

1 **The fate of standing variation and new mutation under climate change**

2 Cheng-Yueh Lu¹, Cheng-Ruei Lee^{1,2,3*}

3
4 ¹Institute of Plant Biology, National Taiwan University

5 ²Institute of Ecology and Evolutionary Biology, National Taiwan University

6 ³Genome and Systems Biology Degree Program, National Taiwan University

7 *Correspondence to: chengrueilee@ntu.edu.tw

8 **Abstract:**

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

27

28 **Introduction:**

29 Anthropogenic climatic change poses an imminent threat to most organisms. For large and
30 sessile plant species with long generation time, the speed of migration may not keep up with
31 environmental change, and therefore phenotypic plasticity and genetic variation in the
32 population may allow their survival under novel environments¹⁻⁴. Adaptive genetic variation
33 originates from standing variants before or new mutations after environmental change. Since
34 anthropogenic climate change greatly outpaces natural mutation, the amount of standing
35 genetic variation is therefore critical for the rapid response of a population to changing
36 environments⁵⁻⁷. It remains unclear, however, whether standing variants are sufficient to
37 guarantee future survival, given that they are mostly adaptive to the present range of climatic
38 variability. While it may be difficult to perform manipulative experiments in the field to
39 compare the effects of new mutations (NM) and standing variation (SV), one could investigate
40 NM and SV during climatic change in the past.

41 Adaptation could happen through genetic variants that differ in their origins (NM or SV)
42 or effect sizes (Mendelian genes with major effects or polygenic variants with minor effects).
43 However, how these factors interact and respond to environmental pressures remains relatively
44 uninvestigated. For example, does SV and NM differ in their relative number or effect sizes
45 towards environmental adaptation, and how does this relationship change with different types
46 of environmental factor? For adaptive new mutations that were fixed when facing environment
47 change, Fisher first predicted primarily small allelic effects⁸ while Kimura emphasized
48 intermediate effects⁹. Orr, later considering the entire adaptive walk, concluded the evolution
49 towards a novel adaptive peak should first happen through fixation of large-effect mutations
50 and later by small-effect polymorphisms¹⁰. While this was supported by some studies, the
51 majority of these are microbial experimental evolution in well-controlled environments^{11,12},
52 and few have specifically compared the effects of NM and SV. To test whether this idea holds
53 in nature, empirical investigations on natural populations are needed.

54 Taiwan is well-suited for such studies: Unlike oceanic islands such as Hawaii, Taiwan is
55 a continental island where most species originated from the East Asian continent with recurrent
56 gene flow¹³. The land bridge between Taiwan and China during the glacial maximum allowed
57 exchange of SV, and the isolation during interglacial periods enabled the development of NM.
58 Here we investigate the genomic basis of environmental adaptation of a wild banana species,
59 *Musa itinerans*, whose habitats in Taiwan are considered peripheral from ancestral area
60 reconstructions¹⁴, providing an opportunity to distinguish SV from NM, as well as their
61 response to ancestral versus novel adaptive landscapes. We investigated how past events (SV
62 versus NM) influence present adaptation and whether local adaptation may persist under future
63 anthropogenic climate change.

64

65

66 **Results:**

67 **Environmental adaptation in *Musa itinerans***

68 We first sampled *Musa itinerans* at 24 populations across Taiwan (Fig. 1a; Supplementary
69 Fig. 1a; Supplementary Table 1) and investigated the population structure using 14
70 microsatellites (Supplementary Table 2). Environmental Niche Modeling (with 483 occurrence
71 points from field survey and Google Street View) reported species distribution (Fig. 1b) in line
72 with the previous statement that *Musa itinerans* inhabits sunny valleys, watersheds, and
73 hillsides with gentle slopes¹⁵. Populations differentiated mostly between east and west
74 (Supplementary Fig. 1b). The most unsuitable environments lay within the Central Mountain
75 Range and the southwestern plains (Fig. 1b), respectively corresponding to low annual mean
76 temperature (BIO1) and low precipitation of driest quarter (BIO17), the two most important
77 bioclimatic variables determining species distribution (MaxEnt permutation importance¹⁶: 36.7
78 for BIO1 and 27.3 for BIO17).

79 To test for local adaptation, we examined the pattern of “isolation by adaptation¹⁷”, a
80 process where differential local adaptation restricted effective gene flow and promoted genetic
81 differentiation among populations, by dissecting geographic and environmental effects on
82 genetic differentiation. The strait-line “fly-over” geographical distance, calculated as the
83 straight-line distance between locations, does not explain patterns of genetic differentiation
84 (Mantel’s $r = 0.146$ and $P = 0.062$). However, if we considered that Central Mountain Range
85 lacks corridors for *M. itinerans* to disperse (Fig. 1b), this fly-over geographical distance could
86 be too unrealistic. We therefore used resistance distance, calculated from the route with least
87 resistance among populations on the niche suitability map (Fig. 1b), to represent the “realized”
88 geographical distance (Fig. 1c) and found that genetic differentiation was significantly
89 associated with resistance (Mantel’s $r = 0.226$ and $P = 0.006$). The environmental Mahalanobis
90 distance of bioclimatic variables also showed strong association with genetic differentiation
91 (Mantel’s $r = 0.298$ and $P = 0.005$). Given that the environmental distance could be strongly
92 dependent on geography, we performed Partial Mantel test to control the geographical effect.
93 After controlling for realized geographic distance (resistance distance), genetic differentiation
94 still correlates with the Mahalanobis environmental distance (Mantel’s $r = 0.250$ and $P = 0.012$),
95 suggesting differential local adaptation is associated with genetic variation.

96

97 **Standing variation versus new mutations**

98 To identify genomic regions associated with environmental adaptation, we performed
99 pooled sequencing for each population. SNPs were separated into standing variation (SV: both
100 alleles exist in Taiwan and China) or new mutations (NM, polymorphic only in Taiwan). SV
101 outnumbered NM in both adaptive (identified with Bayenv¹⁸) and non-adaptive SNPs, and after
102 controlling for the overall number of SNPs in SV and NM, SV were further enriched among
103 adaptive polymorphisms (Supplementary Table 3). However, since adaptive SNPs also have

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

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

129 The results above suggest SV might be more important than NM in terms of enriched
130 number of variants with environment association, and adaptive NM, while lower in number,
131 are more associated with novel environments. On the other hand, the number of candidate SNPs
132 does not necessarily reflect the overall importance of SV over NM, since the effect size also
133 needs to be considered. To investigate the effect size of SV and NM in environmental
134 adaptation, we compared their Bayes factors from Bayenv and focused on two bioclimatic
135 variables that were important in determining species distribution, annual mean temperature
136 (BIO1) and precipitation of driest quarter (BIO17). The two variables exhibit opposite patterns
137 (Fig. 2c): While in BIO17 and other related precipitation variables (Supplementary Fig. 3a)
138 NM consistently had higher Bayes factor and therefore stronger effect size than SV, in BIO1
139 and other related temperature variables (Supplementary Fig. 3a) we observed the reverse. The
140 same trend was observed when we estimated the effect size with gradient forest^{19,20} (Fig. 2b, c;
141 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 results⁸⁻¹⁰.

