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 from the past to infer their future fate in nature. In the wild banana species *Musa itinerans* where the Taiwanese populations were ancient immigrants from the Chinese populations, new mutations exert larger effect size in precipitation-related variables, where Taiwan contains extreme environments beyond the ancestral climatic range, and new alleles have stronger association with novel environments. For temperature-related variables where Taiwan is within the ancestral climatic range, standing variants are more important than new mutations. The effect sizes of adaptive variants therefore differ under distinct environmental pressures, supporting theoretical predictions that natural selection favors new mutations with larger effect sizes in novel environments where the population is far from the adaptive peak. Despite their importance, large-effect variants also have higher mismatch and may be more vulnerable to future environmental perturbation, leaving minor-effect variants the main source of adaptive response to rapid anthropogenic climate change. Our work provides a support in natural environment to the previous conclusions from theoretical modeling and microbial experiments in well-controlled lab conditions.

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 [1,2]. 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 [3]. It remains unclear, however, whether standing variation is sufficient to guarantee future survival, given that they were 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 [4] while Kimura emphasized intermediate effects [5]. 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 [6]. While this was supported by some studies, the majority of these are microbial experimental evolution in well-controlled environments [7,8], 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 [9]. 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 [10], 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; fig. S1a; table S1) and investigated the population structure using 14 microsatellites (table S2). 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 [11]. Populations differentiated mostly between east and west (fig. S1b). 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 [12]: 36.7 for BIO1 and 27.3 for BIO17).

To test for local adaptation, we examined the pattern of "isolation by adaptation" [13], 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 [14]) and non-adaptive SNPs, and after controlling for the overall number of SNPs in SV and NM, SV were further enriched among adaptive polymorphisms (table S3). However, since adaptive SNPs also have higher minor allele frequency (MAF) (fig. S2a), this pattern could be confounded: SV are more likely

107 to have higher MAF than NM, and SNPs with higher MAF may be more likely detected as
108 adaptive due to higher statistical power. We therefore performed the same test with a subset
109 where the adaptive and non-adaptive SNPs have similar allele frequencies (ranging from the
110 first quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each
111 bioclimatic variable; fig. S2a). In this case as well, SV are still disproportionately abundant
112 (table S4), suggesting SV are more likely than NM to become environment-associated SNPs.
113 Another potential confounding factor is the geographic extent of variants: if most NM resulted
114 from mutations restricted to a few local populations, the limited distribution prevents
115 environment association for NM. We therefore compared the number of Taiwanese populations
116 containing the minor alleles for SV and NM, respectively. Contrary to the direction predicted
117 by the aforementioned confounding factor, the geographic extent of minor alleles for SV is
118 slightly smaller than NM (15.8 populations for SV and 16.3 for NM, $P < 0.001$).
119 In addition to SNP number, do NM and SV differ in their directions of effect? Under the
120 null hypothesis that (I) the effects of NM are equally likely to facilitate adaptation to the
121 ancestral or novel environments and (II) natural selection is equally likely to fix NM facilitating
122 adaptation towards either direction, we expected no enrichment of new alleles in either
123 environment. When we separated the Taiwanese populations into those within the Chinese
124 ancestral environmental range and those with novel environments, frequencies of putatively
125 adaptive new alleles in NM SNPs were higher in the latter set of populations, with precipitation
126 of driest quarter (BIO17) and precipitation of coldest quarter (BIO19) showing the strongest
127 effect (Fig. 2a; fig. S2b). Given that the directions of mutation effects should be random, these
128 results suggest that new variants facilitating adaptation to novel environments are more likely
129 to be retained by selection.
130 The results above suggest SV might be more important than NM in terms of enriched
131 number of variants with environment association, and adaptive NM, while lower in number,
132 are more associated with novel environments. On the other hand, the number of candidate SNPs
133 does not necessarily reflect the overall importance of SV over NM, since the effect size also
134 needs to be considered. To investigate the effect size of SV and NM in environmental
135 adaptation, we compared their Bayes factors from Bayenv and focused on two bioclimatic
136 variables that were important in determining species distribution, annual mean temperature
137 (BIO1) and precipitation of driest quarter (BIO17). The two variables exhibit opposite patterns
138 (Fig. 2c): While in BIO17 and other related precipitation variables (fig. S3a) NM consistently
139 had higher Bayes factor and therefore stronger effect size than SV, in BIO1 and other related
140 temperature variables (fig. S3a) we observed the reverse. The same trend was observed when
141 we estimated the effect size with gradient forest [15,16] (Fig. 2b and 2c; fig. S3b and S3c):
142 BIO17 was the most important factor for differential local adaptation, and NM had stronger
143 effects than SV. On the other hand, BIO1 was the least important factor where SV had stronger
144 effects. Finally, the "importance" estimated by gradient forest is analogous to r^2, representing

145 the amount of allele frequency variation explained by environmental gradients. Assuming a
146 simple linear relationship between allele frequency and environment, the value of r^2 only
147 represents how well each data point (a population) fits along the regression line. We were,
148 however, also interested in the regression slope: the amount of allele frequency changes along
149 environmental gradients (fig. S3d). Again, BIO17 had the largest overall slope among all
150 bioclimatic variables (fig. S3e), with NM being significantly higher than SV (Fig. 2c). BIO1
151 had the lowest overall slope, again with the reversed pattern. Given that the MAF of adaptive
152 NM and SV SNPs are similar, there is no need to control for allele frequency in these tests (fig.
153 S4a). Therefore, NM with larger effect size per SNP (as estimated by Bayenv Bayes factor,
154 gradient forest r^2 weighted importance, and gradient forest slope) were associated with the
155 adaptation to novel environments outside of the ancestral niche space, consistent with previous
156 population genetics modeling results [4-6].
157 The observed patterns could be integrated with the unique climate of Taiwan. In
158 comparison to the rest of the species range, northern Taiwan experiences northeastern
159 monsoons during winter and has higher precipitation during the typical winter dry season (Fig.
160 3b). This pattern has been maintained since at least the last glacial maximum (Fig. 3d). The
161 novel environments might impose novel adaptive optima to the immigrant population from
162 China. The response to selection imposed by these environmental gradients is strong (with
163 highest Bayes factor, r^2, and slope among all bioclimatic variables; fig. S3a, S3c, and S3e)
164 especially for NM, where new alleles are strongly associated with novel environments (fig.
165 S2b). More importantly, for this major driver of adaptation (BIO17), the greatest increment of
166 gradient forest importance lies between 200 mm and 300 mm (Fig. 2b), a range also
167 distinguishing the novel Taiwanese versus ancestral Chinese environments (Fig. 3b). This
168 suggests that the majority of differential local adaptation is associated with such novel-versus-
169 ancestral environmental differences. Annual mean temperature (BIO1) is the other extreme:
170 the environmental gradient within Taiwan is well within the ancestral Chinese environmental
171 range (Fig. 3a), which can also be traced back to the last glacial maximum (Fig. 3c). It is likely
172 that SV already contained genetic variants adaptive to such environmental gradients and are
173 therefore more important than NM (Fig. 2c). In summary, we observed adaptation happening
174 through the assortment of SV for a new territory with similar adaptive landscape and optimum
175 (BIO1; Fig. 2c, 3a, and 3c). For adaptation to novel environments and a new adaptive landscape
176 (BIO17; Fig. 2c, 3b, and 3d), NM with larger effect sizes were more likely favored by natural
177 selection. Our results are therefore consistent with Orr's model[3], providing one of the few
178 examples in nature.
179 One key point of this study is the correct designation of SV or NM. It is possible that some
180 SV SNPs were mis-assigned as NM if we missed an allele in China, most likely for SNPs with
181 low MAF. We addressed this issue with the following: (I) The Taiwanese populations were
182 nested within China in the phylogenetic tree (fig. S4b) and contained much less genetic

