# The fate of standing variation and new mutation under climate change

Cheng-Yueh Lu[1], Cheng-Ruei Lee[1,2,3]*

[1]Institute of Plant Biology, National Taiwan University
[2]Institute of Ecology and Evolutionary Biology, National Taiwan University
[3]Genome and Systems Biology Degree Program, National Taiwan University
*Correspondence to: chengrueilee@ntu.edu.tw

**Abstract:**

Many species face existence threat under anthropogenic climate change, and standing genetic variation was proposed as a way for sessile species to adapt to novel environments. However, it is still unknown whether standing genetic variants, being adaptive to current environmental variability, are sufficient to guarantee future survival. Here we investigate the relative importance of standing variation versus new mutations and their relative effect sizes from the past to infer the future. While theoretical and experimental evolution studies provided valuable insights, examples in nature are scarce. In the wild banana *Musa itinerans* in Taiwan, we found enrichment of standing variation with environmental associations. New mutations exert larger effect size per variant in precipitation-related climatic variables, where Taiwan contains extreme environments beyond the ancestral climatic range, and novel alleles have stronger association with novel environments. This pattern is reversed for temperature-related variables, where Taiwan is within the ancestral environmental range. Despite such differences, anthropogenic climate change impacts both types of variants strongly. The results constitute one of the few examples in nature demonstrating that the effect sizes of adaptive variants differ under distinct environmental pressures, and the patterns support theoretical predictions that natural selection favors new mutations with larger effect sizes in a novel environment where the population is far from the adaptive peak. Despite the importance of new mutations, however, the pace of anthropogenic climate change may not allow the accumulation of such mutations.

## Introduction:

Anthropogenic climatic change posts an imminent threat to most organisms. For large and sessile plant species with long generation time, the speed of migration may not keep up with environmental change, and therefore phenotypic plasticity and genetic variation in the population may allow their survival under novel environments[1-4]. Adaptive genetic variation originates from standing variants before or new mutations after environmental change. Since anthropogenic climate change greatly outpaces natural mutation, the amount of standing genetic variation is therefore critical for the rapid response of a population to changing environments[5-7]. It remains unclear, however, whether standing variants are sufficient to guarantee future survival, given that they are mostly adaptive to the present range of climatic variability. While it may be difficult to perform manipulative experiments in the field to compare the effects of new mutations (NM) and standing variation (SV), one could investigate NM and SV during climatic change in the past.

Adaptation could happen through genetic variants that differ in their origins (NM or SV) or effect sizes (Mendelian genes with major effects or polygenic variants with minor effects). However, how these factors interact and respond to environmental pressures remains relatively uninvestigated. For example, does SV and NM differ in their relative number or effect sizes towards environmental adaptation, and how does this relationship change with different types of environmental factor? For adaptive new mutations that were fixed when facing environment change, Fisher first predicted primarily small allelic effects[8] while Kimura emphasized intermediate effects[9]. Orr, later considering the entire adaptive walk, concluded the evolution towards a novel adaptive peak should first happen through fixation of large-effect mutations and later by small-effect polymorphisms[10]. While this was supported by some studies, the majority of these are microbial experimental evolution in well-controlled environments[11,12], and few have specifically compared the effects of NM and SV. To test whether this idea holds in nature, empirical investigations on natural populations are needed.

Taiwan is well-suited for such studies: Unlike oceanic islands such as Hawaii, Taiwan is a continental island where most species originated from the East Asian continent with recurrent gene flow[13]. The land bridge between Taiwan and China during the glacial maximum allowed exchange of SV, and the isolation during interglacial periods enabled the development of NM. Here we investigate the genomic basis of environmental adaptation of a wild banana species, *Musa itinerans*, whose habitats in Taiwan are considered peripheral from ancestral area reconstructions[14], providing an opportunity to distinguish SV from NM, as well as their response to ancestral versus novel adaptive landscapes. We investigated how past events (SV versus NM) influence present adaptation and whether local adaptation may persist under future anthropogenic climate change.

## Results:

### Environmental adaptation in *Musa itinerans*

We first sampled *Musa itinerans* at 24 populations across Taiwan (Fig. 1a; Supplementary Fig. 1a; Supplementary Table 1) and investigated the population structure using 14 microsatellites (Supplementary Table 2). Environmental Niche Modeling (with 483 occurrence points from field survey and Google Street View) reported species distribution (Fig. 1b) in line with the previous statement that *Musa itinerans* inhabits sunny valleys, watersheds, and hillsides with gentle slopes[15]. Populations differentiated mostly between east and west (Supplementary Fig. 1b). The most unsuitable environments lay within the Central Mountain Range and the southwestern plains (Fig. 1b), respectively corresponding to low annual mean temperature (BIO1) and low precipitation of driest quarter (BIO17), the two most important bioclimatic variables determining species distribution (MaxEnt permutation importance[16]: 36.7 for BIO1 and 27.3 for BIO17).

To test for local adaptation, we examined the pattern of "isolation by adaptation[17]", a process where differential local adaptation restricted effective gene flow and promoted genetic differentiation among populations, by dissecting geographic and environmental effects on genetic differentiation. The strait-line "fly-over" geographical distance, calculated as the straight-line distance between locations, does not explain patterns of genetic differentiation (Mantel's $r = 0.146$ and $P = 0.062$). However, if we considered that Central Mountain Range lacks corridors for *M. itinerans* to disperse (Fig. 1b), this fly-over geographical distance could be too unrealistic. We therefore used resistance distance, calculated from the route with least resistance among populations on the niche suitability map (Fig. 1b), to represent the "realized" geographical distance (Fig. 1c) and found that genetic differentiation was significantly associated with resistance (Mantel's $r = 0.226$ and $P = 0.006$). The environmental Mahalanobis distance of bioclimatic variables also showed strong association with genetic differentiation (Mantel's $r = 0.298$ and $P = 0.005$). Given that the environmental distance could be strongly dependent on geography, we performed Partial Mantel test to control the geographical effect. After controlling for realized geographic distance (resistance distance), genetic differentiation still correlates with the Mahalanobis environmental distance (Mantel's $r = 0.250$ and $P = 0.012$), suggesting differential local adaptation is associated with genetic variation.

### Standing variation versus new mutations

To identify genomic regions associated with environmental adaptation, we performed pooled sequencing for each population. SNPs were separated into standing variation (SV: both alleles exist in Taiwan and China) or new mutations (NM, polymorphic only in Taiwan). SV outnumbered NM in both adaptive (identified with Bayenv[18]) and non-adaptive SNPs, and after controlling for the overall number of SNPs in SV and NM, SV were further enriched among adaptive polymorphisms (Supplementary Table 3). However, since adaptive SNPs also have

higher minor allele frequency (MAF) (Supplementary Fig. 2a), this pattern could be confounded: SV are more likely to have higher MAF than NM, and SNPs with higher MAF may be more likely detected as adaptive due to higher statistical power. We therefore performed the same test with a subset where the adaptive and non-adaptive SNPs have similar allele frequencies (ranging from the first quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each bioclimatic variable; Supplementary Fig. 2a). In this case as well, SV are still disproportionately abundant (Supplementary Table 4), suggesting SV are more likely than NM to become environment-associated SNPs. Another potential confounding factor is the geographic extent of variants: if most NM resulted from mutations restricted to a few local populations, the limited distribution prevents environment association for NM. We therefore compared the number of Taiwanese populations containing the minor alleles for SV and NM, respectively. Contrary to the direction predicted by the aforementioned confounding factor, the geographic extent of minor alleles for SV is slightly smaller than NM (15.8 populations for SV and 16.3 for NM, $P < 0.001$).

In addition to SNP number, do NM and SV differ in their directions of effect? Under the null hypothesis that (1) the effects of NM are equally likely to facilitate adaptation to the ancestral or novel environments and (2) natural selection is equally likely to fix NM facilitating adaptation towards either direction, we expected no enrichment of new alleles in either environment. When we separated the Taiwanese populations into those within the Chinese ancestral environmental range and those with novel environments, frequencies of putatively adaptive new alleles in NM SNPs were higher in the latter set of populations, with precipitation of driest quarter (BIO17) and precipitation of coldest quarter (BIO19) showing the strongest effect (Fig. 2a; Supplementary Fig. 2b). Given that the directions of mutation effects should be random, these results suggest that new variants facilitating adaptation to novel environments are more likely to be retained by selection.

The results above suggest SV might be more important than NM in terms of enriched number of variants with environment association, and adaptive NM, while lower in number, are more associated with novel environments. On the other hand, the number of candidate SNPs does not necessarily reflect the overall importance of SV over NM, since the effect size also needs to be considered. To investigate the effect size of SV and NM in environmental adaptation, we compared their Bayes factors from Bayenv and focused on two bioclimatic variables that were important in determining species distribution, annual mean temperature (BIO1) and precipitation of driest quarter (BIO17). The two variables exhibit opposite patterns (Fig. 2c): While in BIO17 and other related precipitation variables (Supplementary Fig. 3a) NM consistently had higher Bayes factor and therefore stronger effect size than SV, in BIO1 and other related temperature variables (Supplementary Fig. 3a) we observed the reverse. The same trend was observed when we estimated the effect size with gradient forest[19,20] (Fig. 2b, c; Supplementary Fig. 3b, c): BIO17 was the most important factor for differential local

adaptation, and NM had stronger effects than SV. On the other hand, BIO1 was the least important factor where SV had stronger effects. Finally, the "importance" estimated by gradient forest is analogous to $r^2$, representing the amount of allele frequency variation explained by environmental gradients. Assuming a simple linear relationship between allele frequency and environment, the value of $r^2$ only represents how well each data point (a population) fits along the regression line. We were, however, also interested in the regression slope: the amount of allele frequency changes along environmental gradients (Supplementary Fig. 3d). Again, BIO17 had the largest overall slope among all bioclimatic variables (Supplementary Fig. 3e), with NM being significantly higher than SV (Fig. 2c). BIO1 had the lowest overall slope, again with the reversed pattern. Given that the MAF of adaptive NM and SV SNPs are similar, there is no need to control for allele frequency in these tests (Supplementary Fig. 4a). Therefore, NM with larger effect size per SNP (as estimated by Bayenv Bayes factor, gradient forest $r^2$ weighted importance, and gradient forest slope) were associated with the adaptation to novel environments outside of the ancestral niche space, consistent with previous population genetics modeling results[8-10].