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 model³, providing one of the few 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: (1) The Taiwanese populations were
182 nested within China in the phylogenetic tree (Supplementary Fig. 4b) and contained much less
183 genetic variation (Supplementary Fig. 4c). Due to the stronger genetic drift in Taiwan than
184 China, it is less likely that an originally SV SNP would retain both alleles in Taiwan but lose
185 one in China. (2) In the extreme case, assuming 50% of NM SNPs were mis-assigned from SV,
186 we performed 100 new analyses, each randomly assigning 50% of NM SNPs back to SV. These
187 new analyses yielded similar results, with NM having higher effect sizes than SV in
188 precipitation-related variables (Supplementary Fig. 5a). (3) Since MAF are correlated between
189 Taiwan and China (Spearman's rank correlation $\rho = 0.35$, $P < 0.001$), we performed analyses
190 with top 50% MAF SNPs in Taiwan, thereby reducing the chance of missing minor alleles in
191 China. The results are qualitatively the same (Supplementary Fig. 5b-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 predicted 16 future outcomes and used Bayenv to
197 investigate the future fate of currently adaptive SNPs. 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 any set of adaptive SNPs
201 enriched towards retention or disruption (Supplementary Table 5), suggesting both SV and NM
202 will be affected by climate change.

203 Under the 16 future scenarios we identified two extreme conditions: CCSM4-2070-
204 RCP8.5 as the niche-expansion extreme (Fig. 4a-d) and MIROC-2070-RCP4.5 as the niche-
205 contraction extreme (Fig. 4e-h). Intending to investigate whether SV and NM have different
206 fates under the two contrasting future predictions of this species, we used the genetic offset
207 value from the gradient forest results to estimate genetic mismatch for SV and NM separately,
208 which is associated with the magnitude of allele frequency turnover perturbed by future
209 climatic conditions²⁰ (Fig. 4b, f). The "genetic offset difference" was calculated by subtracting
210 genetic offset caused by SV from that of NM (Fig. 4c, g). We found no general consensus of
211 whether or where in Taiwan NM or SV SNPs are more affected by climate change (Fig. 4c, g),
212 suggesting NM and SV might both be affected by future anthropogenic climate change despite
213 their distinct genetic architecture in adaptation to current environments.

214 Traditional species distribution models build the species-wide niche model from all
215 occurrence records of a species, thereby assuming all populations in a species reacting equally
216 to the environment and overlooking local adaptation due to within-species polymorphism⁶.

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

227

228 **Discussion:**

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

235 For a population facing environmental change and therefore a novel adaptive landscape,
236 previous population genetics models have documented the effect size distribution of new
237 mutations fixed by natural selection. Considering the entire adaptive walk of a population
238 facing novel environments, Orr’s model predicted the distribution of effect sizes where early
239 substitutions have larger effect than later ones¹⁰. Consistent with Orr’s prediction, we show
240 that NM have stronger effect size than SV in precipitation-related variables, where Taiwan
241 exceeds the ancestral climatic range in China (and therefore the migrating population was far
242 away from the optimum in the new adaptive landscape). This pattern is reversed for
243 temperature-related variables, where Taiwan has similar environmental range as China. Here
244 we provide another perspective to recent research showing that SV contributes to adaptation^{7,25}.
245 We show that SV indeed dominate over NM in number⁷. However, the effect size of SV and
246 NM hinges on environmental conditions: natural selection may prefer new mutations
247 contributing to the adaptation to the new rather than the old environment, and the effect sizes
248 of NM tend to be higher than SV under such conditions. Our results imply that standing genetic
249 variation in a population may not be sufficient for the adaptation to anthropogenic climate
250 change.

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

255 the need to consider local adaptation. Incorporating species distribution modeling (predicting
256 the future range of the whole species) and genetic offset (predicting the genetic mismatch of
257 each local population to future environments), we identified the potential risk of western
258 lowland populations despite the seemingly distinct species distribution modeling results in
259 different future climate scenarios. In other words, despite in some future scenarios the range of
260 environments suitable for the whole species might increase, giving the impression that the
261 species benefits from anthropogenic climate change, such change might be detrimental to each
262 local population uniquely adapted to a much narrower environmental range than the whole
263 species. Finally, we separately calculated the genetic offset originated from the mismatch of
264 SV or NM to future environments and showed that both will be affected by anthropogenic
265 climate change, regardless of their distinct genetic architecture towards environmental
266 adaptation. Since we distinguished SNPs into SV or NM based on past environmental change,
267 all of these genetic variants will be standing variation when facing future climate change. Novel
268 mutations facilitating the adaptation to new environments may happen in the future, although
269 they will be strongly outpaced by anthropogenic climate change.
270

271 **Methods:**

272 **Sample collection and DNA extraction.**

273 Field work was conducted during 2017 (August - December) and 2018 (January - May).
274 We sampled *Musa itinerans* at 24 sites across Taiwan (Supplementary Fig. 1a; Supplementary
275 Table 1). Fresh leaves were harvested from nine to fifteen individuals at each site. Total
276 genomic DNA was extracted using the standard CTAB extraction method²⁶. Since other
277 commercial *Musa* species were also grown in Taiwan, we developed an indel marker for
278 species delimitation. From previous studies²⁷⁻³⁰, we identified a 6-bp insertion specific for the
279 Taiwanese *M. itinerans* in the *atpB-rbcL* region of chloroplast. We designed a primer pair (5'-
280 GAAGGGTAGGATTGATTCTCA-3'; 5'-CGACTTGGCATGGCACTATT-3') and used
281 amplicon size to confirm all collected samples are Taiwanese *M. itinerans*.
282

283 **Simple sequence repeat genotyping and analysis.**

284 SSR primer sequences used in this study were originally developed for the genus *Musa*^{31,32},
285 which were then applied on *Musa itinerans*. Previously documented primer sequences were
286 first searched against the *Musa acuminata* DH-Pahang genome version 2³³ on Banana Genome
287 Hub (<https://banana-genome-hub.southgreen.fr/>) to check specificity as having only one
288 amplicon, resulting in 26 primer pairs. These primers were then experimented to check
289 specificity on *M. itinerans*, resulting in 14 pairs (Supplementary Table 2). We modified each
290 pair of primers by capping the 5' end of forward primers with M13 sequences
291 (CACGACGTTGTAAAACGAC) and inflorescent molecules³⁴. SSR amplicons were run
292 through capillary electrophoresis and the length of each allele was recorded.

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

301 To investigate the association among genetic, geographical, and environmental distance,
302 we generated these distance matrices. Genetic distance was calculated by GenAIEx 6.503^{37,38}
303 from 14 SSR markers; straight geographical distance (the fly-over distance) was generated by
304 ArcGIS 10.5 (<http://desktop.arcgis.com/en/>); environmental distance was measured as
305 Mahalanobis distance to address the correlation among nine bioclimatic variables (below). In
306 addition to the fly-over geographical distance which assumes organism dispersal ignores
307 landscapes, we further calculated as resistance distance the cumulative cost along the least cost
308 path (below). Matrix association was examined under Mantel and Partial Mantel tests.
309 Statistical significance was examined with 1,000 permutations. We performed Mantel tests on
310 (1) genetic distance vs. fly-over distance, (2) genetic distance vs. resistance distance, (3) genetic
311 distance vs. Mahalanobis environmental distance, and Partial Mantel tests on (4) genetic
312 distance vs. Mahalanobis environmental distance while controlling for resistance distance.
313

314 Species distribution modeling.

315 Current and future species distribution models were built for *Musa itinerans* using
316 presence-only data (483 occurrence points) obtained from field survey and Google Street view.
317 Occurrence points were then reduced to 204 cells by the removal of co-occurring presence data
318 within the same 1 × 1 km grid. MaxEnt 3.4.1¹⁶, implemented with the maximum entropy
319 modeling approach, reports an overall niche suitability and the importance of predictors by
320 analyzing the presence-only data as well as background (pseudo-absence) data distribution³⁹⁻
321 ⁴¹. We downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) spatial layers
322 of 19 present-day bioclimatic variables based on high-resolution monthly temperature and
323 rainfall data⁴². Layers were selected at spatial resolution of 30 arc-second and with a mask that
324 ranges 119.25–122.47°E and 21.76–25.49°N covering Taiwan. Variables showing high
325 dependence (Pearson's correlation coefficient > 0.9 calculated from ENMTools⁴³) from each
326 other were removed, resulting in nine final variables: BIO1 – mean annual temperature, BIO2
327 – mean diurnal range, BIO3 – isothermality, BIO7 – temperature annual range, BIO12 –
328 annual precipitation, BIO15 – precipitation seasonality, BIO16 – precipitation of wettest
329 quarter, BIO17 – precipitation of driest quarter, and BIO19 – precipitation of coldest quarter
330 (Supplementary Table 1).

