

國立臺灣大學生命科學院植物科學研究所



碩士論文

Institute of Plant Biology

College of Life Science

National Taiwan University

Master Thesis

台灣芭蕉適應環境的遺傳結構之探究

**The genetic architecture of environmental adaptation  
for the past, present, and future of *Musa itinerans***

陸政鉞

Cheng-Yueh Lu

指導教授：李承叡 博士

Advisor: Cheng-Ruei Lee, Ph.D.

中華民國 108 年 8 月

August 2019

國立臺灣大學碩士學位論文  
口試委員會審定書

台灣芭蕉適應環境的遺傳結構之探究

The genetic architecture of environmental adaptation for  
the past, present, and future of *Musa itinerans*

本論文係陸政鉞君（學號 R06B42017）在國立臺灣大學植物科學研究所完成之碩士學位論文，於 2019 年 7 月 15 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

臺大生態學與演化生物學研究所李承叡博士

李承叡 (簽名)

臺大臨床醫學研究所王弘毅博士

王弘毅 (簽名)

臺大生命科學系王俊能博士

王俊能 (簽名)

臺大生命科學系丁照棟博士

丁照棟 (簽名)

## 摘要

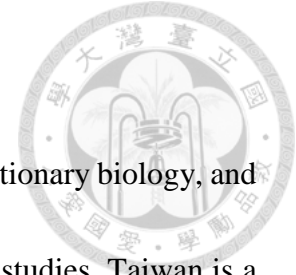


遺傳結構與其對應的效應值向來是演化生物學的研究核心。冰河時期和亞洲大陸間斷相連的台灣恰為絕佳地點進行該類研究，因其不僅具有大陸族群的既有變異，更有因應新環境被保留的新生突變。此研究著眼於野生種香蕉——台灣芭蕉，並利用生態基因體學的方法檢視其過去至未來的適應進程。遺傳組成與環境因子的顯著相關顯示台灣芭蕉存在在地適應現象，且新生突變與既有變異此二者遺傳結構貢獻不一。數量上，既有變異普遍較新生突變為多；而對於雨量相關的氣候因子，效應值則以新生突變為大，且當該因子相對於中國大陸地區為一嶄新的氣候時，此現象尤為明顯。我們亦著眼於台灣芭蕉未來的適應現象，雖未發現氣候變遷傾向保留任一遺傳結構，但透過物種分布模型與效應值的結合，揭露了適存族群潛在的滅絕風險。此研究不僅演示了遺傳結構——既有變異與新生突變——的適應軌跡，且透過不同模型的整合，指出台灣西南部為台灣芭蕉的易危區域。

### 關鍵字：

效應值、在地適應、既有變異、新生突變、滅絕風險

## Abstract

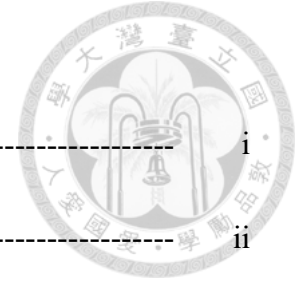


Genetic architecture of adaptation has been the central focus to evolutionary biology, and effect sizes thereof have been investigated from theory to empirical studies. Taiwan is a perfect place to explore such synthesis of the genetic basis and effect size where standing variations (SV) and new mutations (NM) were established through the recurring connection to East Asian continent at glacial periods. Here, we center on a wild banana *Musa itinerans* that distributes along altitudinal and latitudinal gradients in Taiwan, and assess the adaptive course from the past to the future. Significant genetics-environment association indicates local adaptation where the assortment of SV and NM contributes differently. While SV are dominant in number, NM exert larger effect size in precipitation-related climates, especially for those novel to mainland China. Under anthropogenic climate change, both SV and NM have no inclination to retain in the future. Incorporation of effect size into species distribution modeling unveils the indiscernible extinction risk of apparently fitting populations. Our results demonstrate the trajectories of adaptive SV and NM, and identify southwestern Taiwan as the most vulnerable region with the integration of universal and locally differential responses of *M. itinerans*.

### **Keyword:**

effect size, local adaptation, standing variation, new mutation, extinction risk

# Contents



摘要	i
Abstract	ii
Contents	iii
Contents of Figures	v
Contents of Tables	viii
Introduction	1
Materials and Methods	5
Sample Collection and DNA Extraction	5
Simple Sequence Repeat Genotyping	5
Library Construction and SNP Identification	6
Data Analysis	8
Population Structure	8
Species Distribution Modeling	9
Isolation Pattern and Adaptive SNP Identification	11
Adaptive SNP Retention and Disruption under Climate Change	12
Genetic Offset	13
Regression Slope	14

Results -----	16
Population Structure and Evidences of Local Adaptation -----	16
The Genetic Architecture of Local Adaptation -----	18
Local Adaptation in the Face of Future Climate Change -----	23
Discussion -----	27
References -----	32
Figures -----	37
Tables -----	76