The observed patterns could be integrated with the unique climate of Taiwan. In comparison to the rest of the species range, northern Taiwan experiences northeastern monsoons during winter and has higher precipitation during the typical winter dry season (Fig. 3b). This pattern has been maintained since at least the last glacial maximum (Fig. 3d). The novel environments might impose novel adaptive optima to the immigrant population from China. The response to selection imposed by these environmental gradients is strong (with highest Bayes factor, $r^2$, and slope among all bioclimatic variables; Supplementary Fig. 3a, c, e) especially for NM, where new alleles are strongly associated with novel environments (Supplementary Fig. 2b). More importantly, for this major driver of adaptation (BIO17), the greatest increment of gradient forest importance lies between 200 mm and 300 mm (Fig. 2b), a range also distinguishing the novel Taiwanese versus ancestral Chinese environments (Fig. 3b). This suggests that the majority of differential local adaptation is associated with such novel-versus-ancestral environmental differences. Annual mean temperature (BIO1) is the other extreme: the environmental gradient within Taiwan is well within the ancestral Chinese environmental range (Fig. 3a), which can also be traced back to the last glacial maximum (Fig. 3c). It is likely that SV already contained genetic variants adaptive to such environmental gradients and are therefore more important than NM (Fig. 2c). In summary, we observed adaptation happening through the assortment of SV for a new territory with similar adaptive landscape and optimum (BIO1; Figs. 2c, 3a, c). For adaptation to novel environments and a new adaptive landscape (BIO17; Figs. 2c, 3b, d), NM with larger effect sizes were more likely favored by natural selection. Our results are therefore consistent with Orr's model[3], providing one of the few examples in nature.

One key point of this study is the correct designation of SV or NM. It is possible that some SV SNPs were mis-assigned as NM if we missed an allele in China, most likely for SNPs with low MAF. We addressed this issue with the following: (1) The Taiwanese populations were nested within China in the phylogenetic tree (Supplementary Fig. 4b) and contained much less genetic variation (Supplementary Fig. 4c). Due to the stronger genetic drift in Taiwan than China, it is less likely that an originally SV SNP would retain both alleles in Taiwan but lose one in China. (2) In the extreme case, assuming 50% of NM SNPs were mis-assigned from SV, we performed 100 new analyses, each randomly assigning 50% of NM SNPs back to SV. These new analyses yielded similar results, with NM having higher effect sizes than SV in precipitation-related variables (Supplementary Fig. 5a). (3) Since MAF are correlated between Taiwan and China (Spearman's rank correlation $\rho = 0.35$, $P < 0.001$), we performed analyses with top 50% MAF SNPs in Taiwan, thereby reducing the chance of missing minor alleles in China. The results are qualitatively the same (Supplementary Fig. 5b-d).

**Fate of adaptive variants under future climate change**

In addition to understanding how past events (SV versus NM) affected present adaptation, we are also concerned with how these factors affect the future of this species under anthropogenic climate change. We predicted 16 future outcomes and used Bayenv to investigate the future fate of currently adaptive SNPs. Currently adaptive SNPs retaining high association with future environments are classified as "retention", while those no longer associated with future environments are "disruption". Different from the present pattern of SV being enriched in adaptive SNPs, we saw no clear tendency for any set of adaptive SNPs enriched towards retention or disruption (Supplementary Table 5), suggesting both SV and NM will be affected by climate change.

Under the 16 future scenarios we identified two extreme conditions: CCSM4-2070-RCP8.5 as the niche-expansion extreme (Fig. 4a-d) and MIROC-2070-RCP4.5 as the niche-contraction extreme (Fig. 4e-h). Intending to investigate whether SV and NM have different fates under the two contrasting future predictions of this species, we used the genetic offset value from the gradient forest results to estimate genetic mismatch for SV and NM separately, which is associated with the magnitude of allele frequency turnover perturbed by future climatic conditions[20] (Fig. 4b, f). The "genetic offset difference" was calculated by subtracting genetic offset caused by SV from that of NM (Fig. 4c, g). We found no general consensus of whether or where in Taiwan NM or SV SNPs are more affected by climate change (Fig. 4c, g), suggesting NM and SV might both be affected by future anthropogenic climate change despite their distinct genetic architecture in adaptation to current environments.

Traditional species distribution models build the species-wide niche model from all occurrence records of a species, thereby assuming all populations in a species reacting equally to the environment and overlooking local adaptation due to within-species polymorphism[6].

Here we propose a concept, "extinction risk", integrating species-level and population-level responses to future climate change simply by dividing genetic offset (estimated from gradient forest) by suitability (estimated from MaxEnt). While the two future conditions exhibit slight differences in suitability and genetic offset, the extinction risks are surprisingly similar, with very high risk in western Taiwan (Fig. 4d, h). In Western Taiwan, the expansion-extreme model suggested high suitability for the species as a whole but also high local population mismatch, thereby having similar trends of extinction risk as the contraction-extreme model. Taken together, our results emphasize the importance of considering both whole-species suitability as well as adaptation of local populations when considering the effect of anthropogenic climate change on the fate of sessile species.

**Discussion:**

Assessing the within-species variation in climate association is the crucial first step to understand species susceptibility to fluctuating environment, and the relative importance of standing variation (SV) and new mutations (NM) in adaptation has long been debated[7,21-25]. In this study, we investigate how past genetic variation (SV and NM) contribute to the adaptation of present environments as well as how these factors together affect the future fate of a species under anthropogenic climate change.

For a population facing environmental change and therefore a novel adaptive landscape, previous population genetics models have documented the effect size distribution of new mutations fixed by natural selection. Considering the entire adaptive walk of a population facing novel environments, Orr's model predicted the distribution of effect sizes where early substitutions have larger effect than later ones[10]. Consistent with Orr's prediction, we show that NM have stronger effect size than SV in precipitation-related variables, where Taiwan exceeds the ancestral climatic range in China (and therefore the migrating population was far away from the optimum in the new adaptive landscape). This pattern is reversed for temperature-related variables, where Taiwan has similar environmental range as China. Here we provide another perspective to recent research showing that SV contributes to adaptation[7,25]. We show that SV indeed dominate over NM in number[7]. However, the effect size of SV and NM hinges on environmental conditions: natural selection may prefer new mutations contributing to the adaptation to the new rather than the old environment, and the effect sizes of NM tend to be higher than SV under such conditions. Our results imply that standing genetic variation in a population may not be sufficient for the adaptation to anthropogenic climate change.

Ecological niche models are widely used to predict future species distribution, but such models often do not account for the differential adaptation of populations within species. Our observed difference between the major determinant of species-wide range distribution (BIO1) and the driver of differential adaptation of populations within species (BIO17) demonstrates

the need to consider local adaptation. Incorporating species distribution modeling (predicting the future range of the whole species) and genetic offset (predicting the genetic mismatch of each local population to future environments), we identified the potential risk of western lowland populations despite the seemingly distinct species distribution modeling results in different future climate scenarios. In other words, despite in some future scenarios the range of environments suitable for the whole species might increase, giving the impression that the species benefits from anthropogenic climate change, such change might be detrimental to each local population uniquely adapted to a much narrower environmental range than the whole species. Finally, we separately calculated the genetic offset originated from the mismatch of SV or NM to future environments and showed that both will be affected by anthropogenic climate change, regardless of their distinct genetic architecture towards environmental adaptation. Since we distinguished SNPs into SV or NM based on past environmental change, all of these genetic variants will be standing variation when facing future climate change. Novel mutations facilitating the adaptation to new environments may happen in the future, although they will be strongly outpaced by anthropogenic climate change.

## Methods:

### Sample collection and DNA extraction.

Field work was conducted during 2017 (August - December) and 2018 (January - May). We sampled *Musa itinerans* at 24 sites across Taiwan (Supplementary Fig. 1a; Supplementary Table 1). Fresh leaves were harvested from nine to fifteen individuals at each site. Total genomic DNA was extracted using the standard CTAB extraction method[26]. Since other commercial *Musa* species were also grown in Taiwan, we developed an indel marker for species delimitation. From previous studies[27-30], we identified a 6-bp insertion specific for the Taiwanese *M. itinerans* in the *atpB-rbcL* region of chloroplast. We designed a primer pair (5'-GAAGGGGTAGGATTGATTCTCA-3'; 5'-CGACTTGGCATGGCACTATT-3') and used amplicon size to confirm all collected samples are Taiwanese *M. itinerans*.

### Simple sequence repeat genotyping and analysis.

SSR primer sequences used in this study were originally developed for the genus *Musa*[31,32], which were then applied on *Musa itinerans*. Previously documented primer sequences were first searched against the *Musa acuminata* DH-Pahang genome version 2[33] on Banana Genome Hub (https://banana-genome-hub.southgreen.fr/) to check specificity as having only one amplicon, resulting in 26 primer pairs. These primers were then experimented to check specificity on *M. itinerans*, resulting in 14 pairs (Supplementary Table 2). We modified each pair of primers by capping the 5' end of forward primers with M13 sequences (CACGACGTTGTAAAACGAC) and inflorescent molecules[34]. SSR amplicons were run through capillary electrophoresis and the length of each allele was recorded.