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

338 To predict the species distribution under different scenarios of future climate change, we
339 projected the present-day model onto eight future climatic conditions combining two periods
340 (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0,
341 and RCP 8.5). Future climatic layers were obtained from the WorldClim database at spatial
342 resolution of 30 arc-second and were developed based on two general circulation models: the
343 Community Climate System Model⁴⁴, CCSM, and the Model for Interdisciplinary Research on
344 Climate⁴⁵, MIROC. Species distribution models for the future were carried out using the same
345 settings described above.

346 To estimate the least cost path between populations, we first generated the resistance
347 surface by taking the reciprocal of suitability. Resistance and suitability is simply a monotonic
348 transformation in which locations with higher suitability exhibit lower resistance. Pairwise
349 least cost path was then measured among 20 populations from the resistance surface, performed
350 by SDM Toolbox v2.3⁴⁶. While least cost path is the single line with least overall cost, we also
351 constructed the least cost corridor between populations, allowing 1%, 2%, or 5% higher cost
352 than the least cost value. In essence, the least cost corridors represent the realized dispersal
353 routes of organisms along suitable habitats.

354

355 Sequencing library construction and SNP identification.

356 We conducted whole genome pooled-sequencing⁴⁷ for each population (Supplementary
357 Table 1), resulting in 24 pooled-sequencing libraries. Equal amount of DNA from ten
358 individuals at each population were pooled, except for the PTWT population where only nine
359 individuals were available. A library with 300-400 bp insert size for each pool was prepared
360 using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs). Libraries were then
361 sequenced with 150 bp paired-end on the HiSeq X Ten platform.

362 Illumina reads were then trimmed with SolexaQA⁴⁸, followed by the removal of adaptor
363 sequences with cutadapt⁴⁹, subsequently mapped to the *Musa itinerans* reference genome
364 assembly ASM164941v1⁵⁰ with BWA 0.7.15⁵¹. Picard Tools
365 (<http://broadinstitute.github.io/picard>) was used to mark duplicated read pairs, and the
366 genotypes were called following GATK 3.7 best practice⁵².

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

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

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

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

386

387 **Environmentally-associated SNP identification.**

388 We used Bayenv 2.0¹⁸ to search for SNPs highly associated with environmental variables.
389 Bayenv estimates the relationship between SNPs and environments while controlling the
390 whole-genome population structure from a subset of loose linkage-disequilibrium SNPs. Loose
391 linkage-disequilibrium SNPs were formed by sampling (1) one SNP from scaffolds more than
392 10 kb and less than 100 kb, (2) two SNPs from scaffolds more than 100 kb and less than 500
393 kb, (3) three SNPs from scaffolds more than 500 kb and less than 1000 kb, and (4) four SNPs
394 from scaffolds more than 1000 kb. We then, for each bioclimatic variable, defined as the
395 adaptive SNPs ones exhibiting top 1% Bayes factor and top 5% rho value (a nonparametric
396 correlation coefficient capable to reduce outlier effects).

397 We further investigated the fate of currently adaptive SNPs under anthropogenic climate
398 change, performing the same Bayenv analyses of currently adaptive SNPs using future climatic
399 conditions. We included two time periods (2050 and 2070) and four Representative
400 Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) from two general
401 circulation models, CCSM⁴⁴ and MIROC⁴⁵, resulting in 16 future climatic conditions. If a
402 currently adaptive SNP remains strongly associated with environments, it should exhibit Bayes
403 factor above the current threshold. We then defined as “retention” a currently adaptive SNP
404 constantly exhibiting Bayes factor above the current adaptive threshold in all future scenarios,
405 and defined as “disruption” a currently adaptive SNP exhibiting Bayes factor above the current
406 adaptive threshold in none of the future scenarios.

407

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

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

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

438

439 **Regression slope.**

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

449 **References:**

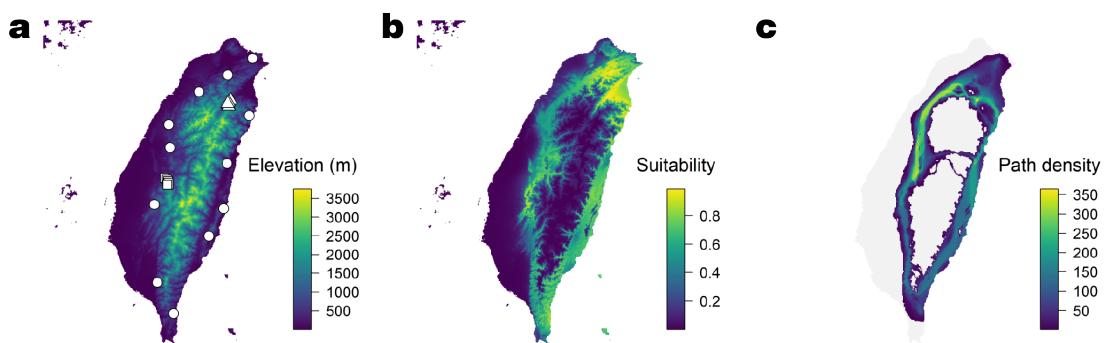
- 450 1 Bradshaw, A. D. Evolutionary significance of phenotypic plasticity in plants. in
451 *Advances in Genetics* **13** (eds Caspary, E. W. & Thoday, J. M.) 115-155 (Academic
452 Press, 1965).
- 453 2 Ghalambor, C. K., McKay, J. K., Carroll, S. P. & Reznick, D. N. Adaptive versus non-
454 adaptive phenotypic plasticity and the potential for contemporary adaptation in new
455 environments. *Funct. Ecol.* **21**, 394-407, doi:10.1111/j.1365-2435.2007.01283.x
456 (2007).
- 457 3 Jump, A. S. & Peñuelas, J. Running to stand still: adaptation and the response of
458 plants to rapid climate change. *Ecol. Lett.* **8**, 1010-1020, doi:10.1111/j.1461-
459 0248.2005.00796.x (2005).
- 460 4 Franks, S. J., Weber, J. J. & Aitken, S. N. Evolutionary and plastic responses to
461 climate change in terrestrial plant populations. *Evol. Appl.* **7**, 123-139,
462 doi:10.1111/eva.12112 (2014).
- 463 5 Barrett, R. D. H. & Schlüter, D. Adaptation from standing genetic variation. *Trends
464 Ecol. Evol.* **23**, 38-44, doi:10.1016/j.tree.2007.09.008 (2008).
- 465 6 Razgour, O. *et al.* Considering adaptive genetic variation in climate change
466 vulnerability assessment reduces species range loss projections. *Proc. Natl. Acad. Sci.
467 USA* **116**, 10418-10423, doi:10.1073/pnas.1820663116 (2019).
- 468 7 Lai, Y. T. *et al.* Standing genetic variation as the predominant source for adaptation of
469 a songbird. *Proc. Natl. Acad. Sci. USA* **116**, 2152-2157, doi:10.1073/pnas.1813597116
470 (2019).
- 471 8 Fisher, R. A. *The Genetical Theory of Natural Selection*. (Clarendon Press, Oxford,
472 England, 1930).
- 473 9 Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University
474 Press, Cambridge, England, 1983).
- 475 10 Orr, H. A. The population genetics of adaptation: The distribution of factors fixed
476 during adaptive evolution. *Evolution* **52**, 935-949, doi:10.2307/2411226 (1998).
- 477 11 Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with
478 *Escherichia coli*. *Nature* **461**, 1243-1247, doi:10.1038/nature08480 (2009).
- 479 12 Gifford, D. R., Schoustra, S. E. & Kassen, R. The length of adaptive walks is
480 insensitive to starting fitness in *Aspergillus nidulans*. *Evolution* **65**, 3070-3078,
481 doi:10.1111/j.1558-5646.2011.01380.x (2011).
- 482 13 Bramwell, D. & Caujapé-Castells, J. *The Biology of Island Floras*. (Cambridge
483 University Press, Cambridge, 2011).
- 484 14 Janssens, S. B. *et al.* Evolutionary dynamics and biogeography of Musaceae reveal a
485 correlation between the diversification of the banana family and the geological and
486 climatic history of Southeast Asia. *New Phytol.* **210**, 1453-1465,