183 variation (fig. S4c). Due to the stronger genetic drift in Taiwan than China, it is less likely that
184 an originally SV SNP would retain both alleles in Taiwan but lose one in China. (II) In the
185 extreme case, assuming 50% of NM SNPs were mis-assigned from SV, we performed 100 new
186 analyses, each randomly assigning 50% of NM SNPs back to SV. These new analyses yielded
187 similar results, with NM having higher effect sizes than SV in precipitation-related variables
188 (fig. S5a). (III) Since MAF are correlated between Taiwan and China (Spearman's rank
189 correlation $\rho = 0.35$, $P < 0.001$), we performed analyses with top 50% MAF SNPs in Taiwan,
190 thereby reducing the chance of missing minor alleles in China. The results are qualitatively the
191 same (fig. S5b-d).
192
193 **Fate of adaptive variants under future climate change**
194 In addition to understanding how past events (SV versus NM) affected present adaptation,
195 we are also concerned with how these factors affect the future of this species under
196 anthropogenic climate change. We used Bayenv to investigate the fate of currently adaptive
197 SNPs under 16 different future climatic scenarios. Currently adaptive SNPs retaining high
198 association with future environments are classified as "retention", while those no longer
199 associated with future environments are "disruption". Different from the present pattern of SV
200 being enriched in adaptive SNPs, we saw no clear tendency for SV or NM enriched towards
201 retention or disruption (table S5), suggesting both types of genetic variants will be affected by
202 climate change regardless of their previous origin.
203 We further used the genetic offset values from gradient forest to estimate genetic
204 mismatch for SV and NM separately, which is associated with the magnitude of allele
205 frequency turnover perturbed by future climatic conditions [16]. After projecting the genetic
206 mismatch of SV and NM on the map of Taiwan for all 16 predicted future scenarios, we found
207 SV generally has higher mean (paired t-test, $t = 3.296$, $P = 0.005$, Fig. 4A) while NM has higher
208 maximum genetic offset across Taiwan (paired t-test, $t = -8.608$, $P < 0.001$, Fig. 4B). We further
209 averaged the genetic offset map across all 16 future scenarios (Fig. 4C-E). There are more
210 geographical grids with SV offset higher than NM (12,711 square kilometers) and less grids
211 with NM offset higher than SV (8,707 square kilometers), consistent with the higher mean
212 genetic offset for SV than NM. On the other hand, the most extreme genetic offset values were
213 often observed for NM rather than SV (Fig. 4D-E).
214 Geographic regions where NM has much higher genetic offset (maladaptation towards
215 anthropogenic climate change) than SV include northwestern mountainous and eastern coastal
216 regions. Interestingly, these regions appear to have BIO17 values (precipitation of driest
217 quarter) close to 200, where NM has much sharper transition of adaptive allele frequency than
218 SV (Fig. 2B and 3B). To further examine such association, we cut the Taiwan map into 20 bins
219 of equal grid numbers based on their current BIO17 value (excluding grids with species-wide
220 Maxent suitability < 0.2) and estimated the mean genetic offset values of grids within each bin.

Indeed, NM has much higher genetic offset in regions with BIO17 value near 200 (Fig. 4F). Such pattern exists not only for BIO17 but also for many other bioclimatic variables (fig. S6) where genetic variants (either SV or NM) tend to have higher genetic offset in environments with sharp increase in gradient forest's cumulative importance (high delta importance per unit environmental gradient in fig. S6), indicating populations closer to the threshold of sharp adaptive allele frequency change might be more vulnerable to climate perturbation.

Taken together, for adaptation in the past, NM is relatively less enriched in number but has higher effect size per SNP. For maladaptation towards the future, here we show a similar pattern: NM has weaker genetic offset than SV in the majority of geographic regions, but NM tends to have the overall highest genetic offset values in locations near the sharp transition of gradient forest's cumulative importance. In other words, while NM has higher effect size and sharper allele frequency transition for adaptation in the past to novel environments outside of the ancestral range, such property may also make NM more vulnerable to climate change in the future, especially near the transition zone.

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 [17-22]. 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 [6]. 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 [21,22]. We show that SV indeed dominate over NM in number [21]. 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. Looking towards the future, our results imply that standing genetic variation in a population may not be sufficient for the adaptation to

rapid anthropogenic climate change.

Projecting into the future, we observed similar patterns. Both SV and NM could be affected by anthropogenic climate change depending on their genetic architecture of climatic adaptation: SV has slightly larger genetic offset (maladaptation to anthropogenic climate change) than NM in the majority of geographic regions, but the largest genetic offset values were mostly observed for NM SNPs in locations near the sharp transition of gradient forest's cumulative importance. On one hand, the higher effect sizes of NM SNPs suggest their importance in past adaptation towards novel environments. On the other, the high effect size is accompanied by sharp allele frequency transitions across environmental gradients, making NM SNPs highly sensitive (or even more vulnerable) to environmental perturbations in the future. Our results also suggest that, under frequent environmental changes such as repeated glacial cycles, although natural selection might favor new mutations with higher effect sizes under novel environments, such genetic variants might more likely be mal-adaptive during the next round of environmental change due to their high effect sizes. In other words, adaptive variants of high effect sizes may be either strongly favored or disfavored by different rounds of environmental changes, making these high-effect variants less likely to be retained as the standing variation for future. This is consistent with previous notions that genetic variants with higher effect sizes might be more likely to "overshoot" the adaptive peak [4] and may partially contribute to the observations that most adaptations have polygenic architecture [23]. The results also echo a recent experimental evolution study showing non-synonymous mutations are more difficult to be retained in antagonistic changing environments [24]. For the future fate of organisms living in natural environments, further large-effect 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 (fig. S1a; table S1). Fresh leaves were harvested from nine to fifteen individuals at each site. Total genomic DNA was extracted using the standard CTAB extraction method [25]. Since other commercial *Musa* species were also grown in Taiwan, we developed an indel marker for species delimitation. From previous studies [26-29], 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* [30,31], which were then applied on *Musa itinerans*. Previously documented primer sequences were first searched against the *Musa acuminata* DH-Pahang genome version 2 [32] 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 (table S2). We modified each pair of primers by capping the 5' end of forward primers with M13 sequences (CACGACGTTGTAAAACGAC) and inflorescent molecules [33]. SSR amplicons were run through capillary electrophoresis and the length of each allele was recorded.