Population structure of 20 populations (Supplementary Table 1) was analyzed with 14 SSR markers (Supplementary Table 2). Lowland populations (C35H, WFL, THNL, PTWT, P199H, MLLYT, HDPG, TTL, NAJY, HLCN, NXIR, and DFR), east transect populations (TPS300, TPS500, TPS700, and TPS900), and west transect populations (XT400, XT700, XT1200, and XT1500) were used in the analysis. We inferred the ancestry of 244 individuals with STRUCTURE 2.3.4[35,36], parameterizing a run to have (1) run length of burnin and after-burnin period of 100,000, (2) admixture ancestry model, and (3) independent allele frequency model, further setting 20 runs for each *K* value.

To investigate the association among genetic, geographical, and environmental distance, we generated these distance matrices. Genetic distance was calculated by GenAlEx 6.503[37,38] from 14 SSR markers; straight geographical distance (the fly-over distance) was generated by ArcGIS 10.5 (http://desktop.arcgis.com/en/); environmental distance was measured as Mahalanobis distance to address the correlation among nine bioclimatic variables (below). In addition to the fly-over geographical distance which assumes organism dispersal ignores landscapes, we further calculated as resistance distance the cumulative cost along the least cost path (below). Matrix association was examined under Mantel and Partial Mantel tests. Statistical significance was examined with 1,000 permutations. We performed Mantel tests on (1) genetic distance *vs*. fly-over distance, (2) genetic distance *vs*. resistance distance, (3) genetic distance *vs*. Mahalanobis environmental distance, and Partial Mantel tests on (4) genetic distance *vs*. Mahalanobis environmental distance while controlling for resistance distance.

**Species distribution modeling.**

Current and future species distribution models were built for *Musa itinerans* using presence-only data (483 occurrence points) obtained from field survey and Google Street view. Occurrence points were then reduced to 204 cells by the removal of co-occurring presence data within the same 1×1 km grid. MaxEnt 3.4.1[16], implemented with the maximum entropy modeling approach, reports an overall niche suitability and the importance of predictors by analyzing the presence-only data as well as background (psuedo-absence) data distribution[39-41]. We downloaded from WorldClim database version 1.4 (http://worldclim.org/) spatial layers of 19 present-day bioclimatic variables based on high-resolution monthly temperature and rainfall data[42]. Layers were selected at spatial resolution of 30 arc-second and with a mask that ranges 119.25-122.47°E and 21.76-25.49°N covering Taiwan. Variables showing high dependence (Pearson's correlation coefficient > 0.9 calculated from ENMTools[43]) from each other were removed, resulting in nine final variables: BIO1 — mean annual temperature, BIO2 — mean diurnal range, BIO3 — isothermality, BIO7 — temperature annual range, BIO12 — annual precipitation, BIO15 — precipitation seasonality, BIO16 — precipitation of wettest quarter, BIO17 — precipitation of driest quarter, and BIO19 — precipitation of coldest quarter (Supplementary Table 1).

Present species distribution model was constructed using the default optimization settings in MaxEnt, except the regularization set to three. We tested the predictive model by ten-fold cross-validation which was carried out by randomly partitioning the data into ten equally sized subsets and then replicating models while omitting one subset in turn. In each turn, the predictive model was built using nine subsets as training data and evaluated using the other subset as test data. The output of the predictive model is the probability of presence, or called suitability, and we averaged the ten runs to have an averaged suitability.

To predict the species distribution under different scenarios of future climate change, we projected the present-day model onto eight future climatic conditions combining two periods (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5). Future climatic layers were obtained from the WorldClim database at spatial resolution of 30 arc-second and were developed based on two general circulation models: the Community Climate System Model[44], CCSM, and the Model for Interdisciplinary Research on Climate[45], MIROC. Species distribution models for the future were carried out using the same settings described above.

To estimate the least cost path between populations, we first generated the resistance surface by taking the reciprocal of suitability. Resistance and suitability is simply a monotonic transformation in which locations with higher suitability exhibit lower resistance. Pairwise least cost path was then measured among 20 populations from the resistance surface, performed by SDM Toolbox v2.3[46]. While least cost path is the single line with least overall cost, we also constructed the least cost corridor between populations, allowing 1%, 2%, or 5% higher cost than the least cost value. In essence, the least cost corridors represent the realized dispersal routes of organisms along suitable habitats.

**Sequencing library construction and SNP identification.**

We conducted whole genome pooled-sequencing[47] for each population (Supplementary Table 1), resulting in 24 pooled-sequencing libraries. Equal amount of DNA from ten individuals at each population were pooled, except for the PTWT population where only nine individuals were available. A library with 300-400 bp insert size for each pool was prepared using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs). Libraries were then sequenced with 150 bp paired-end on the HiSeq X Ten platform.

Illumina reads were then trimmed with SolexaQA[48], followed by the removal of adaptor sequences with cutadapt[49], subsequently mapped to the *Musa itinerans* reference genome assembly ASM164941v1[50] with BWA 0.7.15[51]. Picard Tools (http://broadinstitute.github.io/picard) was used to mark duplicated read pairs, and the genotypes were called following GATK 3.7 best practice[52].

For the 24 pooled samples, we filtered out sites with (1) more than two alleles, (2) indels, (3) quality (QUAL) < 30, (4) quality by depth (QD) < 2, (5) call rate < 0.74, and (6) depth (DP)

> genome-wide average depth plus three standard deviations, resulting in 4,200,177 SNPs. SNPs with (1) minor allele frequency (MAF) < 0.05, (2) missing data in any of the pooled-seq sample, and (3) DP per sample < 20 were further filtered out, resulting in 1,256,894 SNPs.

To investigate the relationship between Taiwanese and Chinese *M. itinerans*, we downloaded public data from 24 Chinese accessions (SRR6382516 - SRR6382539)[53]. SNPs were called using all 24 Chinese and the 24 Taiwanese samples together following the pipeline described above. We did not perform any site filtering for this joint data set since the main objective is to investigate whether specific SNPs in Taiwan also existed in China as SV. This dataset has 18,442,853 SNPs. SRR6382532 was excluded due to high missing rate. Only when evaluating the averaged expected heterozygosity between Taiwanese and Chinese populations did we filter out sites with (1) indels and (2) QUAL < 30, resulting in 15,591,923 SNPs.

To assess the phylogeny of our Taiwanese populations and Chinese accessions, we downloaded *Musa acuminata* sequence (SRR7013754) as an outgroup. SNPs were called using one *M. acuminata* species, 24 Chinese and 24 Taiwanese samples together following the pipeline described above. We filtered out sites with (1) more than two alleles, (2) indels, (3) QUAL < 30, and (4) call rate < 0.9, resulting in 12,693,687 SNPs. This dataset also excluded SRR6382532.

**Environmentally-associated SNP identification.**

We used Bayenv 2.0[18] to search for SNPs highly associated with environmental variables. Bayenv estimates the relationship between SNPs and environments while controlling the whole-genome population structure from a subset of loose linkage-disequilibrium SNPs. Loose linkage-disequilibrium SNPs were formed by sampling (1) one SNP from scaffolds more than 10 kb and less than 100 kb, (2) two SNPs from scaffolds more than 100 kb and less than 500 kb, (3) three SNPs from scaffolds more than 500 kb and less than 1000 kb, and (4) four SNPs from scaffolds more than 1000 kb. We then, for each bioclimatic variable, defined as the adaptive SNPs ones exhibiting top 1% Bayes factor and top 5% rho value (a nonparametric correlation coefficient capable to reduce outlier effects).

We further investigated the fate of currently adaptive SNPs under anthropogenic climate change, performing the same Bayenv analyses of currently adaptive SNPs using future climatic conditions. We included two time periods (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) from two general circulation models, CCSM[44] and MIROC[45], resulting in 16 future climatic conditions. If a currently adaptive SNP remains strongly associated with environments, it should exhibit Bayes factor above the current threshold. We then defined as “retention” a currently adaptive SNP constantly exhibiting Bayes factor above the current adaptive threshold in all future scenarios, and defined as “disruption” a currently adaptive SNP exhibiting Bayes factor above the current adaptive threshold in none of the future scenarios.

**The gradient forest method and genetic offset.**

We used a novel method, gradient forest[19,20], to estimate the effect of environmental gradients on allele frequency differences among populations. Gradient forest is a regression-tree based machine-learning algorithm using environmental variables to partition SNP allele frequencies. The analysis was done separately for each SNP. The resulting "importance" measures how much of the variation in allele frequency was explained by partitioning the populations based on a specific value in an environmental variable. By making multiple regression trees (thus generating a random forest) for a SNP, the goodness-of-fit $r^2$ of a random forest is measured as the proportion of variance explained by this random forest, which is then partitioned among environmental variables in proportion to their conditional importance. Such SNP-wise importance of each environmental variable is then averaged across SNPs belonging to the standing variation (SV) or new mutations (NM) group, resulting in the overall importance (of each environmental variable). In the end, one could obtain the relation curve between environmental gradient and cumulative importance (analogous to the cumulative $r^2$, proportion of allele frequency differences among populations explained by environments). This curve has two properties. First, the highest point of the cumulative importance curve denotes the overall association between a climatic variable and allele frequency, and we used this to represent the effect size of these SNPs. Second, when traversing along the environmental gradient, a sudden increase of cumulative importance at some environmental value (for example, 20°C) denotes populations on either side of this environmental cutoff have very different allele frequency compositions. In other words, this represents a threshold factor for local adaptation.

One can use this cumulative importance curve to estimate the effect of future environmental change on local populations. In the example above, a population's local temperature increased from 19°C to 21°C due to climate change would require larger allele frequency shift than another population whose local temperature changed from 17°C to 19°C. The "genetic offset"[10] could then be calculated as the Euclidean distance between cumulative importance corresponding to the contemporary environmental value and that corresponding to the future environmental value, considering all bioclimatic variables together. Genetic offset can then be considered to be the magnitude of genetic change needed for a population to be still adaptive in the face of climate change.