- 487 doi:10.1111/nph.13856 (2016).
- 488 15 Chiu, H. L. The collection, evaluation and analysis of genetic diversity of *Musa*
489 *formosana* (Warb.) Hayata native in Taiwan. (National Taiwan University, 2004).
- 490 16 Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of
491 species geographic distributions. *Ecol. Model.* **190**, 231-259,
492 doi:10.1016/j.ecolmodel.2005.03.026 (2006).
- 493 17 Nosil, P., Egan, S. P. & Funk, D. J. Heterogeneous genomic differentiation between
494 walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent
495 selection. *Evolution* **62**, 316-336, doi:10.1111/j.1558-5646.2007.00299.x (2008).
- 496 18 Gunther, T. & Coop, G. Robust identification of local adaptation from allele
497 frequencies. *Genetics* **195**, 205-220, doi:10.1534/genetics.113.152462 (2013).
- 498 19 Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance
499 gradients on physical predictors. *Ecology* **93**, 156-168, doi:10.1890/11-0252.1 (2012).
- 500 20 Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level
501 modelling of biodiversity: mapping the genomic landscape of current and future
502 environmental adaptation. *Ecol. Lett.* **18**, 1-16, doi:10.1111/ele.12376 (2015).
- 503 21 Barton, N. H. Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **355**,
504 1553-1562, doi:10.1098/rstb.2000.0716 (2000).
- 505 22 Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. On the origin and
506 spread of an adaptive allele in deer mice. *Science* **325**, 1095-1098,
507 doi:10.1126/science.1175826 (2009).
- 508 23 Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks.
509 *Nature* **484**, 55-61, doi:10.1038/nature10944 (2012).
- 510 24 Reid, N. M. *et al.* The genomic landscape of rapid repeated evolutionary adaptation to
511 toxic pollution in wild fish. *Science* **354**, 1305-1308, doi:10.1126/science.aah4993
512 (2016).
- 513 25 Ahrens, C. W., Byrne, M. & Rymer, P. D. Standing genomic variation within coding
514 and regulatory regions contributes to the adaptive capacity to climate in a foundation
515 tree species. *Mol. Ecol.* **28**, 2502-2516, doi:10.1111/mec.15092 (2019).
- 516 26 Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of
517 fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-15 (1987).
- 518 27 White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of
519 fungal ribosomal RNA genes for phylogenetics. in *PCR Protocols* (eds Innis, M. A.,
520 Gelfand, D. H., Sninsky, J. J. & White, T. J.) 315-322 (Academic Press, San Diego,
521 California, 1990).
- 522 28 Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for amplification of
523 three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105-1109,
524 doi:10.1007/BF00037152 (1991).

- 525 29 Oxelman, B., Lidén, M. & Berglund, D. Chloroplast *rps16* intron phylogeny of the
526 tribe Sileneae (Caryophyllaceae). *Plant Syst. Evol.* **206**, 393-410,
527 doi:10.1007/BF00987959 (1997).
- 528 30 Hurr, K. A., Lockhart, P. J., Heenan, P. B. & Penny, D. Evidence for the recent
529 dispersal of *Sophora* (Leguminosae) around the Southern Oceans: molecular data. *J.*
530 *Biogeogr.* **26**, 565-577, doi:10.1046/j.1365-2699.1999.00302.x (1999).
- 531 31 Lagoda, P. J. *et al.* Sequence tagged microsatellite site (STMS) markers in the
532 Musaceae. *Mol. Ecol.* **7**, 659-663 (1998).
- 533 32 Crouch, H. K., Crouch, J. H., Jarret, R. L., Cregan, P. B. & Ortiz, R. Segregation at
534 microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Sci.* **38**, 211-217,
535 doi:10.2135/cropsci1998.0011183X003800010035x (1998).
- 536 33 Martin, G. *et al.* Improvement of the banana "*Musa acuminata*" reference sequence
537 using NGS data and semi-automated bioinformatics methods. *BMC Genomics* **17**,
538 243, doi:10.1186/s12864-016-2579-4 (2016).
- 539 34 Schuelke, M. An economic method for the fluorescent labeling of PCR fragments.
540 *Nat. Biotechnol.* **18**, 233-234, doi:10.1038/72708 (2000).
- 541 35 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using
542 multilocus genotype data. *Genetics* **155**, 945-959 (2000).
- 543 36 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using
544 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**,
545 1567-1587 (2003).
- 546 37 Peakall, R. O. D. & Smouse, P. E. Genalex 6: genetic analysis in Excel. Population
547 genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288-295,
548 doi:10.1111/j.1471-8286.2005.01155.x (2006).
- 549 38 Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population
550 genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537-2539,
551 doi:10.1093/bioinformatics/bts460 (2012).
- 552 39 Bean, W. T., Stafford, R. & Brashares, J. S. The effects of small sample size and
553 sample bias on threshold selection and accuracy assessment of species distribution
554 models. *Ecography* **35**, 250-258, doi:10.1111/j.1600-0587.2011.06545.x (2012).
- 555 40 Elith, J. *et al.* Novel methods improve prediction of species' distributions from
556 occurrence data. *Ecography* **29**, 129-151, doi:10.1111/j.2006.0906-7590.04596.x
557 (2006).
- 558 41 Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**,
559 43-57, doi:10.1111/j.1472-4642.2010.00725.x (2011).
- 560 42 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high
561 resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**,
562 1965-1978, doi:10.1002/joc.1276 (2005).