Population structure of 20 populations (table S1) was analyzed with 14 SSR markers (table S2). 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 [34,35], parameterizing a run to have (I) run length of burnin and after-burnin period of 100,000, (II) admixture ancestry model, and (III) 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 [36,37] 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 (I) genetic distance *vs*. fly-over distance, (II) genetic distance *vs*. resistance distance, (III) genetic distance *vs*. Mahalanobis environmental distance, and Partial Mantel tests on (IV) 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 [12], implemented with the maximum entropy

335 modeling approach, reports an overall niche suitability and the importance of predictors by
336 analyzing the presence-only data as well as background (psuedo-absence) data distribution [38-
337 40]. We downloaded from WorldClim database version 1.4 (http://worldclim.org/) spatial
338 layers of 19 present-day bioclimatic variables based on high-resolution monthly temperature
339 and rainfall data [41]. Layers were selected at spatial resolution of 30 arc-second and with a
340 mask that ranges 119.25-122.47°E and 21.76-25.49°N covering Taiwan. Variables showing
341 high dependence (Pearson's correlation coefficient > 0.9 calculated from ENMTools [42]) from
342 each other were removed, resulting in nine final variables: BIO1—mean annual temperature,
343 BIO2—mean diurnal range, BIO3—isothermality, BIO7—temperature annual range, BIO12
344 —annual precipitation, BIO15—precipitation seasonality, BIO16—precipitation of wettest
345 quarter, BIO17—precipitation of driest quarter, and BIO19—precipitation of coldest quarter
346 (table S1).
347 Present species distribution model was constructed using the default optimization settings
348 in MaxEnt, except the regularization set to three. We tested the predictive model by ten-fold
349 cross-validation which was carried out by randomly partitioning the data into ten equally sized
350 subsets and then replicating models while omitting one subset in turn. In each turn, the
351 predictive model was built using nine subsets as training data and evaluated using the other
352 subset as test data. The output of the predictive model is the probability of presence, or called
353 suitability, and we averaged the ten runs to have an averaged suitability.
354 To predict the species distribution under different scenarios of future climate change, we
355 projected the present-day model onto eight future climatic conditions combining two periods
356 (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0,
357 and RCP 8.5). Future climatic layers were obtained from the WorldClim database at spatial
358 resolution of 30 arc-second and were developed based on two general circulation models: the
359 Community Climate System Model [43], CCSM, and the Model for Interdisciplinary Research
360 on Climate [44], MIROC. Species distribution models for the future were carried out using the
361 same settings described above.
362 To estimate the least cost path between populations, we first generated the resistance
363 surface by taking the reciprocal of suitability. Resistance and suitability is simply a monotonic
364 transformation in which locations with higher suitability exhibit lower resistance. Pairwise
365 least cost path was then measured among 20 populations from the resistance surface, performed
366 by SDM Toolbox v2.3 [45]. While least cost path is the single line with least overall cost, we
367 also constructed the least cost corridor between populations, allowing 1%, 2%, or 5% higher
368 cost than the least cost value. In essence, the least cost corridors represent the realized dispersal
369 routes of organisms along suitable habitats.
370
371 **Sequencing library construction and SNP identification.**
372 We conducted whole genome pooled-sequencing [46] for each population (table S1),

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 [47], followed by the removal of adaptor sequences with cutadapt [48], subsequently mapped to the *Musa itinerans* reference genome assembly ASM164941v1 [49] with BWA 0.7.15 [50]. 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 [51].

For the 24 pooled samples, we filtered out sites with more than two alleles, indels, quality (QUAL) < 30, quality by depth (QD) < 2, call rate < 0.74, and depth (DP) > genome-wide average depth plus three standard deviations, resulting in 4,200,177 SNPs. SNPs with minor allele frequency (MAF) < 0.05, missing data in any of the pooled-seq sample, and 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) [52]. 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 indels and 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 more than two alleles, indels, QUAL < 30, and call rate < 0.9, resulting in 12,693,687 SNPs. This dataset also excluded SRR6382532.

Environmentally-associated SNP identification.

We used Bayenv 2.0 [14] 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 one SNP from scaffolds more than 10 kb and less than 100 kb, two SNPs from scaffolds more than 100 kb and less than 500 kb, three SNPs from scaffolds more than 500 kb and less than 1000 kb, and four SNPs from scaffolds more than 1000 kb. We then, for each bioclimatic variable, defined as the adaptive

410 SNPs ones exhibiting top 1% Bayes factor and top 5% rho value (a nonparametric correlation
411 coefficient capable to reduce outlier effects).
412 We further investigated the fate of currently adaptive SNPs under anthropogenic climate
413 change, performing the same Bayenv analyses of currently adaptive SNPs using future climatic
414 conditions. We included two time periods (2050 and 2070) and four Representative
415 Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) from two general
416 circulation models, CCSM [43] and MIROC [44], resulting in 16 future climatic conditions. If
417 a currently adaptive SNP remains strongly associated with environments, it should exhibit
418 Bayes factor above the current threshold. We then defined as "retention" a currently adaptive
419 SNP constantly exhibiting Bayes factor above the current adaptive threshold in all future
420 scenarios, and defined as "disruption" a currently adaptive SNP exhibiting Bayes factor above
421 the current adaptive threshold in none of the future scenarios.
422
423 **The gradient forest method and genetic offset.**
424 We used a novel method, gradient forest [15,16], to estimate the effect of environmental
425 gradients on allele frequency differences among populations. Gradient forest is a regression-
426 tree based machine-learning algorithm using environmental variables to partition SNP allele
427 frequencies. The analysis was done separately for each SNP. The resulting "importance"
428 measures how much of the variation in allele frequency was explained by partitioning the
429 populations based on a specific value in an environmental variable. By making multiple
430 regression trees (thus generating a random forest) for a SNP, the goodness-of-fit r^2 of a random
431 forest is measured as the proportion of variance explained by this random forest, which is then
432 partitioned among environmental variables in proportion to their conditional importance. Such
433 SNP-wise importance of each environmental variable is then averaged across SNPs belonging
434 to the standing variation (SV) or new mutations (NM) group, resulting in the overall importance
435 (of each environmental variable). In the end, one could obtain the relation curve between
436 environmental gradient and cumulative importance (analogous to the cumulative r^2, proportion
437 of allele frequency differences among populations explained by environments). This curve has
438 two properties. First, the highest point of the cumulative importance curve denotes the overall
439 association between a climatic variable and allele frequency, and we used this to represent the
440 effect size of these SNPs. Second, when traversing along the environmental gradient, a sudden
441 increase of cumulative importance at some environmental value (for example, 20°C) denotes
442 populations on either side of this environmental cutoff have very different allele frequency
443 compositions. In other words, this represents a threshold factor for local adaptation.
444 One can use this cumulative importance curve to estimate the effect of future
445 environmental change on local populations. In the example above, a population's local
446 temperature increased from 19°C to 21°C due to climate change would require larger allele
447 frequency shift than another population whose local temperature changed from 17°C to 19°C.

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

References:

1. Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive Versus Non-Adaptive Phenotypic Plasticity and the Potential for Contemporary Adaptation in New Environments. Functional Ecology 21: 394-407.
2. Jump AS, Peñuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters 8: 1010-1020.
3. Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends Ecol Evol 23: 38-44.
4. Fisher RA (1930) The genetical theory of natural selection. Oxford, England: Clarendon Press. xiv, 272-xiv, 272 p.
5. Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press.
6. Orr HA (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution 52: 935-949.
7. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461: 1243-1247.
8. Gifford DR, Schoustra SE, Kassen R (2011) The Length of Adaptive Walks Is Insensitive to Starting Fitness in Aspergillus Nidulans. Evolution 65: 3070-3078.
9. Bramwell D, Caujapé-Castells J (2011) The Biology of Island Floras. Cambridge: Cambridge University Press.
10. Janssens SB, Vandelook F, De Langhe E, Verstraete B, Smets E, et al. (2016) 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.
11. Chiu HL (2004) The collection, evaluation and analysis of genetic diversity of Musa formosana (Warb.) Hayata native in Taiwan [thesis]: National Taiwan University. 1-180 p.
12. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.
13. Nosil P, Egan SP, Funk DJ (2008) HETEROGENEOUS GENOMIC DIFFERENTIATION BETWEEN WALKING-STICK ECOTYPES: “ISOLATION BY ADAPTATION” AND MULTIPLE ROLES FOR DIVERGENT SELECTION. Evolution 62: 316-336.
14. Gunther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. Genetics 195: 205-220.
15. Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests: calculating importance gradients on physical predictors. Ecology 93: 156-168.
16. Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling

502 of biodiversity: mapping the genomic landscape of current and future environmental
503 adaptation. Ecol Lett 18: 1-16.
504 17. Barton NH (2000) Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci 355: 1553-
505 1562.
506 18. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an
507 adaptive allele in deer mice. Science 325: 1095-1098.
508 19. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. (2012) The genomic basis
509 of adaptive evolution in threespine sticklebacks. Nature 484: 55-61.
510 20. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, et al. (2016) The genomic
511 landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish.
512 Science 354: 1305-1308.
513 21. Lai YT, Yeung CKL, Omland KE, Pang EL, Hao Y, et al. (2019) Standing genetic
514 variation as the predominant source for adaptation of a songbird. P Natl Acad Sci
515 USA 116: 2152-2157.
516 22. Ahrens CW, Byrne M, Rymer PD (2019) Standing genomic variation within coding and
517 regulatory regions contributes to the adaptive capacity to climate in a foundation tree
518 species. Molecular Ecology 28: 2502-2516.
519 23. Rockman MV (2012) The QTN program and the alleles that matter for evolution: all
520 that's gold does not glitter. Evolution 66: 1-17.
521 24. Chen P, Zhang J (2020) Antagonistic pleiotropy conceals molecular adaptations in
522 changing environments. Nature ecology & evolution 4: 461-469.
523 25. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh
524 leaf tissue. Phytochemical Bulletin 19(1):11-15.
525 26. White TJ, Bruns T, Lee S, Taylor J (1990) 38 - AMPLIFICATION AND DIRECT
526 SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR
527 PHYLOGENETICS. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR
528 Protocols. San Diego: Academic Press. pp. 315-322.
529 27. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of
530 three non-coding regions of chloroplast DNA. Plant molecular biology 17: 1105-1109.
531 28. Oxelman B, Lidén M, Berglund D (1997) Chloroplast rps16 intron phylogeny of the tribe
532 Sileneae (Caryophyllaceae). Plant Systematics and Evolution 206: 393-410.
533 29. Hurr KA, Lockhart PJ, Heenan PB, Penny D (1999) Evidence for the recent dispersal of
534 Sophora (Leguminosae) around the Southern Oceans: molecular data. Journal of
535 Biogeography 26: 565-577.
536 30. Lagoda PJ, Noyer JL, Dambier D, Baurens FC, Grapin A, et al. (1998) Sequence tagged
537 microsatellite site (STMS) markers in the Musaceae. Mol Ecol 7: 659-663.
538 31. Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R (1998) Segregation at
539 microsatellite loci in haploid and diploid gametes of Musa. Crop Sci 38: 211-217.

32. Martin G, Baurens FC, Droc G, Rouard M, Cenci A, et al. (2016) Improvement of the banana "Musa acuminata" reference sequence using NGS data and semi-automated bioinformatics methods. BMC genomics 17: 243.
33. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18: 233-234.
34. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.
35. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.
36. Peakall ROD, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.
37. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28: 2537-2539.
38. Bean WT, Stafford R, Brashares JS (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography 35: 250-258.
39. Elith* J, H. Graham* C, P. Anderson R, Dudík M, Ferrier S, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.
40. Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, et al. (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43-57.
41. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965-1978.
42. Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 10.1111/j.1600-0587.2009.06142.x.
43. Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, et al. (2011) The Community Climate System Model Version 4. J Climate 24: 4973-4991.
44. Watanabe S, Hajima T, Sudo K, Nagashima T, Takemura T, et al. (2011) MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. Geosci Model Dev 4: 845-872.
45. Brown JL, Anderson B (2014) SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods in Ecology and Evolution 5: 694-700.
46. Schlotterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet 15: 749-763.
47. Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of

578 Illumina second-generation sequencing data. Bmc Bioinformatics 11.
579 48. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing
580 reads. EMBnetjournal 17: 10-12.
581 49. Wu W, Yang YL, He WM, Rouard M, Li WM, et al. (2016) Whole genome sequencing of
582 a banana wild relative Musa itinerans provides insights into lineage-specific
583 diversification of the Musa genus. Sci Rep 6: 31586.
584 50. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
585 transform. Bioinformatics 25: 1754-1760.
586 51. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome
587 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
588 sequencing data. Genome Res 20: 1297-1303.
589 52. Wu W, Ng WL, Yang JX, Li WM, Ge XJ (2018) High cryptic species diversity is revealed
590 by genome-wide polymorphisms in a wild relative of banana, Musa itinerans, and
591 implications for its conservation in subtropical China. BMC Plant Biol 18: 194.
592
593

Figures:

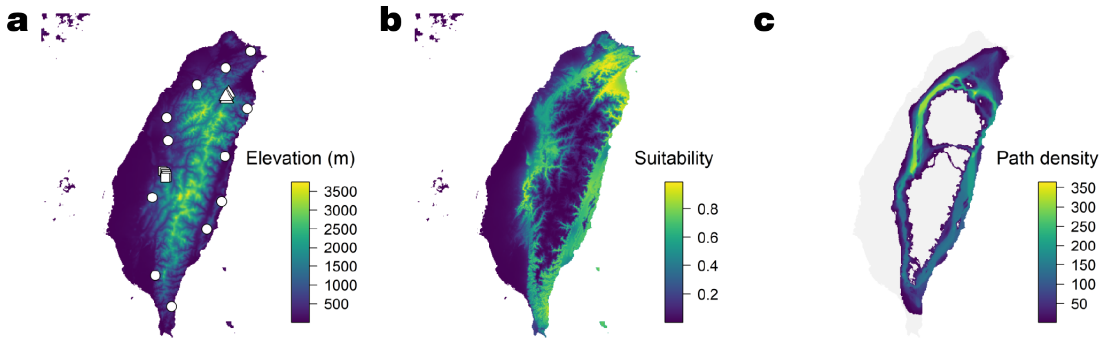


Fig. 1 | Sample distribution and niche modeling of *Musa itinerans*. a, Sampling sites are distributed along the latitudinal and altitudinal gradient. Lowland populations are represented as white circles; transect populations are represented as white triangles and squares. **b**, Suitability is derived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

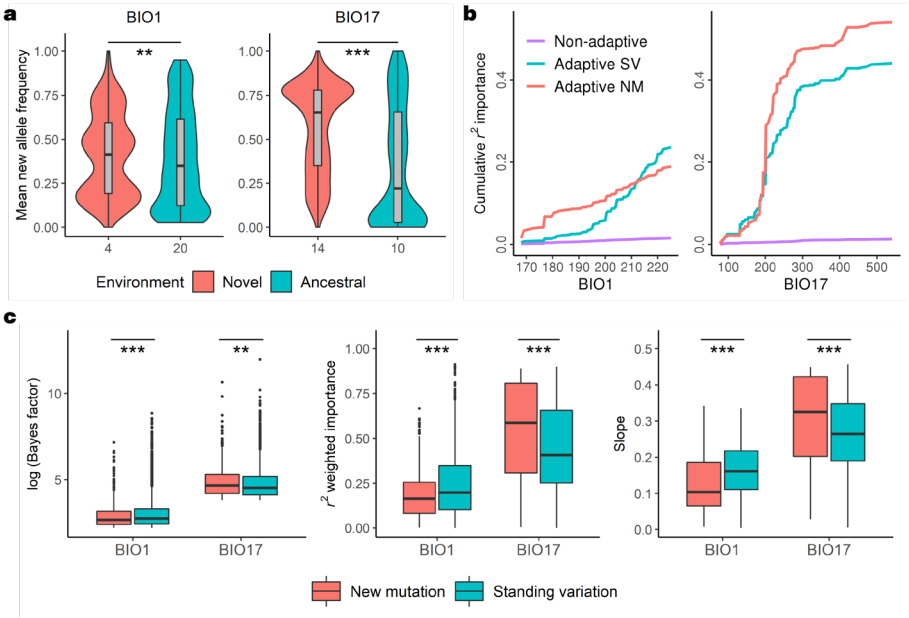


Fig. 2 | The environment-dependent enrichment of new alleles and the distinct effect sizes of standing variation (SV) and new mutations (NM) in contrasting climatic factors. BIO1 reflects annual mean temperature, and BIO17 indicates precipitation of driest quarter. **a**, Mean frequency of new alleles 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).