**Regression slope.**

The regression slope is not given by gradient forest, since it only reports the $r^2$ importance estimate. Thus, we introduced the simple linear regression $y=\alpha+\beta x$ to measure the regression slope. We took $y$ as the allele frequency, $x$ as the standardized bioclimatic variable, and $\beta$ (slope) as the measurement of the amount of allele frequency changes along environmental gradients. By fitting simple linear regression with the general least-square approach, $\beta$ can

445 then be expanded to $r_{xy}\frac{s_y}{s_x}$, where $r_{xy}$ is the correlation coefficient (the square root of gradient
446 forest measured "importance") between $x$ (environment value) and $y$ (allele frequency), and
447 $s_x$ and $s_y$ are the standard deviation of $x$ and $y$.
448

**References:**


1 Bradshaw, A. D. Evolutionary significance of phenotypic plasticity in plants. in *Advances in Genetics* **13** (eds Caspari, E. W. & Thoday, J. M.) 115-155 (Academic Press, 1965).

2 Ghalambor, C. K., McKay, J. K., Carroll, S. P. & Reznick, D. N. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* **21**, 394-407, doi:10.1111/j.1365-2435.2007.01283.x (2007).

3 Jump, A. S. & Peñuelas, J. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* **8**, 1010-1020, doi:10.1111/j.1461-0248.2005.00796.x (2005).

4 Franks, S. J., Weber, J. J. & Aitken, S. N. Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evol. Appl.* **7**, 123-139, doi:10.1111/eva.12112 (2014).

5 Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38-44, doi:10.1016/j.tree.2007.09.008 (2008).

6 Razgour, O. *et al.* Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl. Acad. Sci. USA* **116**, 10418-10423, doi:10.1073/pnas.1820663116 (2019).

7 Lai, Y. T. *et al.* Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl. Acad. Sci. USA* **116**, 2152-2157, doi:10.1073/pnas.1813597116 (2019).

8 Fisher, R. A. *The Genetical Theory of Natural Selection*. (Clarendon Press, Oxford, England, 1930).

9 Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, Cambridge, England, 1983).

10 Orr, H. A. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935-949, doi:10.2307/2411226 (1998).

11 Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243-1247, doi:10.1038/nature08480 (2009).

12 Gifford, D. R., Schoustra, S. E. & Kassen, R. The length of adaptive walks is insensitive to starting fitness in *Aspergillus nidulans*. *Evolution* **65**, 3070-3078, doi:10.1111/j.1558-5646.2011.01380.x (2011).

13 Bramwell, D. & Caujapé-Castells, J. *The Biology of Island Floras*. (Cambridge University Press, Cambridge, 2011).

14 Janssens, S. B. *et al.* Evolutionary dynamics and biogeography of Musaceae reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytol.* **210**, 1453-1465,

doi:10.1111/nph.13856 (2016).

15 Chiu, H. L. The collection, evaluation and analysis of genetic diversity of *Musa formosana* (Warb.) Hayata native in Taiwan. (National Taiwan University, 2004).

16 Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231-259, doi:10.1016/j.ecolmodel.2005.03.026 (2006).

17 Nosil, P., Egan, S. P. & Funk, D. J. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316-336, doi:10.1111/j.1558-5646.2007.00299.x (2008).

18 Gunther, T. & Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205-220, doi:10.1534/genetics.113.152462 (2013).

19 Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156-168, doi:10.1890/11-0252.1 (2012).

20 Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1-16, doi:10.1111/ele.12376 (2015).

21 Barton, N. H. Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **355**, 1553-1562, doi:10.1098/rstb.2000.0716 (2000).

22 Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095-1098, doi:10.1126/science.1175826 (2009).

23 Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61, doi:10.1038/nature10944 (2012).

24 Reid, N. M. *et al.* The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305-1308, doi:10.1126/science.aah4993 (2016).

25 Ahrens, C. W., Byrne, M. & Rymer, P. D. Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. *Mol. Ecol.* **28**, 2502-2516, doi:10.1111/mec.15092 (2019).

26 Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19,** 11-15 (1987).

27 White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. in *PCR Protocols* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 315-322 (Academic Press, San Diego, California, 1990).

28 Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105-1109, doi:10.1007/BF00037152 (1991).

29 Oxelman, B., Lidén, M. & Berglund, D. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Syst. Evol.* **206**, 393-410, doi:10.1007/BF00987959 (1997).

30 Hurr, K. A., Lockhart, P. J., Heenan, P. B. & Penny, D. Evidence for the recent dispersal of *Sophora* (Leguminosae) around the Southern Oceans: molecular data. *J. Biogeogr.* **26**, 565-577, doi:10.1046/j.1365-2699.1999.00302.x (1999).

31 Lagoda, P. J. *et al.* Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Mol. Ecol.* **7**, 659-663 (1998).

32 Crouch, H. K., Crouch, J. H., Jarret, R. L., Cregan, P. B. & Ortiz, R. Segregation at microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Sci.* **38**, 211-217, doi:10.2135/cropsci1998.0011183X003800010035x (1998).

33 Martin, G. *et al.* Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* **17**, 243, doi:10.1186/s12864-016-2579-4 (2016).

34 Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **18**, 233-234, doi:10.1038/72708 (2000).

35 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).

36 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587 (2003).

37 Peakall, R. O. D. & Smouse, P. E. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288-295, doi:10.1111/j.1471-8286.2005.01155.x (2006).

38 Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537-2539, doi:10.1093/bioinformatics/bts460 (2012).

39 Bean, W. T., Stafford, R. & Brashares, J. S. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* **35**, 250-258, doi:10.1111/j.1600-0587.2011.06545.x (2012).

40 Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129-151, doi:10.1111/j.2006.0906-7590.04596.x (2006).

41 Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43-57, doi:10.1111/j.1472-4642.2010.00725.x (2011).

42 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965-1978, doi:10.1002/joc.1276 (2005).

43 Warren, D. L., Glor, R. E. & Turelli, M. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**, 607-611, doi:10.1111/j.1600-0587.2009.06142.x (2010).

44 Gent, P. R. *et al.* The Community Climate System Model Version 4. *J. Climate* **24**, 4973-4991, doi:10.1175/2011jcli4083.1 (2011).

45 Watanabe, S. *et al.* MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.* **4**, 845-872, doi:10.5194/gmd-4-845-2011 (2011).

46 Brown, J. L. & Anderson, B. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol.* **5**, 694-700, doi:10.1111/2041-210x.12200 (2014).

47 Schlotterer, C., Tobler, R., Kofler, R. & Nolte, V. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**, 749-763, doi:10.1038/nrg3803 (2014).

48 Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485, doi:10.1186/1471-2105-11-485 (2010).

49 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12, doi:10.14806/ej.17.1.200 (2011).

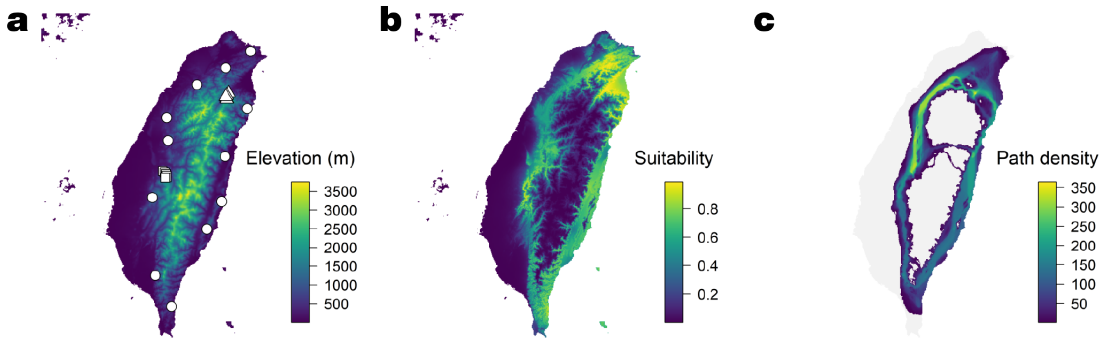
50 Wu, W. *et al.* Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Sci. Rep.* **6**, 31586, doi:10.1038/srep31586 (2016).

51 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

52 McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

53 Wu, W., Ng, W. L., Yang, J. X., Li, W. M. & Ge, X. J. High cryptic species diversity is revealed by genome-wide polymorphisms in a wild relative of banana, *Musa itinerans*, and implications for its conservation in subtropical China. *BMC Plant Biol.* **18**, 194, doi:10.1186/s12870-018-1410-6 (2018).

595 **Figures:**




596

597 **Fig. 1 | Sample distribution and niche modeling of *Musa itinerans*. a**, Sampling sites are
598 distributed along the latitudinal and altitudinal gradient. Lowland populations are represented
599 as white circles; transect populations are represented as white triangles and squares. **b**,
600 Suitability is derived from MaxEnt niche modeling. **c**, Least-cost-corridor landscape is
601 constructed from pairwise least-cost paths among 20 populations from the resistance surface
602 (the reciprocal of niche suitability).