- 563 43 Warren, D. L., Glor, R. E. & Turelli, M. ENMTools: a toolbox for comparative studies
564 of environmental niche models. *Ecography* **33**, 607-611, doi:10.1111/j.1600-
565 0587.2009.06142.x (2010).
- 566 44 Gent, P. R. *et al.* The Community Climate System Model Version 4. *J. Climate* **24**,
567 4973-4991, doi:10.1175/2011jcli4083.1 (2011).
- 568 45 Watanabe, S. *et al.* MIROC-ESM 2010: model description and basic results of
569 CMIP5-20c3m experiments. *Geosci. Model Dev.* **4**, 845-872, doi:10.5194/gmd-4-845-
570 2011 (2011).
- 571 46 Brown, J. L. & Anderson, B. SDMtoolbox: a python-based GIS toolkit for landscape
572 genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol.*
573 **5**, 694-700, doi:10.1111/2041-210x.12200 (2014).
- 574 47 Schlotterer, C., Tobler, R., Kofler, R. & Nolte, V. Sequencing pools of individuals -
575 mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**,
576 749-763, doi:10.1038/nrg3803 (2014).
- 577 48 Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment
578 of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485,
579 doi:10.1186/1471-2105-11-485 (2010).
- 580 49 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
581 reads. *EMBnet.journal* **17**, 10-12, doi:10.14806/ej.17.1.200 (2011).
- 582 50 Wu, W. *et al.* Whole genome sequencing of a banana wild relative *Musa itinerans*
583 provides insights into lineage-specific diversification of the *Musa* genus. *Sci. Rep.* **6**,
584 31586, doi:10.1038/srep31586 (2016).
- 585 51 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
586 transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 587 52 McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for
588 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303,
589 doi:10.1101/gr.107524.110 (2010).
- 590 53 Wu, W., Ng, W. L., Yang, J. X., Li, W. M. & Ge, X. J. High cryptic species diversity is
591 revealed by genome-wide polymorphisms in a wild relative of banana, *Musa*
592 *itinerans*, and implications for its conservation in subtropical China. *BMC Plant Biol.*
593 **18**, 194, doi:10.1186/s12870-018-1410-6 (2018).
- 594

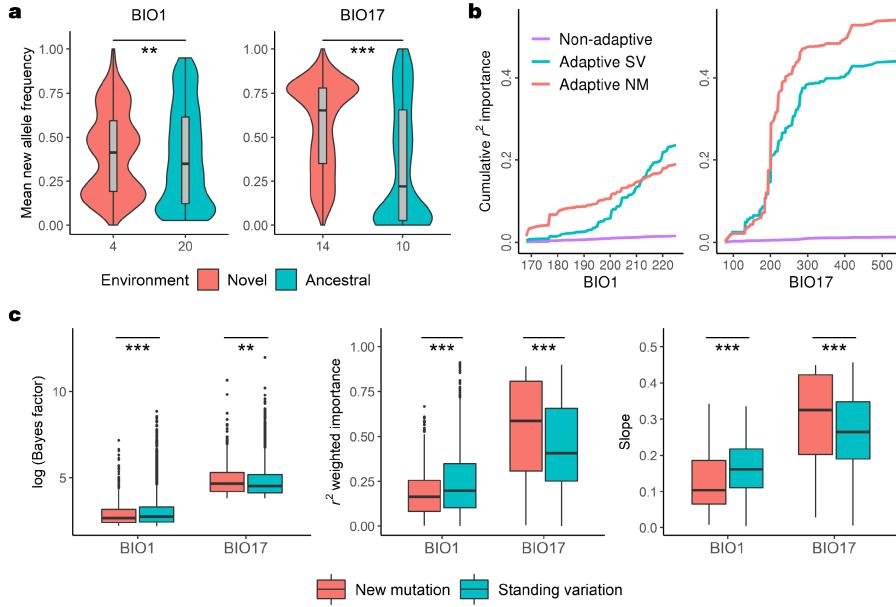
595 **Figures:**



596

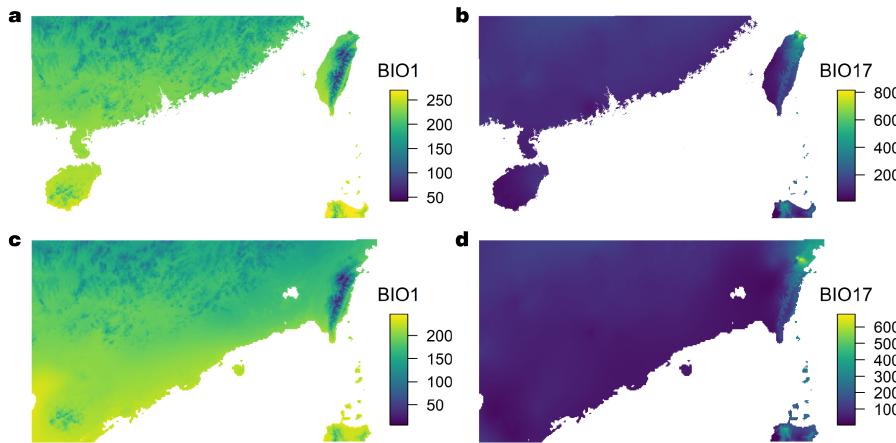
597 **Fig. 1 | Sample distribution and niche modeling of *Musa itinerans*.** **a**, Sampling sites are
598 distributed along the latitudinal and altitudinal gradient. Lowland populations are represented
599 as white circles; transect populations are represented as white triangles and squares. **b**,
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



604
605 **Fig. 2 | The environment-dependent enrichment of new alleles and the distinct effect sizes**
606 **of standing variation (SV) and new mutations (NM) in contrasting climatic factors.** BIO1
607 reflects annual mean temperature, and BIO17 indicates precipitation of driest quarter. **a**, Mean
608 frequency of new alleles among NM SNPs compared between populations that have local
609 environments within or outside of the ancestral climatic range. New alleles are enriched in
610 novel environments (** $P < 0.01$, *** $P < 0.001$, t -test). Values on the horizontal axis denote the
611 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
612 sizes as estimated from Bayes factor, gradient forest r^2 importance, and gradient forest slope
613 all show that SV exhibit higher effect sizes in BIO1 but the reverse in BIO17 (** $P < 0.01$, *** P
614 < 0.001, Wilcoxon rank-sum test for Bayes factor and t -test for r^2 importance and slope).