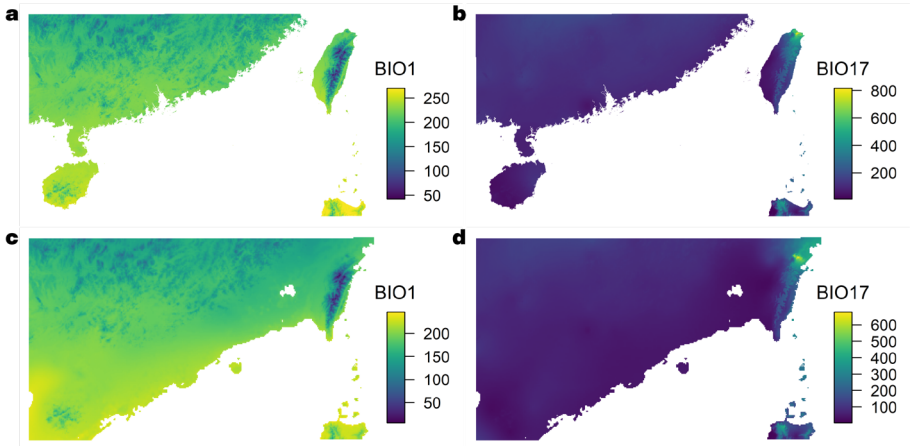


Fig. 3 | Annual mean temperature (BIO1) and precipitation of driest quarter (BIO17) for the present and last glacial maximum. a, **b**, Present environments. **c**, **d**, Last-glacial-maximum environments, showing the environments on the extent of land. Maps on all panels have the same range in latitude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

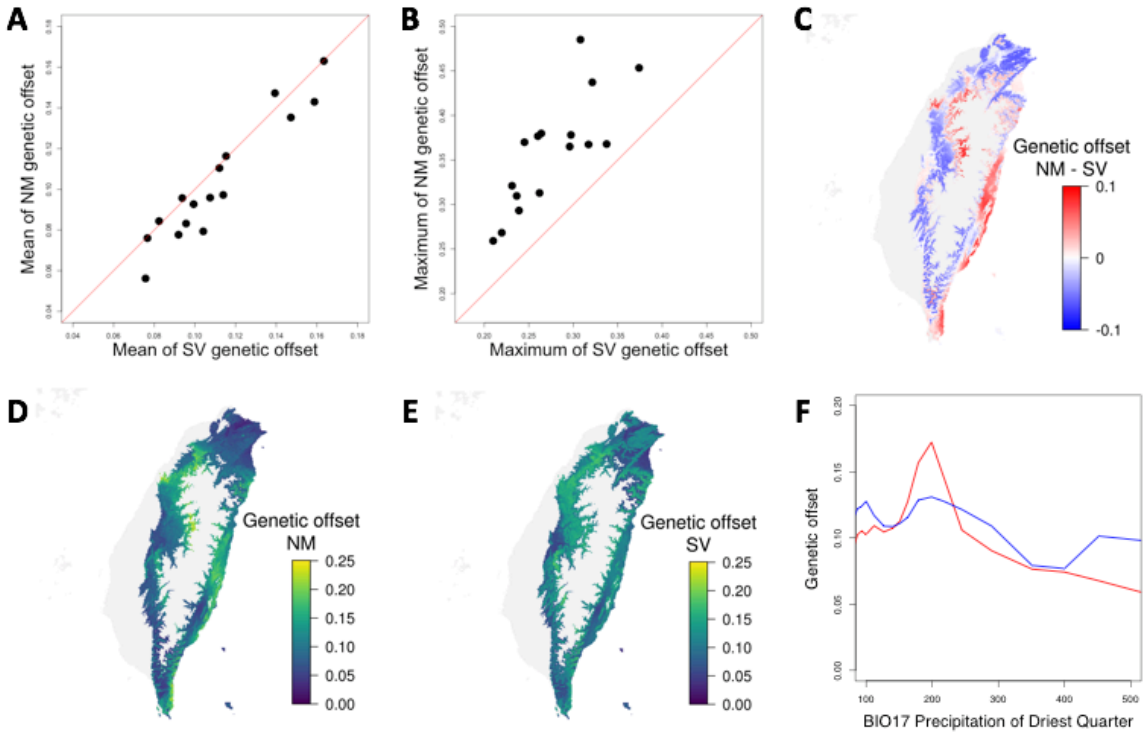


Fig. 4 | The maladaptation of standing variation (SV) and new mutations (NM) towards anthropogenic climate change. (**A**) the mean and (**B**) the maximum genetic offset of currently adaptive SV and NM SNPs across map of Taiwan for each of the 16 future climatic scenarios. In general, SV has higher mean genetic offset and NM has higher maximum genetic offset. (**C**), for the genetic offset map averaged across all 16 future scenarios, this graph shows the offset from NM minus SV. The values range from -0.07 to 0.1, representing NM has more extreme genetic offset values than SV as shown in (**D**) and (**E**). (**D** and **E**), the genetic offset value for NM and SV, averaged across all 16 future scenarios. (**F**), genetic offset values for NM (red) and SV (blue) across current gradients of BIO17 (Precipitation of Driest Quarter) across Taiwan. Grids with current niche suitability < 0.2 are excluded in all calculations and colored in gray in all maps.

Data Availability:

Population pooled sequencing reads are available under NCBI BioProject PRJNA575344.

Acknowledgements:

We thank Thomas Mitchell-Olds and Sergey Nuzhdin 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).

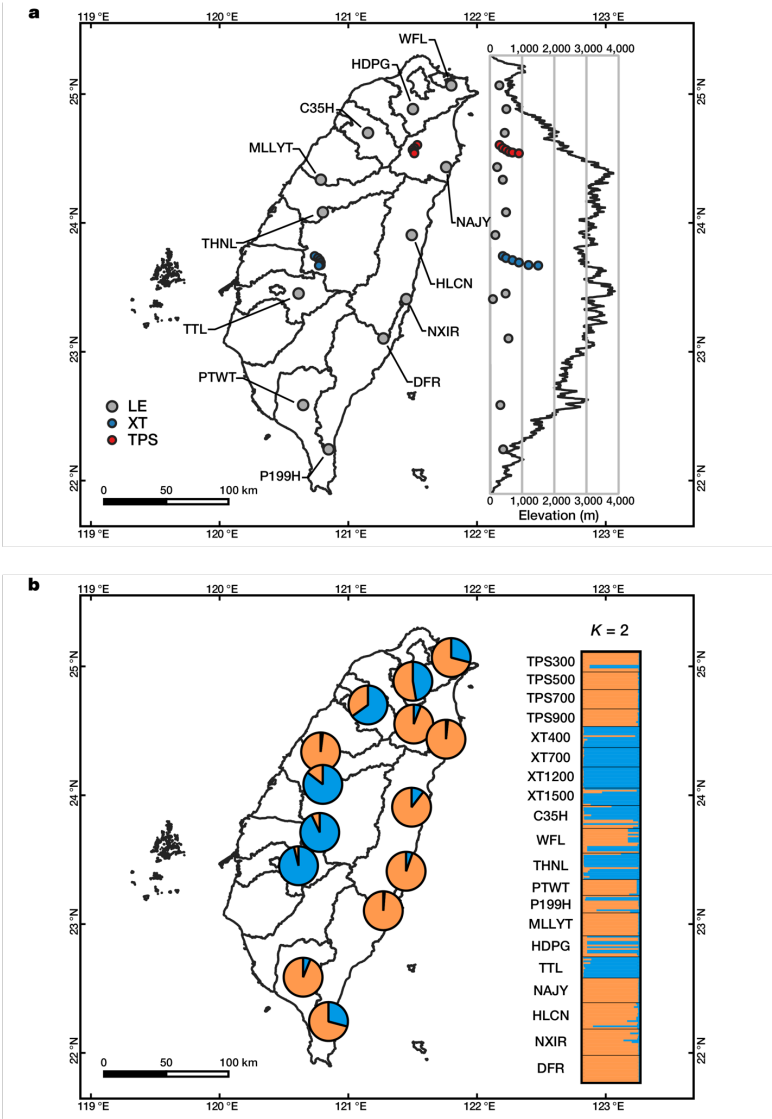
Author Contributions:

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

Competing Interest Declaration:

None

652 **Supplements**



653

654 **Fig. S1 | Sampling site and population structure of Taiwanese *Musa itinerans* under**
655 **STRUCTURE *K* = 2. a**, Solid circles represent collection locations corresponding to their
656 coordinates and elevation: Gray circles indicate populations of low elevation (LE); blue circles
657 indicate populations of Xitou transect (XT); red circles indicate populations of Taipingshan
658 transect (TPS). **b**, Individual ancestry is plotted on the right side, while population ancestry is
659 plotted on map with a pie chart. Map template is provided by *Cheng-Tao Lin.
660 *Cheng-Tao Lin (2018) QGIS template for displaying species distribution by horizontal and
661 vertical view in Taiwan. URL: https://github.com/mutolisp/distrmap_tw.qgis. DOI:
662 10.5281/zenodo.1493690
663

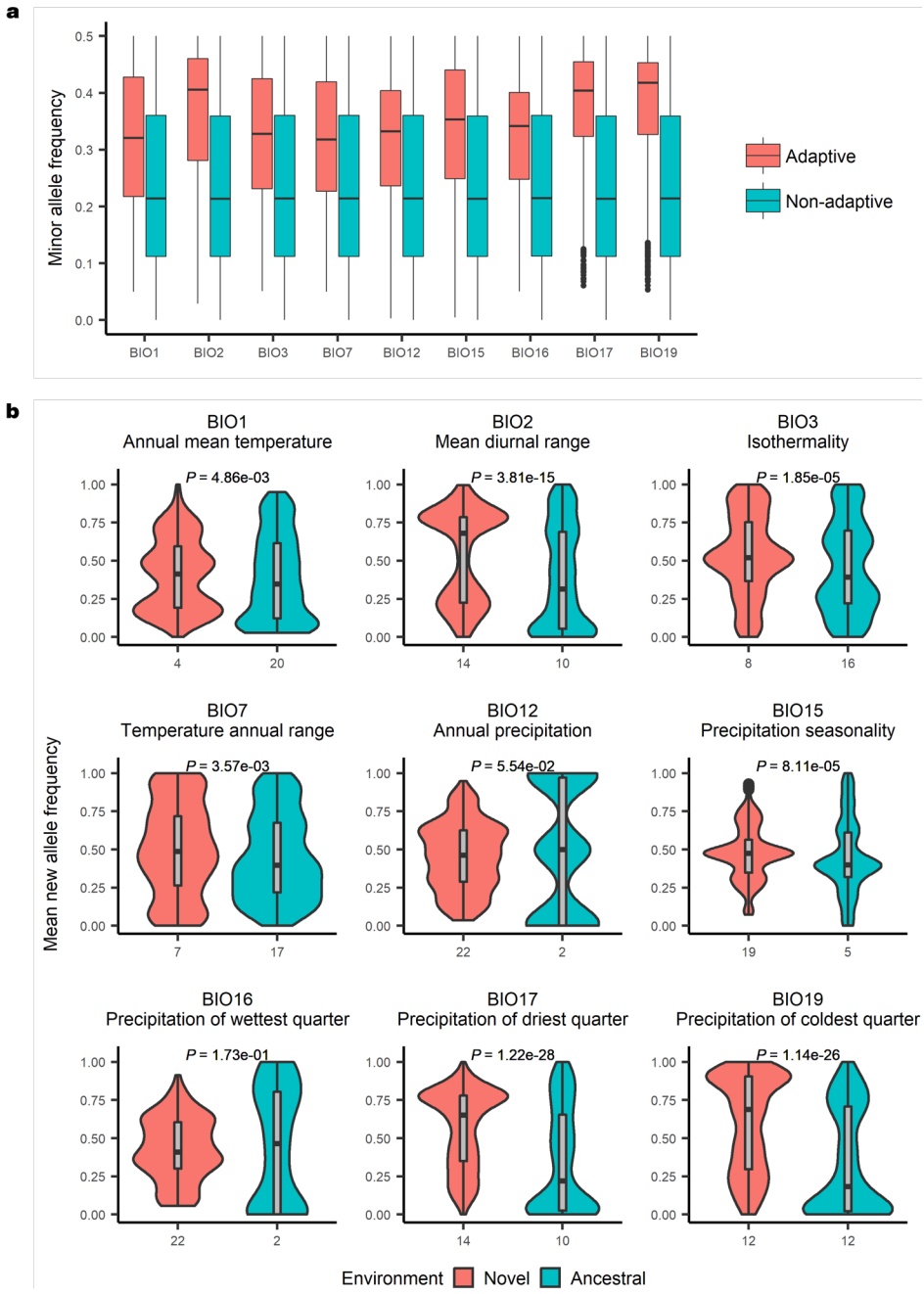


Fig. S2 | Allele frequency distribution. a, Minor allele frequency distribution of adaptive and non-adaptive SNPs. **b**, Frequency of new alleles in Taiwanese populations within the ancestral or novel environmental range. Statistical significance from Student's *t*-test between the novel and ancestral 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.