603

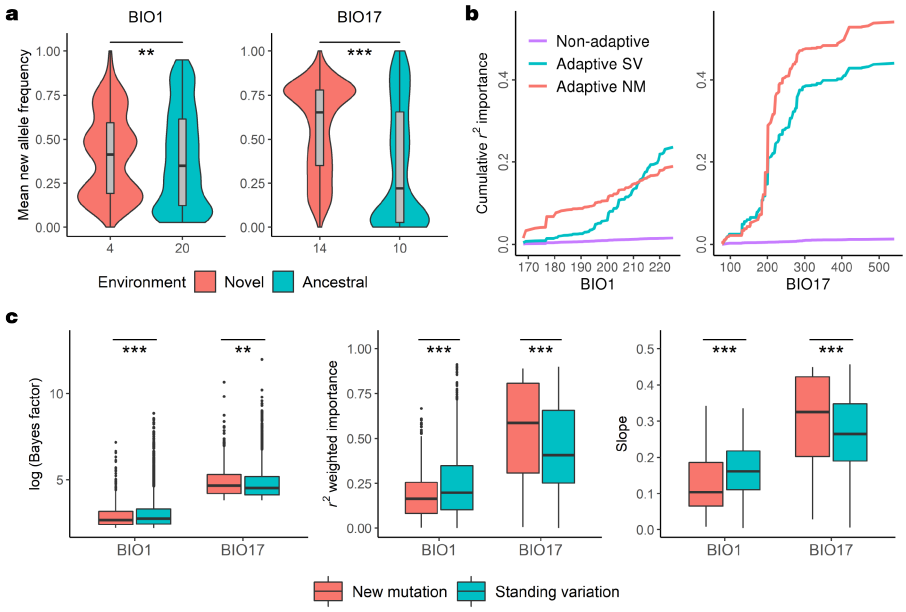



**Fig. 2 | The environment-dependent enrichment of new alleles and the distinct effect sizes of standing variation (SV) and new mutations (NM) in contrasting climatic factors.** BIO1 reflects annual mean temperature, and BIO17 indicates precipitation of driest quarter. **a**, Mean frequency of new alleles among NM SNPs compared between populations that have local environments within or outside of the ancestral climatic range. New alleles are enriched in novel environments (**$P < 0.01$, ***$P < 0.001$, *t*-test). Values on the horizontal axis denote the number of Taiwanese populations within the ancestral or novel environmental range. **b**, Gradient forest cumulative $r^2$ importance is shown along environmental gradients. **c**, Effect sizes as estimated from Bayes factor, gradient forest $r^2$ importance, and gradient forest slope all show that SV exhibit higher effect sizes in BIO1 but the reverse in BIO17 (**$P < 0.01$, ***$P < 0.001$, Wilcoxon rank-sum test for Bayes factor and *t*-test for $r^2$ importance and slope).

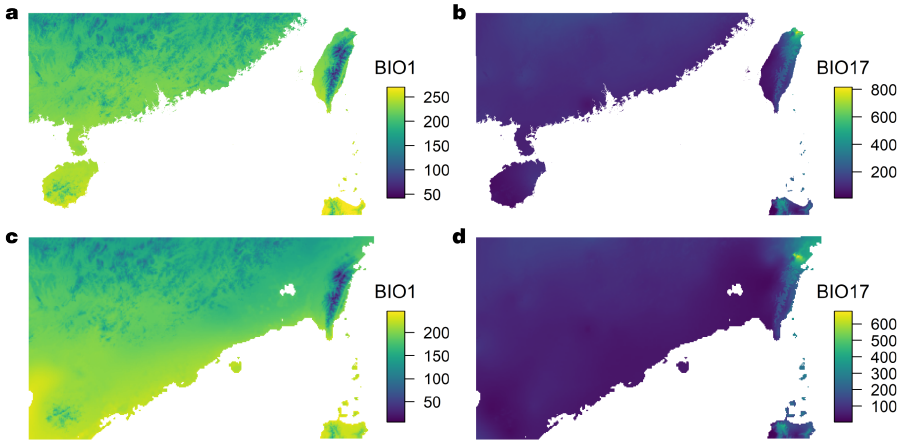



**Fig. 3 | Annual mean temperature (BIO1) and precipitation of driest quarter (BIO17) for the present and last glacial maximum. a**, **b**, Present environments. **c**, **d**, Last-glacial-maximum environments, showing the environments on the extent of land. Maps on all panels have the same range in latitude and longitude.

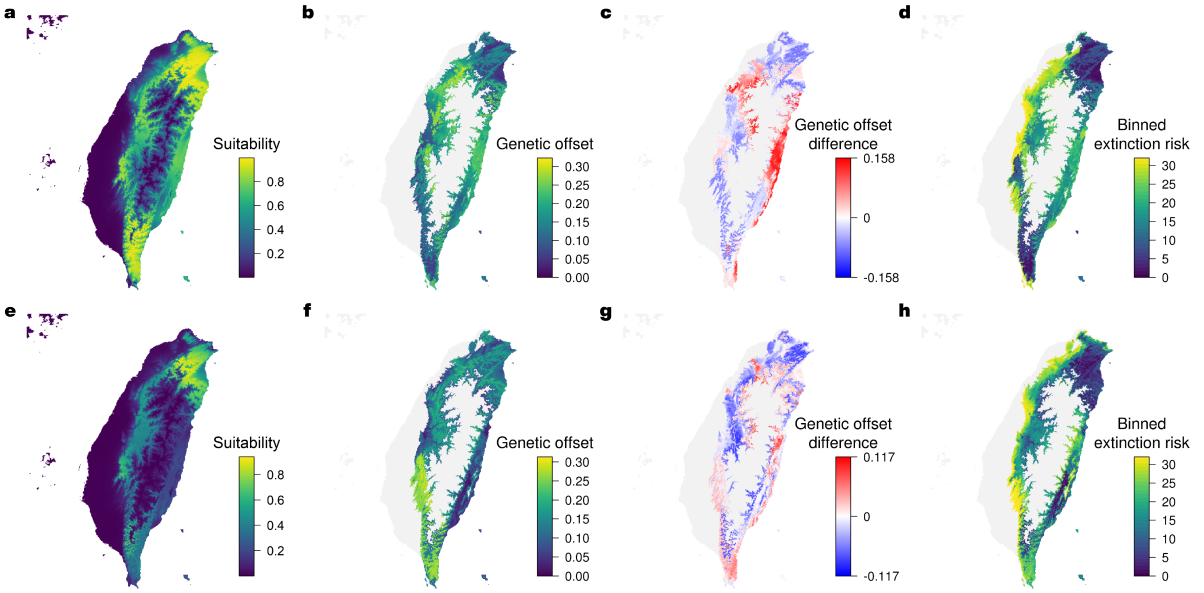



**Fig. 4 | The fate of *Musa itinerans* under two extreme future conditions. a-d,** Niche-expansion extreme (CCSM4-2070-RCP8.5) reports the species-average niche suitability from MaxEnt modeling (**a**), the genetic mismatch of locally adaptive populations to future environments based on the genetic offset from gradient forest (**b**), genetic offset difference calculated as the genetic offset experienced by new mutations minus that from standing variation (**c**), and extinction risk estimated as genetic offset divided by niche suitability (shown as 32 bins with equal grid counts; **d**). **e-h**, Niche-contraction extreme (MIROC-2070-RCP4.5) reports the suitability (**e**), genetic offset (**f**), genetic offset difference (**g**), and extinction risk (**h**). Grids with present niche suitability < 0.2 are excluded and colored in gray in **b-d** and **f-h**.

**Data Availability:**

Population pooled sequencing reads are available under NCBI BioProject PRJNA575344.

**Acknowledgements:**

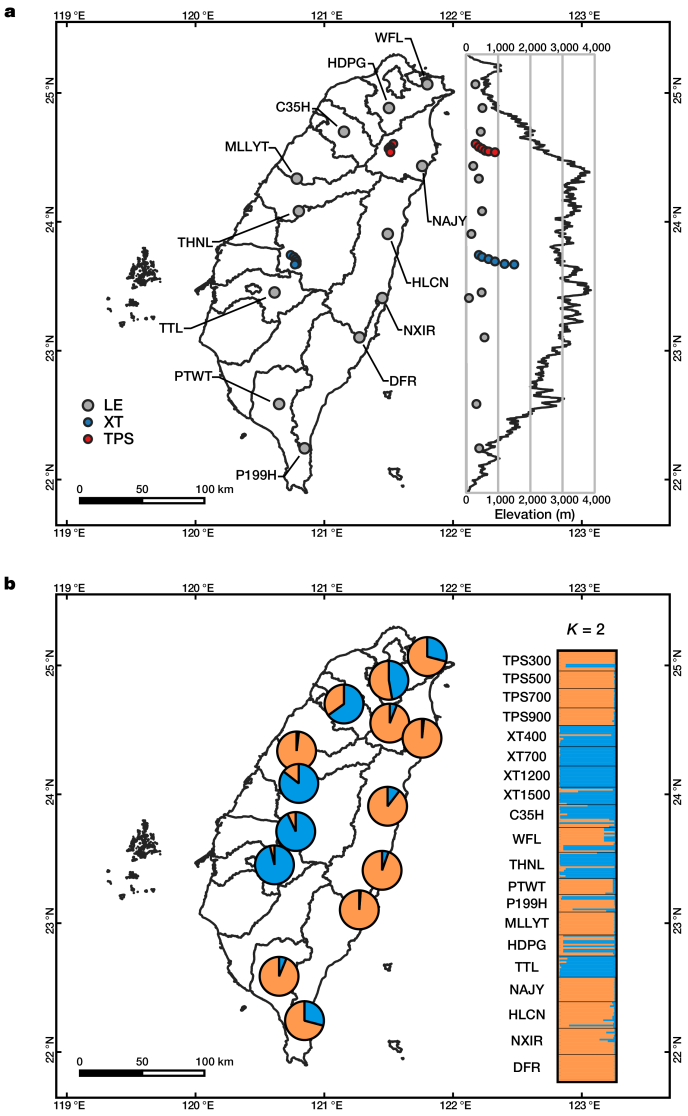
We thank Thomas Mitchell-Olds for valuable comments, Hui-Long Chiu for the knowledge of *Musa itinerans* ecology in Taiwan, Chia-Yu Chen, Zhe-Ting Kuo, and Jo-Wei Hsieh for the assistance of sample preparation, Hao-Chih Kuo for the introduction to species distribution modeling, Cheng-Tao Lin for providing the R template of some maps, the Lee lab member for valuable discussions, and the Computer and Information Networking Center of National Taiwan University for the support of high-performance computing facilities. This work is supported by the Ministry of Science and Technology of Taiwan (106-2628-B-002-001-MY3, 107-2636-B-002-004, and 108-2636-B-002-004 to CRL).