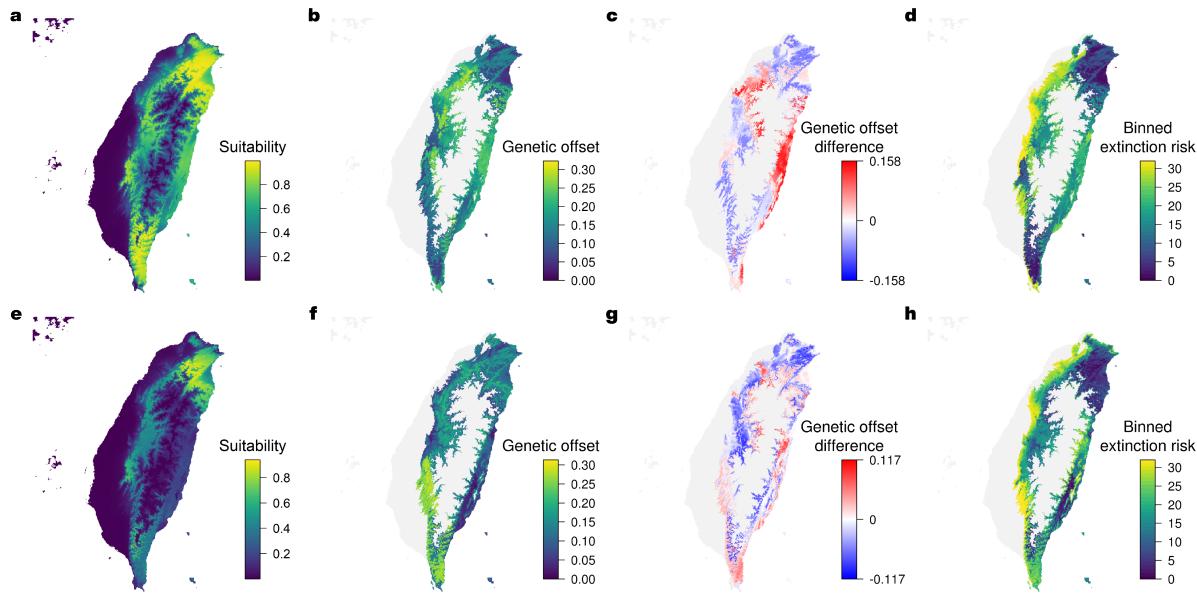
615
616



617

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

622



623

624 **Fig. 4 | The fate of *Musa itinerans* under two extreme future conditions.** a-d, Niche-
625 expansion extreme (CCSM4-2070-RCP8.5) reports the species-average niche suitability from
626 MaxEnt modeling (a), the genetic mismatch of locally adaptive populations to future
627 environments based on the genetic offset from gradient forest (b), genetic offset difference
628 calculated as the genetic offset experienced by new mutations minus that from standing
629 variation (c), and extinction risk estimated as genetic offset divided by niche suitability (shown
630 as 32 bins with equal grid counts; d). e-h, Niche-contraction extreme (MIROC-2070-RCP4.5)
631 reports the suitability (e), genetic offset (f), genetic offset difference (g), and extinction risk (h).
632 Grids with present niche suitability < 0.2 are excluded and colored in gray in b-d and f-h.
633

634 **Data Availability:**

635 Population pooled sequencing reads are available under NCBI BioProject PRJNA575344.

636 **Acknowledgements:**

637 We thank Thomas Mitchell-Olds for valuable comments, Hui-Long Chiu for the knowledge of
638 *Musa itinerans* ecology in Taiwan, Chia-Yu Chen, Zhe-Ting Kuo, and Jo-Wei Hsieh for the
639 assistance of sample preparation, Hao-Chih Kuo for the introduction to species distribution
640 modeling, Cheng-Tao Lin for providing the R template of some maps, the Lee lab member for
641 valuable discussions, and the Computer and Information Networking Center of National
642 Taiwan University for the support of high-performance computing facilities. This work is
643 supported by the Ministry of Science and Technology of Taiwan (106-2628-B-002-001-MY3,
644 107-2636-B-002-004, and 108-2636-B-002-004 to CRL).