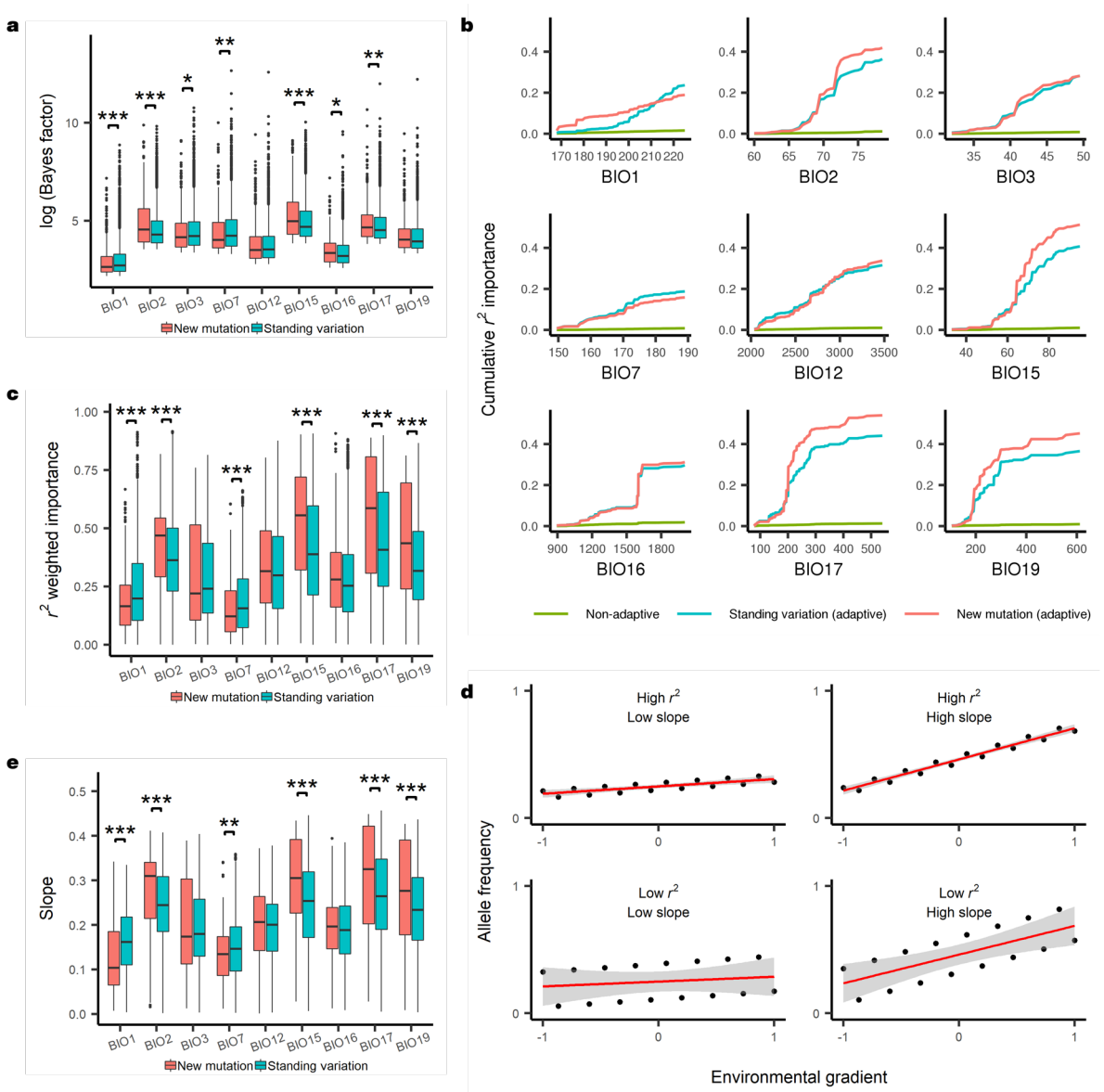
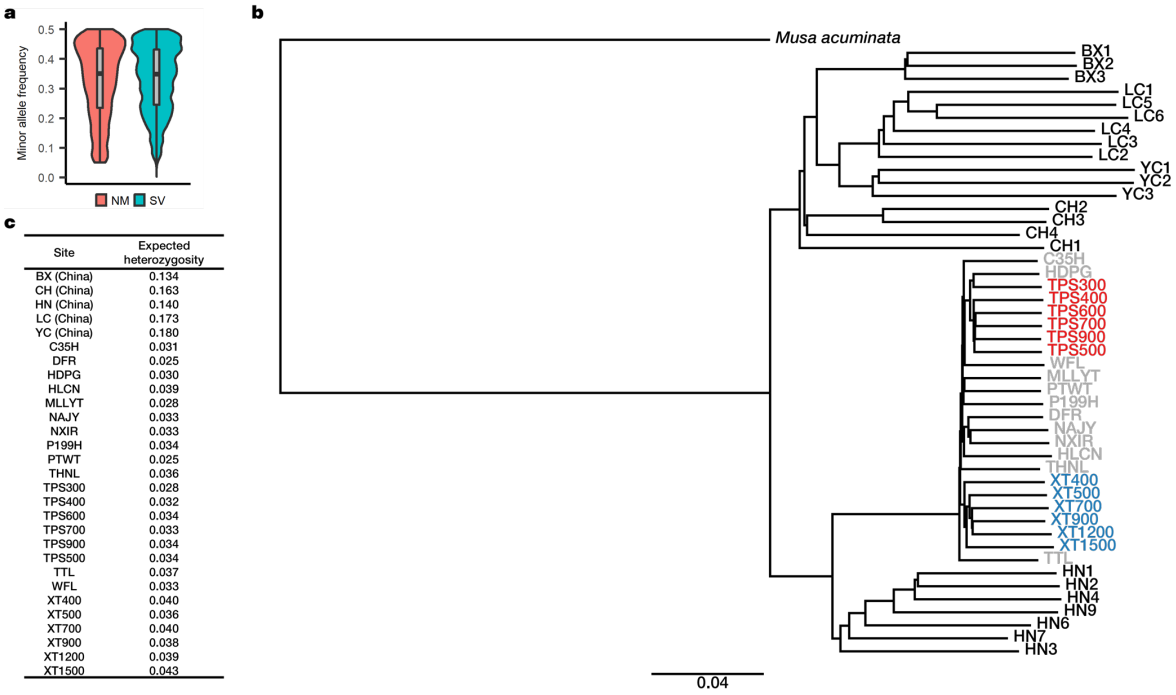


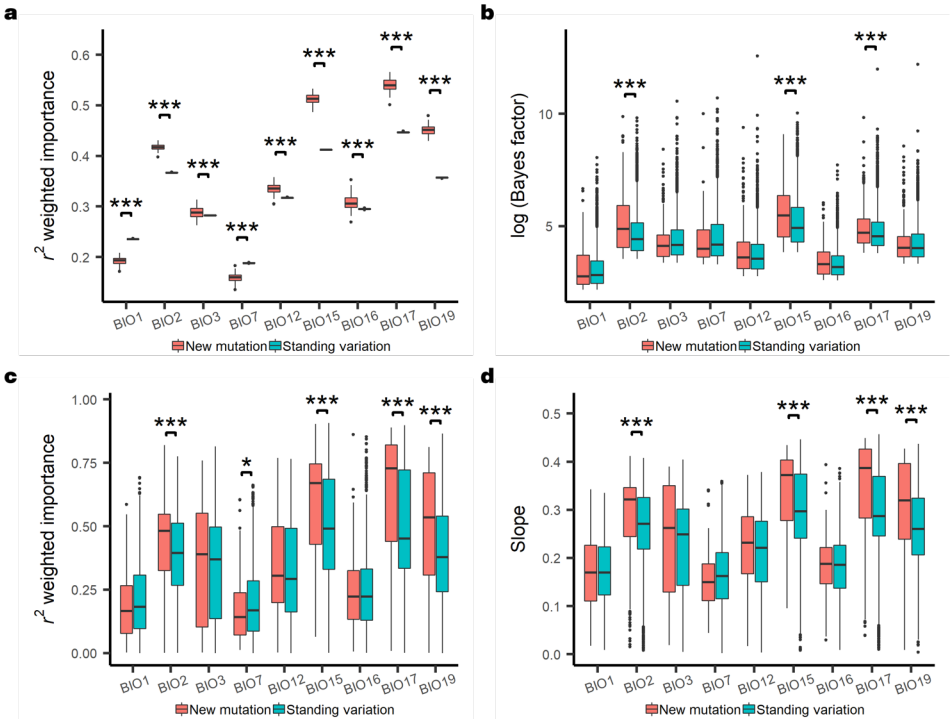
Fig. S3 | Profiling of effect sizes. a, Bayenv Bayes factor distribution for adaptive new mutations and standing variation (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, Wilcoxon rank-sum test). **b**, Cumulative r^2 importance from gradient forest along environmental gradients. **c**, Gradient forest r^2 importance distribution (***$P < 0.001$, Welch's *t*-test). **d**, Example figure showing the relationship between r^2 importance and slope. r^2 indicates the extent that allele frequency fits a linear model, while slope indicates the amount of allele frequency changes along the linear relationship. The graphs indicate one should also investigate regression slopes in addition to the gradient forest r^2. Values on the horizontal axis show the range of standardized environmental variables. **e**, The distribution of regression slopes when one regresses adaptive SNP allele frequency onto environmental gradients (**$P < 0.01$, ***$P < 0.001$, Welch's *t*-test).