**Author Contributions:**

C-RL designed the research. Both authors performed experiments and analyses and wrote the manuscript.

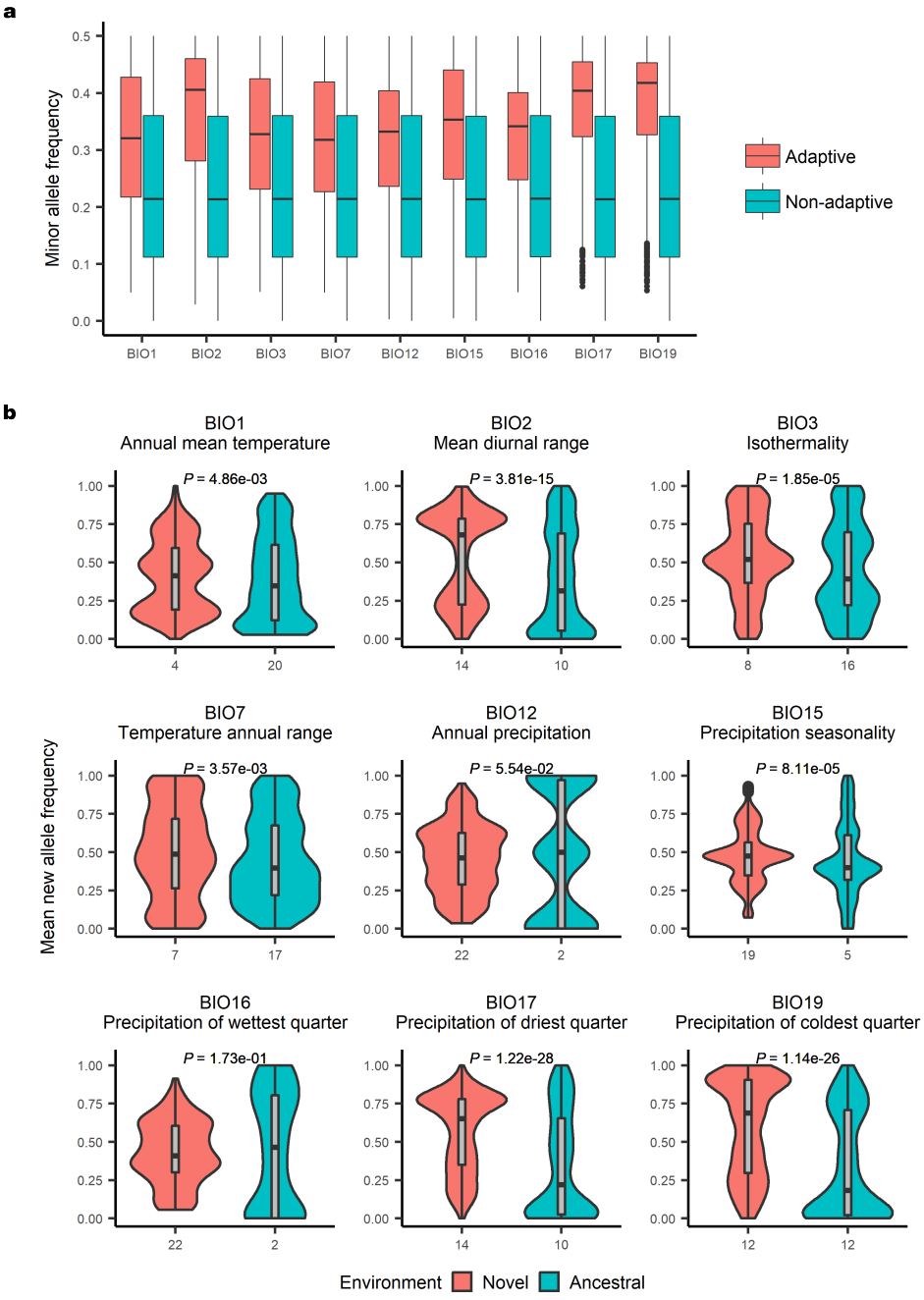
**Competing Interest Declaration:**

None

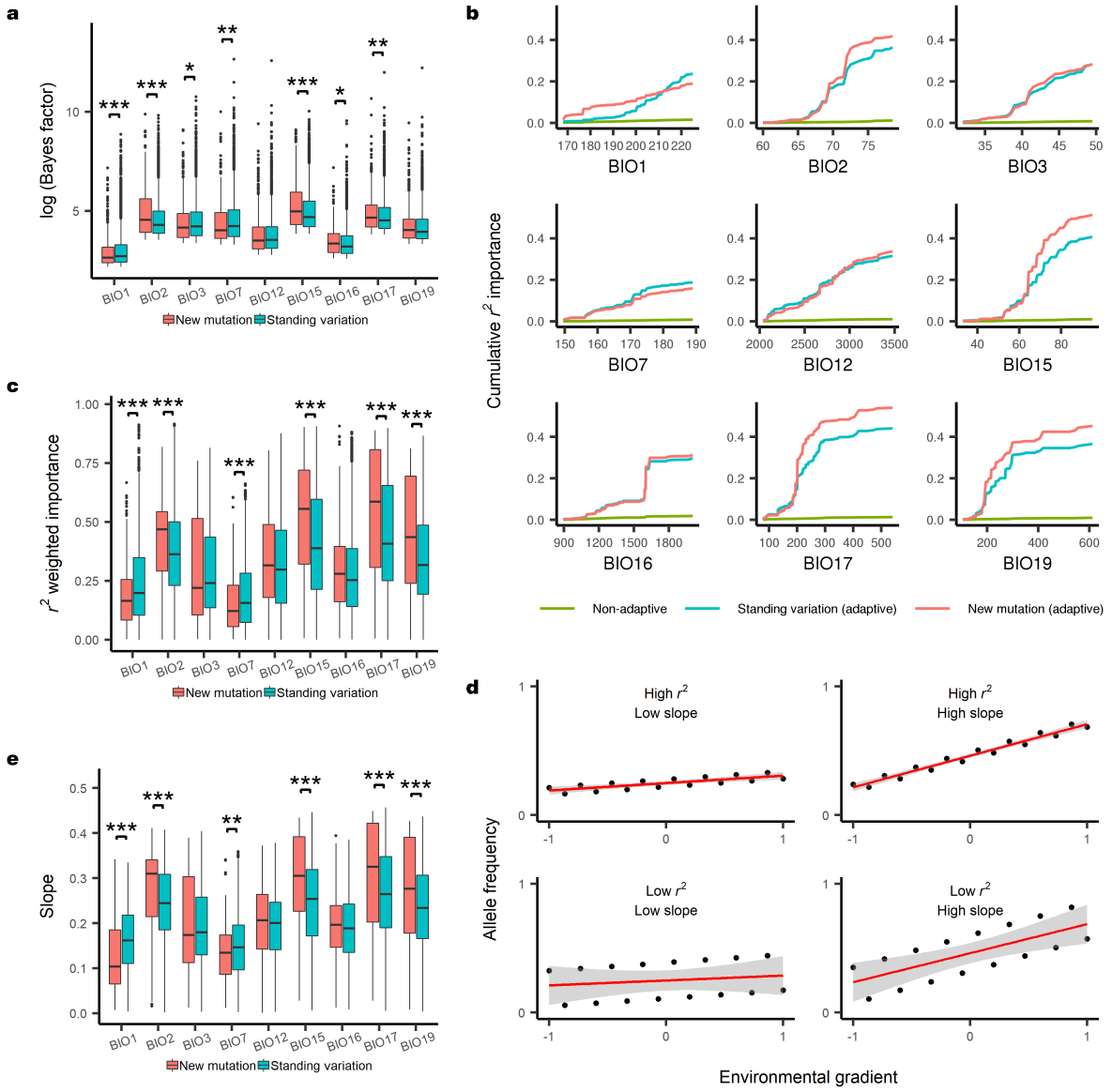



**Supplementary Fig. 1 | Sampling site and population structure of Taiwanese *Musa itinerans* under STRUCTURE *K* = 2. a**, Solid circles represent collection locations corresponding to their coordinates and elevation: Gray circles indicate populations of low elevation (LE); blue circles indicate populations of Xitou transect (XT); red circles indicate populations of Taipingshan transect (TPS). **b**, Individual ancestry is plotted on the right side, while population ancestry is plotted on map with a pie chart. Map template is provided by *Cheng-Tao Lin.