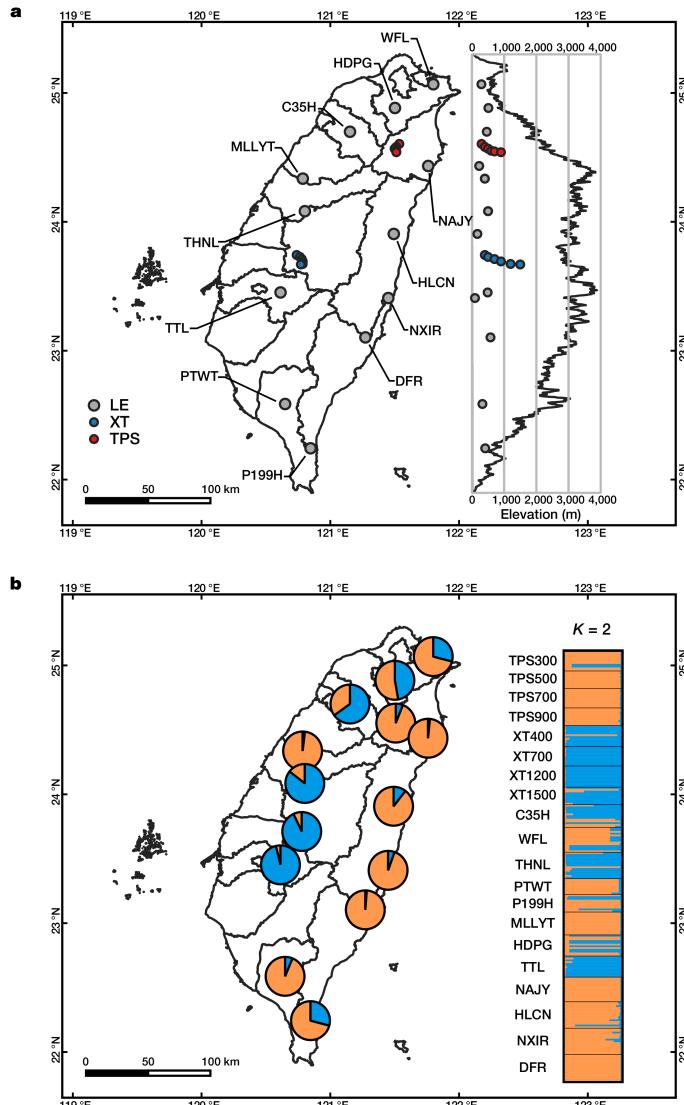
645 **Author Contributions:**

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

648 **Competing Interest Declaration:**

649 None

650

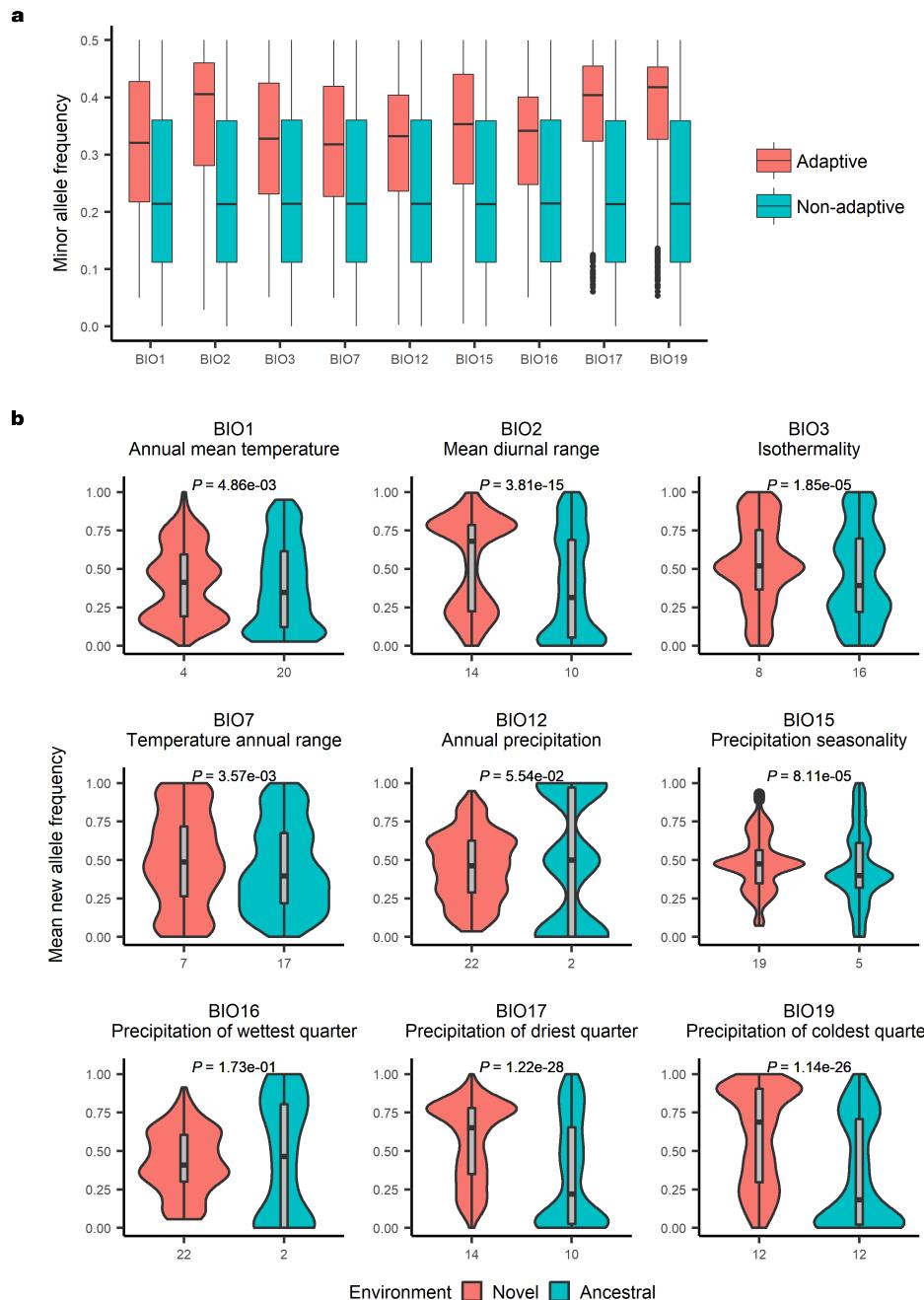


651

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

659 *Cheng-Tao Lin (2018) QGIS template for displaying species distribution by horizontal and
660 vertical view in Taiwan. URL: https://github.com/mutolisp/distrmap_tw.qgis. DOI:
661 10.5281/zenodo.1493690

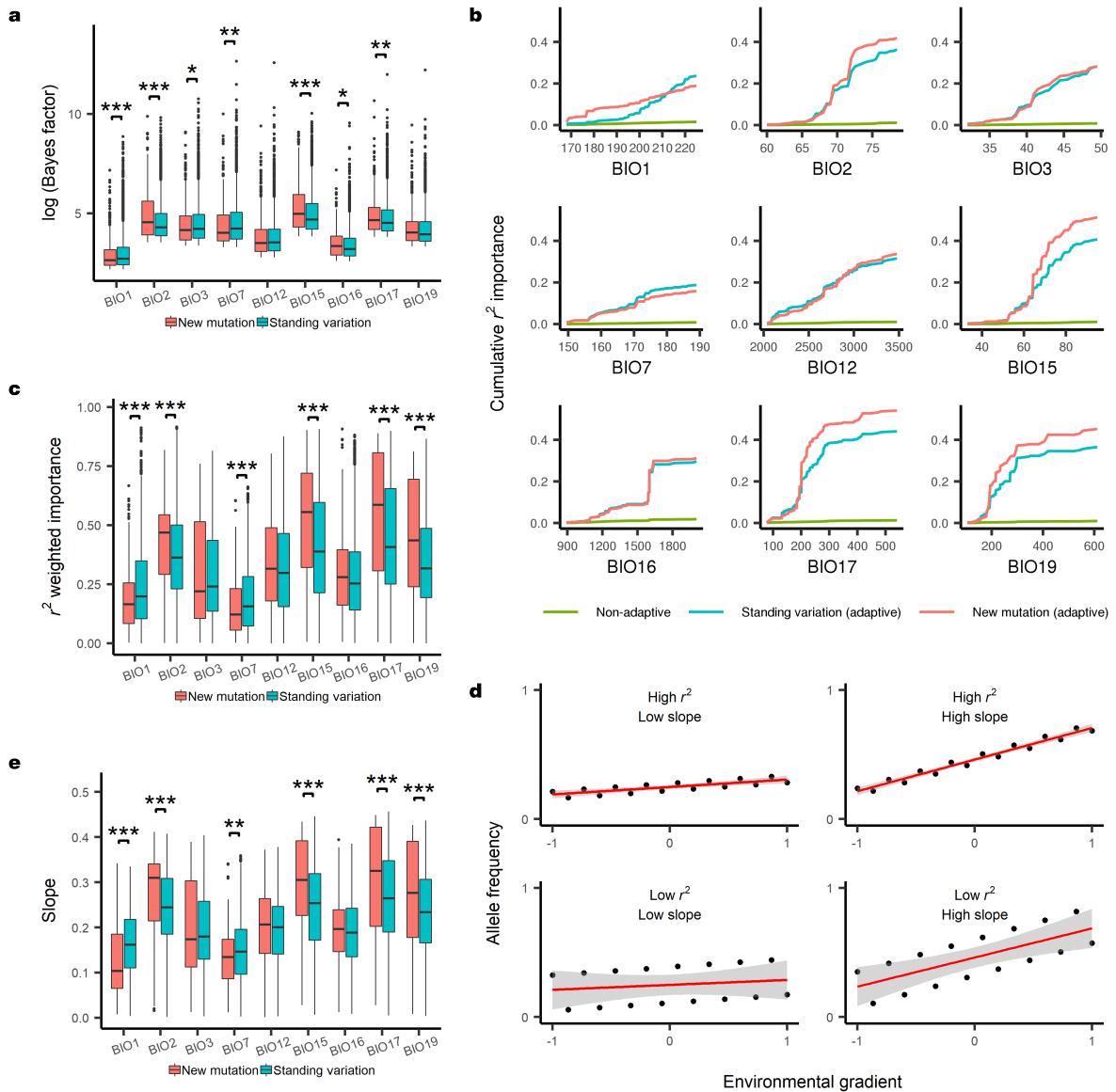
662



663

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

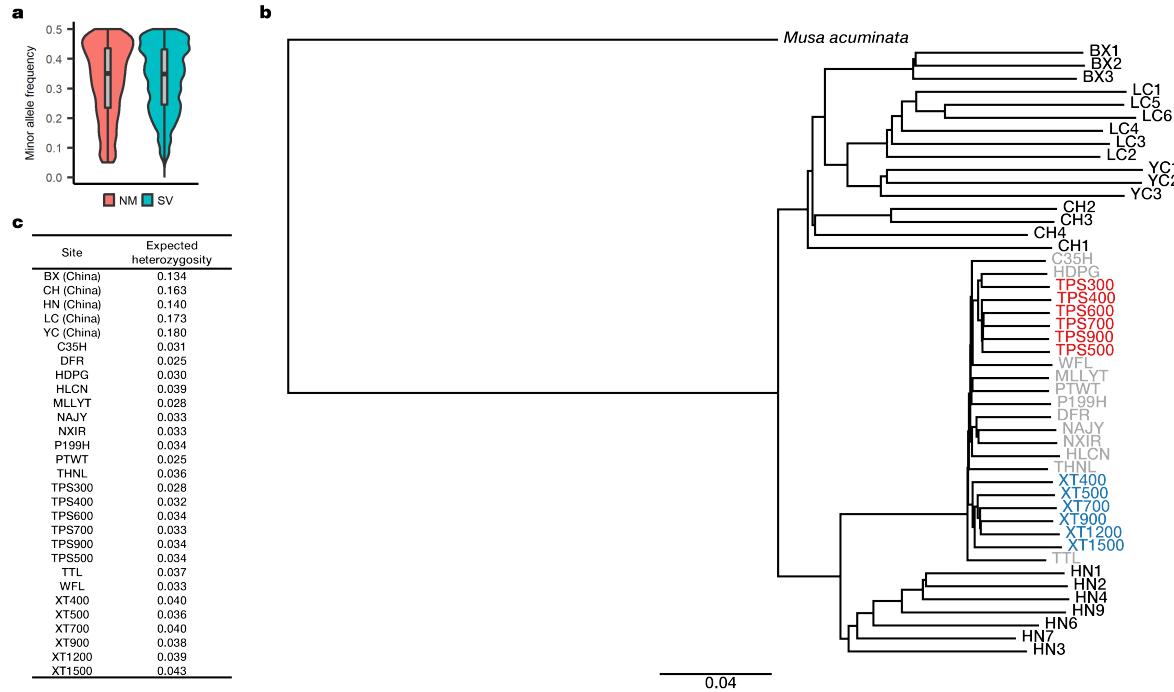
670



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).

