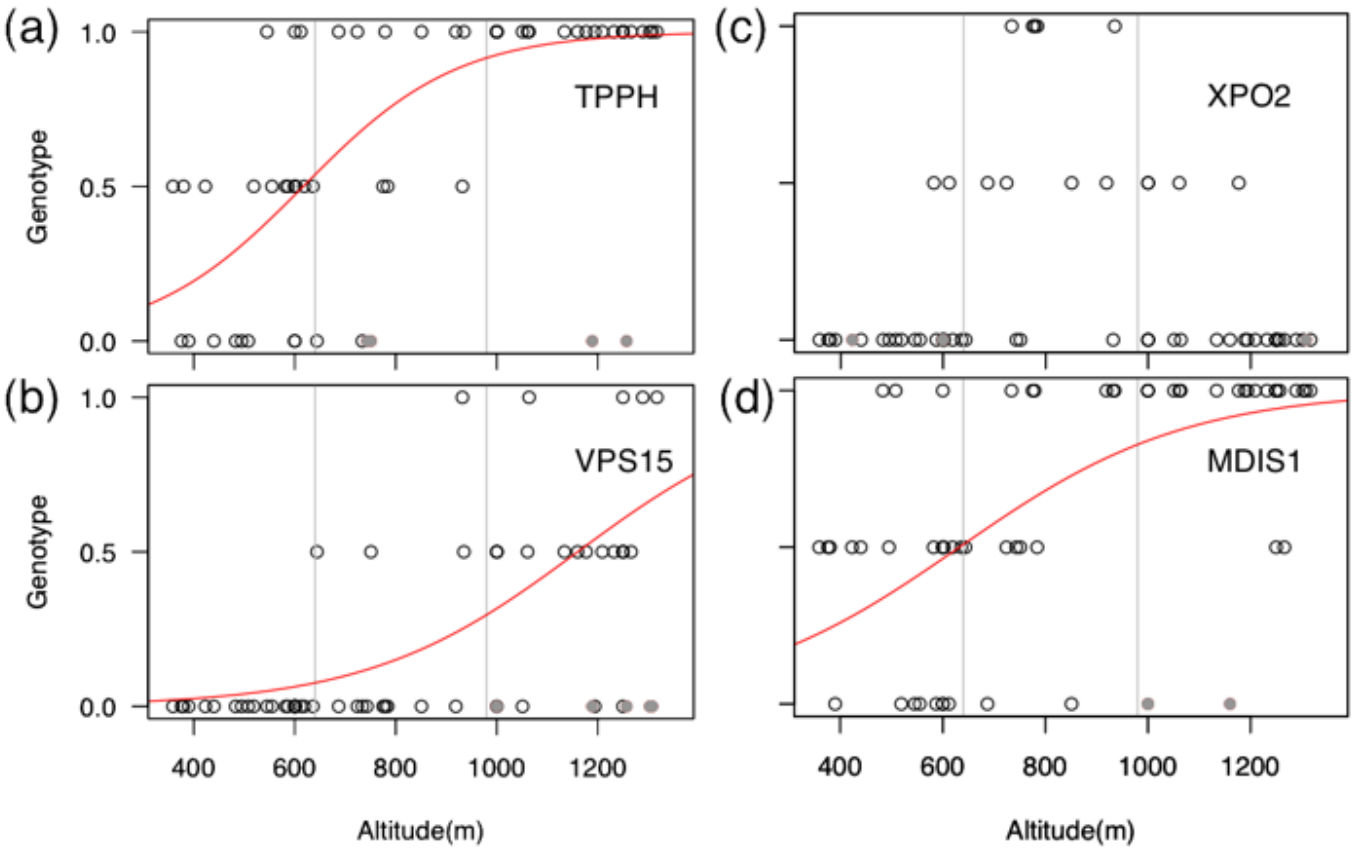

Figure 7

The genotype patterns of variants contained in candidate genome windows. Each circle represents an individual. The vertical axis represents its genotype (homozygote of lowland allele = 0, heterozygote = 0.5, and homozygote of highland allele = 1). The gray dots represent the individuals whose genotype data were eliminated from the analysis due to their low coverage depth. The curvilinear is the logistic regression.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementalmaterialNS1.pdf
- supplementalmaterialNS2.xlsx