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

Fig. S4 | 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.



694

695 **Fig. S5 | Multi-analyses addressing ascertainment of standing variation and new mutation.**
696 **a**, The distribution of gradient forest r^2 importance across 100 re-sampling trials. In each re-
697 sampling trials, a random set of 50% new mutations were assigned as standing variation, and
698 the mean r^2 importance was reported for each trial. All comparisons show strong statistical
699 significance from Student's *t*-test between new mutations and introduced standing variation
700 (***$P < 0.001$). **b-d**, Profiling of effect sizes of adaptive SNPs with top 50% minor allele
701 frequency. Distribution of Bayes factor (***$P < 0.001$, Wilcoxon rank-sum test; **b**), gradient
702 forest r^2 importance (*$P < 0.05$, ***$P < 0.001$, Welch's *t*-test; **c**), and regression slopes of
703 adaptive SNP allele frequency onto environmental gradients (***$P < 0.001$, Welch's *t*-test; **d**)
704 is analyzed between new mutations and standing variation for each bioclimatic variable.
705

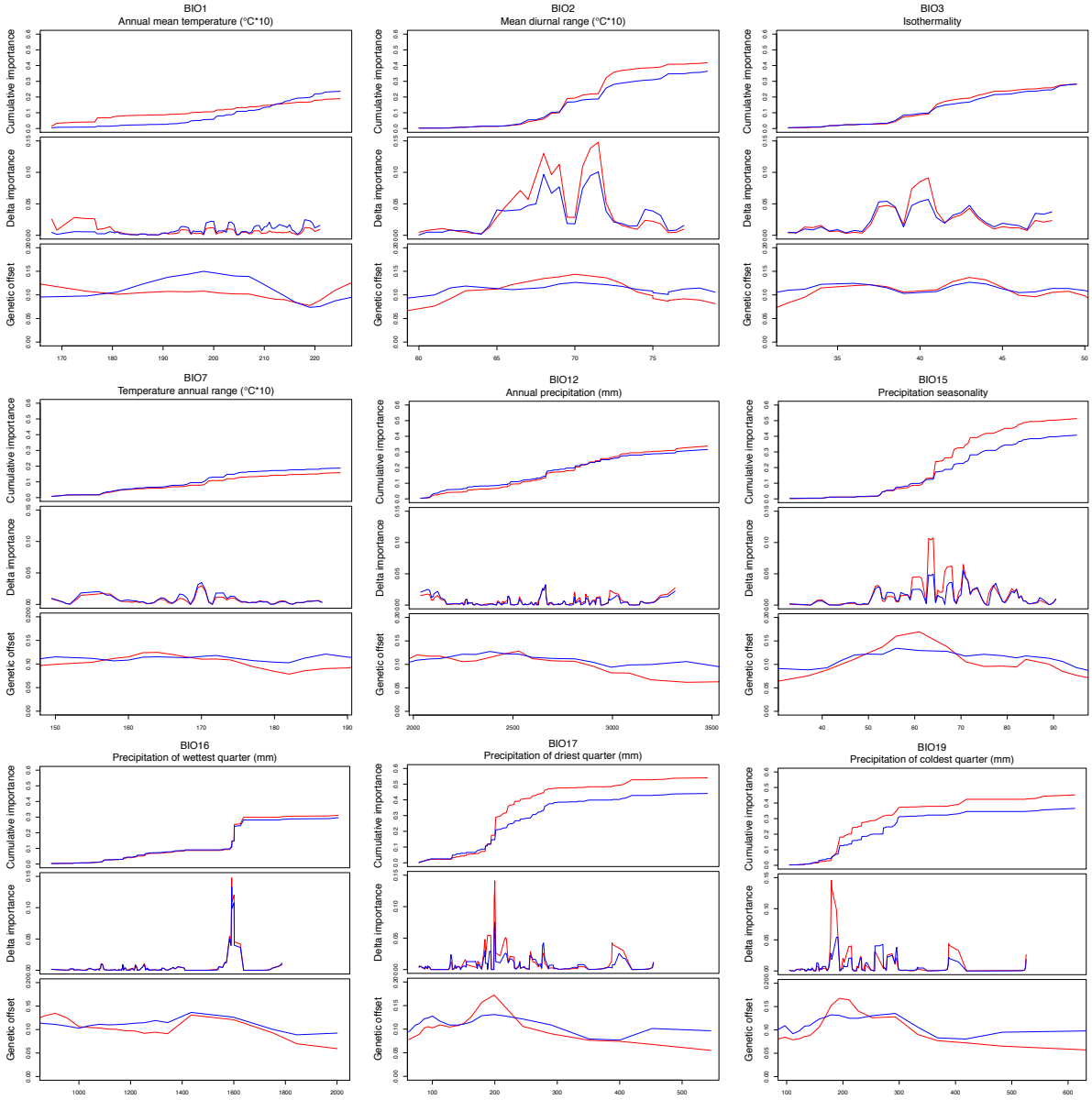


Fig. S6 | Association among gradient forest cumulative importance, delta importance per unit environmental gradient, and genetic offset for each bioclimatic variable. In each figure, red line represents new mutation (NM), and blue line represents standing variation (SV). The "delta importance" is analogous to the instantaneous slope on the cumulative importance graph. Genetic offset was calculated for SV and NM separately, where the genetic offset maps of 16 future climatic scenarios were averaged into one map for SV and one for NM. Geographic grids were separated into 20 bins based on their bioclimatic values, and the mean genetic offset in each bin was plotted. Grids with current niche suitability < 0.2 are excluded in all calculations and colored in gray in all maps.

717 **Table S1 | 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

718

719

720 **Table S2 | SSR primer information**

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

721

722 Primer sequences were searched against the *Musa acuminata* DH-Pahang version 2[32] on
723 Banana Genome Hub (https://banana-genome-hub.southgreen.fr/).
724 *The chromosome and position indicate locations on *Musa acuminata* where primer sequences
725 were found by blast.
726

727 **Table S3 | Number of new-mutation (NM) or standing-variation (SV) SNPs with (adaptive)**
728 **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

729
730 Statistical significance from χ^2 test is shown for each bioclimatic variable. Odds ratio is
731 calculated as (adaptive SV / adaptive NM) / (non-adaptive SV / non-adaptive NM).
732

733 **Table S4 | Number of new-mutation (NM) or standing-variation (SV) SNPs with (adaptive)**
734 **or without (non-adaptive) significant associations with bioclimatic variables (controlled**
735 **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

736
737 The number of four sets of SNPs whose minor allele frequency (MAF) ranges from the first
738 quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each
739 bioclimatic variable is shown. Statistical significance from χ^2 test is shown for each
740 bioclimatic variable. Odds ratio is calculated as (adaptive SV / adaptive NM) / (non-adaptive
741 SV / non-adaptive NM).
742

743 **Table S5 | Number of currently adaptive new-mutation (NM) or standing-variation (SV)**
744 **SNPs that remain (retention) or lose (disruption) significant associations with**
745 **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

746
747 Statistical significance from χ^2 test is shown for each bioclimatic variable. Odds ratio is
748 calculated as (retention SV / retention NM) / (disruption SV / disruption NM).