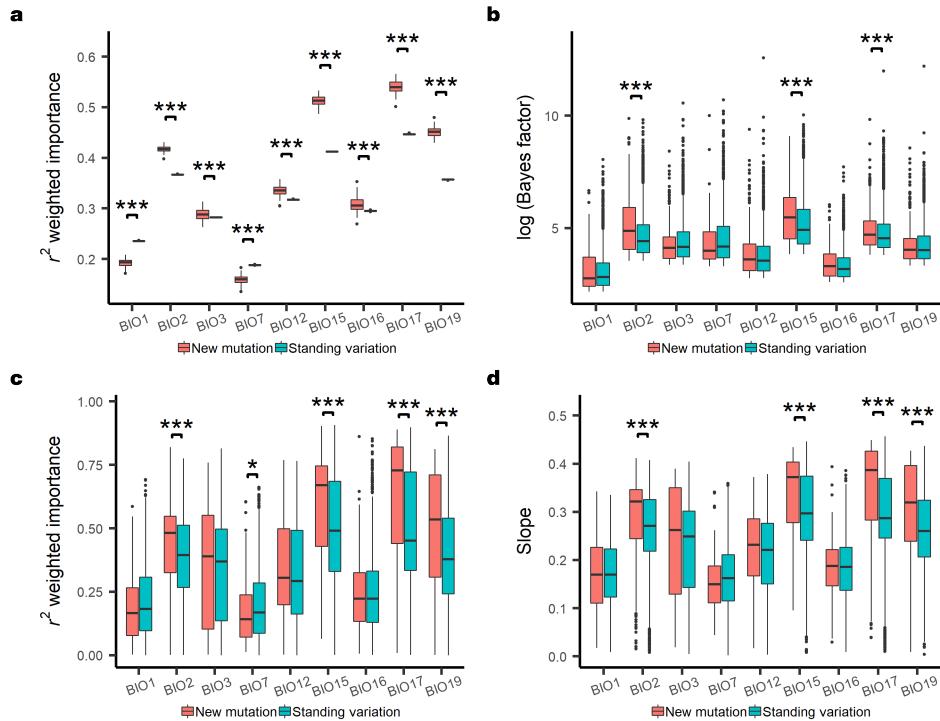
683



684

685 **Supplementary Fig. 4 | Evolutionary analyses on Taiwanese and Chinese *Musa itinerans*.**

686 **a**, Minor allele frequency distribution of adaptive SNPs in Taiwan. **b**, Phylogeny of Taiwanese
687 and Chinese *Musa itinerans*. The gray-colored indicates Taiwanese lowland populations (LE);
688 the blue-colored indicates Taiwanese populations of Xitou transect (XT); the red-colored
689 indicates Taiwanese populations of Taipingshan transect (TPS); the black-colored indicates
690 Chinese accessions and the outgroup *Musa acuminata*. **c**, Table showing the expected
691 heterozygosity among Taiwanese and Chinese populations.
692



693

694 **Supplementary Fig. 5 | Multi-analyses addressing ascertainment of standing variation**
695 **and new mutation.** **a**, The distribution of gradient forest r^2 importance across 100 re-sampling
696 trials. In each re-sampling trials, a random set of 50% new mutations were assigned as standing
697 variation, and the mean r^2 importance was reported for each trial. All comparisons show strong
698 statistical significance from Student's t -test between new mutations and introduced standing
699 variation ($***P < 0.001$). **b-d**, Profiling of effect sizes of adaptive SNPs with top 50% minor
700 allele frequency. Distribution of Bayes factor ($***P < 0.001$, Wilcoxon rank-sum test; **b**),
701 gradient forest r^2 importance ($*P < 0.05$, $***P < 0.001$, Welch's t -test; **c**), and regression slopes
702 of adaptive SNP allele frequency onto environmental gradients ($***P < 0.001$, Welch's t -test;
703 **d**) is analyzed between new mutations and standing variation for each bioclimatic variable.
704

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 CTGCAGCAACCAAATTATTCT | 1 | 4,917,979 - 4,918,003 4,918,057 - 4,918,080 |
| CIR276 | CTCCTCCATAGCCTGACTGC TGACCCACGAGAAAAGAACG | 1 | 13,588,759 - 13,588,778 13,588,847 - 13,588,866 |
| CIR1113 | ACTCTGCCCATCTTCATCC ACTTATCCCCCGCACTCAA | 1 | 28,009,948 - 28,009,967 28,010,173 - 28,010,192 |
| Ma-3-132 | TCCCTTCAACCAAAGCAC AACCGAATGTGTGTTTCA | 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 | AGGCGGGAAATCGGTAGA GGCGGGAGACAGATGGAGT | 2 | 24,383,391 - 24,383,408 24,383,488 - 24,383,506 |
| CIR332a | ATGACCTGTGAAACATCCTT TCCCAACCCCTGCAACCCT | 3 | 8,978,575 - 8,978,595 8,978,831 - 8,978,848 |
| Ma-3-48 | CCCGTCCCATTCTCA TTCGTTGTTCATGGAATCA | 5 | 32,672,883 - 32,672,898 32,673,018 - 32,673,036 |
| CIR631a | ATTAGATCACCGAAGAACTC ATCTTTCTTATCCTTCTAACG | 6 | 33,942,751 - 33,942,770 33,943,017 - 33,943,038 |
| CIR16a | TCATCTCACAATGCTTCTAGTT TGTTGAGTAGATCTTCTTGTT | 8 | 1,234,609 - 1,234,632 1,234,700 - 1,234,722 |
| Ma-3-103 | TCGCCTCTCTTAGCTCTG TGTTGGAGGATCTGAGATTG | 8 | 40,127,782 - 40,127,800 40,127,910 - 40,127,929 |
| CIR348b | ACAGAACGCTAACCTTAATCCTCA CCCTTGGCGTGCCTCTAA | 10 | 27,228,162 - 27,228,186 27,228,325 - 27,228,342 |
| Ma-3-139 | ACTGCTGCTCTCCACCTAAC GTCCCCCAAGAACCATATGATT | 10 | 30,237,591 - 30,237,611 30,237,717 - 30,237,735 |
| CIR550a | ACCGCACCTCCACCTCTG TGCTGCCTCATCGCTACTA | 10 | 31,914,219 - 31,914,237 31,914,459 - 31,914,478 |

709

710 Primer sequences were searched against the *Musa acuminata* DH-Pahang version 2³³ 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

715 **Supplementary Table 3 | Number of new-mutation (NM) or standing-variation (SV) SNPs**
716 **with (adaptive) or without (non-adaptive) significant associations with bioclimatic**
717 **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 |

718 Statistical significance from χ^2 test is shown for each bioclimatic variable. Odds ratio is
719 calculated as (adaptive SV / adaptive NM) / (non-adaptive SV / non-adaptive NM).
720
721

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 χ^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

732 **Supplementary Table 5 | Number of currently adaptive new-mutation (NM) or standing-
733 variation (SV) SNPs that remain (retention) or lose (disruption) significant associations
734 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 |

735
736 Statistical significance from χ^2 test is shown for each bioclimatic variable. Odds ratio is
737 calculated as (retention SV / retention NM) / (disruption SV / disruption NM).