*Cheng-Tao Lin (2018) QGIS template for displaying species distribution by horizontal and vertical view in Taiwan. URL: https://github.com/mutolisp/distrmap_tw.qgis. DOI: 10.5281/zenodo.1493690

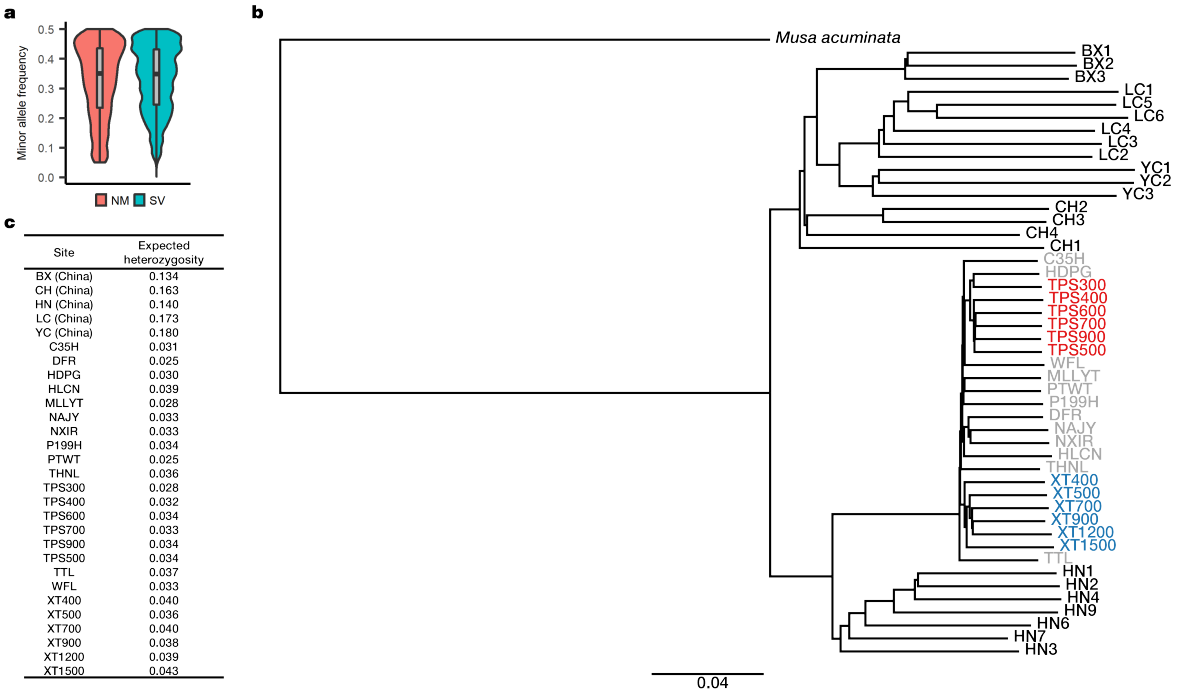



**Supplementary Fig. 2 | Allele frequency distribution. a**, Minor allele frequency distribution of adaptive and non-adaptive SNPs. **b**, Frequency of new alleles in Taiwanese populations within the ancestral or novel environmental range. Statistical significance from Student's *t*-test between the novel and ancestral environmental range for each bioclimatic variable is shown. Values on the horizontal axis denote the number of Taiwanese populations within the ancestral or novel environmental range.




671

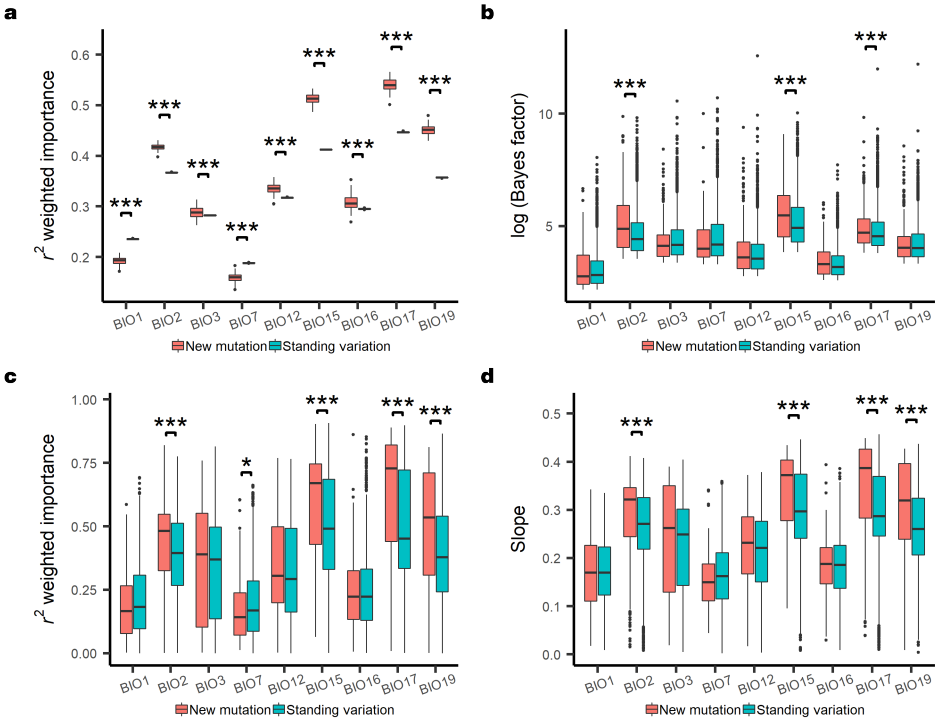
672 **Supplementary Fig. 3 | Profiling of effect sizes. a**, Bayenv Bayes factor distribution for
673 adaptive new mutations and standing variation (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, Wilcoxon
674 rank-sum test). **b**, Cumulative $r^2$ importance from gradient forest along environmental
675 gradients. **c**, Gradient forest $r^2$ importance distribution (***$P < 0.001$, Welch's $t$-test). **d**,
676 Example figure showing the relationship between $r^2$ importance and slope. $r^2$ indicates the
677 extent that allele frequency fits a linear model, while slope indicates the amount of allele
678 frequency changes along the linear relationship. The graphs indicate one should also investigate
679 regression slopes in addition to the gradient forest $r^2$. Values on the horizontal axis show the
680 range of standardized environmental variables. **e**, The distribution of regression slopes when
681 one regresses adaptive SNP allele frequency onto environmental gradients (**$P < 0.01$, ***$P$
682 $< 0.001$, Welch's $t$-test).
683




c

| Site | Expected heterozygosity |
|---|---|
| BX (China) | 0.134 |
| CH (China) | 0.163 |
| HN (China) | 0.140 |
| LC (China) | 0.173 |
| YC (China) | 0.180 |
| C35H | 0.031 |
| DFR | 0.025 |
| HDPG | 0.030 |
| HLCN | 0.039 |
| MLLYT | 0.028 |
| NAJY | 0.033 |
| NXIR | 0.033 |
| P199H | 0.034 |
| PTWT | 0.025 |
| THNL | 0.036 |
| TPS300 | 0.028 |
| TPS400 | 0.032 |
| TPS600 | 0.034 |
| TPS700 | 0.033 |
| TPS900 | 0.034 |
| TPS500 | 0.034 |
| TTL | 0.037 |
| WFL | 0.033 |
| XT400 | 0.040 |
| XT500 | 0.036 |
| XT700 | 0.040 |
| XT900 | 0.038 |
| XT1200 | 0.039 |
| XT1500 | 0.043 |

**Supplementary Fig. 4 | Evolutionary analyses on Taiwanese and Chinese *Musa itinerans*.** **a**, Minor allele frequency distribution of adaptive SNPs in Taiwan. **b**, Phylogeny of Taiwanese and Chinese *Musa itinerans*. The gray-colored indicates Taiwanese lowland populations (LE); the blue-colored indicates Taiwanese populations of Xitou transect (XT); the red-colored indicates Taiwanese populations of Taipingshan transect (TPS); the black-colored indicates Chinese accessions and the outgroup *Musa acuminata*. **c**, Table showing the expected heterozygosity among Taiwanese and Chinese populations.




**Supplementary Fig. 5 | Multi-analyses addressing ascertainment of standing variation and new mutation. a**, The distribution of gradient forest $r^2$ importance across 100 re-sampling trials. In each re-sampling trials, a random set of 50% new mutations were assigned as standing variation, and the mean $r^2$ importance was reported for each trial. All comparisons show strong statistical significance from Student's *t*-test between new mutations and introduced standing variation (***$P < 0.001$). **b-d**, Profiling of effect sizes of adaptive SNPs with top 50% minor allele frequency. Distribution of Bayes factor (***$P < 0.001$, Wilcoxon rank-sum test; **b**), gradient forest $r^2$ importance (*$P < 0.05$, ***$P < 0.001$, Welch's *t*-test; **c**), and regression slopes of adaptive SNP allele frequency onto environmental gradients (***$P < 0.001$, Welch's *t*-test; **d**) is analyzed between new mutations and standing variation for each bioclimatic variable.

705 **Supplementary Table 1 | Population coordinates and bioclimatic information**

| Site | Latitude | Longitude | BIO1 | BIO2 | BIO3 | BIO7 | BIO12 | BIO15 | BIO16 | BIO17 | BIO19 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TPS300 | 24.60488 | 121.5359 | 208 | 63 | 35 | 180 | 2,970 | 46 | 1,223 | 364 | 364 |
| TPS400 | 24.58135 | 121.5124 | 194 | 65 | 36 | 177 | 2,769 | 50 | 1,145 | 301 | 301 |
| TPS500 | 24.56765 | 121.4992 | 199 | 65 | 36 | 179 | 2,763 | 50 | 1,142 | 299 | 299 |
| TPS600 | 24.55172 | 121.5090 | 193 | 65 | 36 | 176 | 2,709 | 51 | 1,124 | 289 | 289 |
| TPS700 | 24.54638 | 121.5125 | 185 | 65 | 37 | 174 | 2,699 | 49 | 1,107 | 302 | 302 |
| TPS900 | 24.54021 | 121.5126 | 178 | 66 | 38 | 173 | 2,757 | 46 | 1,110 | 336 | 336 |
| XT400 | 23.74292 | 120.7386 | 217 | 77 | 42 | 180 | 2,468 | 86 | 1,395 | 87 | 127 |
| XT500 | 23.72780 | 120.7641 | 206 | 77 | 44 | 173 | 2,600 | 84 | 1,420 | 97 | 141 |
| XT700 | 23.71137 | 120.7782 | 199 | 77 | 45 | 169 | 2,412 | 83 | 1,311 | 94 | 138 |
| XT900 | 23.69294 | 120.7857 | 189 | 77 | 47 | 162 | 2,357 | 83 | 1,287 | 97 | 145 |
| XT1200 | 23.67333 | 120.7854 | 176 | 77 | 48 | 159 | 2,478 | 82 | 1,346 | 97 | 148 |
| XT1500 | 23.66936 | 120.7710 | 160 | 78 | 48 | 160 | 2,637 | 81 | 1,402 | 98 | 149 |
| C35H | 24.69907 | 121.1528 | 202 | 65 | 34 | 188 | 2,500 | 55 | 1,073 | 192 | 289 |
| WFL | 25.06752 | 121.7998 | 203 | 58 | 31 | 185 | 3,440 | 29 | 1,229 | 606 | 749 |
| THNL | 24.08273 | 120.8021 | 212 | 74 | 40 | 184 | 2,188 | 83 | 1,220 | 78 | 138 |
| PTWT | 22.58684 | 120.6489 | 227 | 78 | 49 | 158 | 2,912 | 98 | 1,784 | 79 | 87 |
| P199H | 22.24302 | 120.8465 | 223 | 73 | 50 | 144 | 3,114 | 92 | 1,856 | 178 | 178 |
| MLLYT | 24.33526 | 120.7864 | 211 | 71 | 37 | 190 | 2,056 | 76 | 1,084 | 79 | 148 |
| HDPG | 24.88347 | 121.5019 | 197 | 62 | 33 | 185 | 3,201 | 37 | 1,201 | 475 | 475 |
| TTL | 23.45230 | 120.6134 | 215 | 79 | 45 | 174 | 3,515 | 98 | 2,154 | 81 | 123 |
| NAJY | 24.43352 | 121.7602 | 216 | 62 | 35 | 174 | 2,860 | 61 | 1,328 | 271 | 296 |
| HLCN | 23.90623 | 121.4928 | 213 | 68 | 40 | 168 | 2,127 | 46 | 849 | 251 | 255 |
| NXIR | 23.40887 | 121.4503 | 227 | 71 | 43 | 163 | 2,147 | 57 | 938 | 211 | 211 |
| DFR | 23.10356 | 121.2724 | 206 | 75 | 48 | 155 | 2,017 | 68 | 1,026 | 166 | 166 |

706

707

708 **Supplementary Table 2 | SSR primer information**

| Locus | Primer (5' to 3') | Chromosome* | Position* |
|---|---|---|---|
| CIR436 | ATAAGCTCATATGGGTACAGTCACA<br>CTGCAGCAACCAAATTTATTTCT | 1 | 4,917,979 - 4,918,003<br>4,918,057 - 4,918,080 |
| CIR276 | CTCCTCCATAGCCTGACTGC<br>TGACCCACGAGAAAAGAAGC | 1 | 13,588,759 - 13,588,778<br>13,588,847 - 13,588,866 |
| CIR1113 | ACTCTCGCCCATCTTCATCC<br>ACTTATTCCCCCGCACTCAA | 1 | 28,009,948 - 28,009,967<br>28,010,173 - 28,010,192 |
| Ma-3-132 | TCCCTCTTCAACCAAAGCAC<br>AACGCGAATGTGTGTTTTCA | 2 | 20,365,901 - 20,365,920<br>20,366,041 - 20,366,060 |
| CIR646 | AACACCGTACAGGGAGTCAC<br>GATACATAAGGCAGTCACATTG | 2 | 23,940,966 - 23,940,985<br>23,941,276 - 23,941,297 |
| Ma-1-17 | AGGCGGGGAATCGGTAGA<br>GGCGGGAGACAGATGGAGT | 2 | 24,383,391 - 24,383,408<br>24,383,488 - 24,383,506 |
| CIR332a | ATGACCTGTCGAACATCCTTT<br>TCCCAACCCCTGCAACCACT | 3 | 8,978,575 - 8,978,595<br>8,978,831 - 8,978,848 |
| Ma-3-48 | CCCGTCCCATTTCTCA<br>TTCGTTGTTCATGGAATCA | 5 | 32,672,883 - 32,672,898<br>32,673,018 - 32,673,036 |
| CIR631a | ATTAGATCACCGAAGAACTC<br>ATCTTTTCTTATCCTTCTAACG | 6 | 33,942,751 - 33,942,770<br>33,943,017 - 33,943,038 |
| CIR16a | TCATCTCACAATGCTTTCATAGTT<br>TGGTTGAGTAGATCTTCTTGTGT | 8 | 1,234,609 - 1,234,632<br>1,234,700 - 1,234,722 |
| Ma-3-103 | TCGCCTCTCTTTAGCTCTG<br>TGTTGGAGGATCTGAGATTG | 8 | 40,127,782 - 40,127,800<br>40,127,910 - 40,127,929 |
| CIR348b | ACAGAATCGCTAACCCTAATCCTCA<br>CCCTTTGCGTGCCCCTAA | 10 | 27,228,162 - 27,228,186<br>27,228,325 - 27,228,342 |
| Ma-3-139 | ACTGCTGCTCTCCACCTCAAC<br>GTCCCCCAAGAACCATATGATT | 10 | 30,237,591 - 30,237,611<br>30,237,717 - 30,237,735 |
| CIR550a | ACCGCACCTCCACCTCCTG<br>TGCTGCCTTCATCGCTACTA | 10 | 31,914,219 - 31,914,237<br>31,914,459 - 31,914,478 |

709

710 Primer sequences were searched against the *Musa acuminata* DH-Pahang version 2[33] on
711 Banana Genome Hub (https://banana-genome-hub.southgreen.fr/).
712 *The chromosome and position indicate locations on *Musa acuminata* where primer sequences
713 were found by blast.
714

**Supplementary Table 3 | Number of new-mutation (NM) or standing-variation (SV) SNPs with (adaptive) or without (non-adaptive) significant associations with bioclimatic variables**

| Variable | Non-adaptive NM | Non-adaptive SV | Adaptive NM | Adaptive SV | *P* | Odds ratio |
|---|---|---|---|---|---|---|
| BIO1 | 293,348 | 954,367 | 816 | 7,113 | 8.92e-169 | 2.68 |
| BIO2 | 293,274 | 953,198 | 890 | 8,282 | 8.43e-213 | 2.86 |
| BIO3 | 293,551 | 953,901 | 613 | 7,579 | 6.47e-256 | 3.80 |
| BIO7 | 293,644 | 953,981 | 520 | 7,499 | 1.32e-282 | 4.44 |
| BIO12 | 293,452 | 954,272 | 712 | 7,208 | 3.12e-203 | 3.11 |
| BIO15 | 293,168 | 951,447 | 996 | 10,033 | 2.33e-281 | 3.10 |
| BIO16 | 293,826 | 957,605 | 338 | 3,875 | 1.95e-123 | 3.52 |
| BIO17 | 293,461 | 954,969 | 703 | 6,511 | 1.61e-166 | 2.85 |
| BIO19 | 293,516 | 955,326 | 648 | 5,142 | 1.06e-107 | 2.44 |

Statistical significance from $\chi^2$ test is shown for each bioclimatic variable. Odds ratio is calculated as (adaptive SV / adaptive NM) / (non-adaptive SV / non-adaptive NM).

722 **Supplementary Table 4 | Number of new-mutation (NM) or standing-variation (SV) SNPs**
723 **with (adaptive) or without (non-adaptive) significant associations with bioclimatic**
724 **variables (controlled for minor allele frequency)**

| Variable | Non-adaptive NM | Non-adaptive SV | Adaptive NM | Adaptive SV | *P* | Odds ratio |
|---|---|---|---|---|---|---|
| BIO1 | 44,920 | 260,113 | 238 | 2,573 | 1.27e-20 | 1.87 |
| BIO2 | 23,475 | 148,796 | 124 | 1,073 | 1.18e-03 | 1.37 |
| BIO3 | 39,432 | 234,613 | 182 | 2,689 | 2.21e-34 | 2.48 |
| BIO7 | 41,020 | 242,313 | 182 | 2,794 | 4.35e-38 | 2.60 |
| BIO12 | 37,266 | 224,455 | 220 | 2,514 | 3.36e-20 | 1.90 |
| BIO15 | 32,906 | 201,729 | 290 | 2,647 | 1.36e-10 | 1.49 |
| BIO16 | 33,685 | 206,816 | 113 | 1,282 | 2.88e-10 | 1.85 |
| BIO17 | 12,258 | 72,175 | 64 | 614 | 2.26e-04 | 1.60 |
| BIO19 | 11,237 | 66,547 | 41 | 362 | 1.81e-02 | 1.50 |

725
726 The number of four sets of SNPs whose minor allele frequency (MAF) ranges from the first
727 quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each
728 bioclimatic variable is shown. Statistical significance from $\chi^2$ test is shown for each
729 bioclimatic variable. Odds ratio is calculated as (adaptive SV / adaptive NM) / (non-adaptive
730 SV / non-adaptive NM).
731

**Supplementary Table 5 | Number of currently adaptive new-mutation (NM) or standing-variation (SV) SNPs that remain (retention) or lose (disruption) significant associations with environments under all future climate-change scenarios**

| Variable | Disruption NM | Disruption SV | Retention NM | Retention SV | *P* | Odds ratio |
|---|---|---|---|---|---|---|
| BIO1 | 92 | 635 | 127 | 1,176 | 5.11e-02 | 1.30 |
| BIO2 | 34 | 374 | 348 | 2,010 | 6.85e-04 | 0.53 |
| BIO3 | 21 | 269 | 119 | 2,277 | 1.32e-01 | 1.50 |
| BIO7 | 29 | 251 | 128 | 2,393 | 4.53e-04 | 2.20 |
| BIO12 | 65 | 481 | 136 | 1,509 | 1.37e-02 | 1.50 |
| BIO15 | 37 | 365 | 437 | 2,982 | 4.79e-02 | 0.69 |
| BIO16 | 26 | 289 | 25 | 237 | 6.93e-01 | 0.85 |
| BIO17 | 37 | 354 | 159 | 1,182 | 2.21e-01 | 0.78 |
| BIO19 | 28 | 235 | 45 | 480 | 4.14e-01 | 1.30 |

Statistical significance from $\chi^2$ test is shown for each bioclimatic variable. Odds ratio is calculated as (retention SV / retention NM) / (disruption SV / disruption NM).