## Contents of Figures

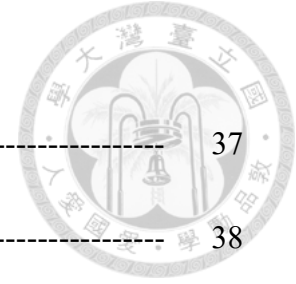


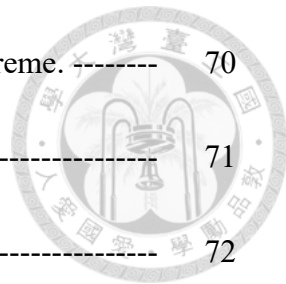
Figure 1.	a. Collection site. -----	37
	b. Population structure under $K = 2$ . -----	38
Figure 2.	a. Current suitability of <i>Musa itinerans</i> . -----	39
	b. Current resistance of <i>Musa itinerans</i> . -----	40
	c. Current least cost corridor landscape. -----	41
Figure 3.	Adaptive candidate SNPs. -----	42
Figure 4.	Minor allele frequency distribution in adaptive and non-adaptive SNPs. -----	43
Figure 5.	Allele frequency of newly mutated alleles within the ancestral and novel environmental range. -----	44
Figure 6.	Bayenv Bayes factor distribution. -----	45
Figure 7.	Cumulative importance distribution. -----	46
Figure 8.	Gradient Forest $r^2$ importance distribution. -----	47
Figure 9.	Minor allele frequency distribution of adaptive SNPs. -----	48
Figure 10.	The relationship between $r^2$ importance and slope. -----	49
Figure 11.	Slope distribution. -----	50
Figure 12.	a. BIO1 environmental map. -----	51



	b. BIO2 environmental map. -----	52
	c. BIO3 environmental map. -----	53
	d. BIO7 environmental map. -----	54
	e. BIO12 environmental map. -----	55
	f. BIO15 environmental map. -----	56
	g. BIO16 environmental map. -----	57
	h. BIO17 environmental map. -----	58
	i. BIO19 environmental map. -----	59
Figure 13.	a. BIO1 environmental map during the last glacial maximum. -----	60
	b. BIO7 environmental map during the last glacial maximum. -----	61
	c. BIO15 environmental map during the last glacial maximum. -----	62
	d. BIO17 environmental map during the last glacial maximum. -----	63
	e. BIO19 environmental map during the last glacial maximum. -----	64
Figure 14.	a. Expanding-extreme suitability of <i>Musa itinerans</i> . -----	65
	b. Expanding-extreme resistance of <i>Musa itinerans</i> . -----	66
	c. Least cost corridor landscape for the expanding extreme. -----	67
Figure 15.	a. Contracting-extreme suitability of <i>Musa itinerans</i> . -----	68
	b. Contracting-extreme resistance of <i>Musa itinerans</i> . -----	69



	c. Least cost corridor landscape for the contracting extreme. -----	70
Figure 16.	a. Genetic offset for the expanding extreme. -----	71
	b. Genetic offset for the contracting extreme. -----	72
Figure 17.	a. Extinction risk for the expanding extreme. -----	73
	b. Extinction risk for the contracting extreme. -----	74
Figure 18.	Phylogeny of Taiwanese and Chinese <i>Musa itinerans</i> . -----	75



## Contents of Tables

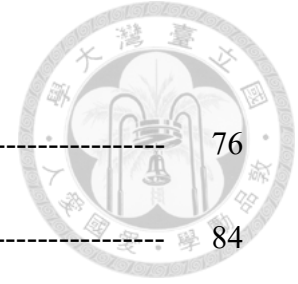


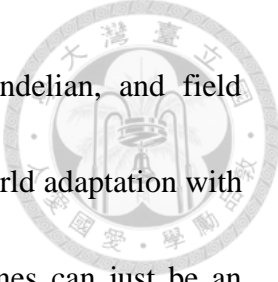
Table 1.	<i>Musa itinerans</i> collection information. -----	76
Table 2.	SSR primer information. -----	84
Table 3.	Integrated presence-only data. -----	85
Table 4.	Population coordinates and bioclimatic information. -----	92
Table 5.	MaxEnt permutation importance. -----	93
Table 6.	SNP composition. -----	94
Table 7.	SNP composition (controlled for minor allele frequency). -----	95
Table 8.	Significance test for allele frequency distribution within the novel and ancestral environmental range. -----	96
Table 9.	Significance test for Bayenv Bayes factor distribution. -----	97
Table 10.	Significance test for Gradient Forest $r^2$ importance and slope distribution. -----	98
Table 11.	SNP composition for the future. -----	99
Table 12.	Significance test for Gradient Forest $r^2$ importance and slope distribution (controlled for minor allele frequency). -----	100

## Introduction




There have long been interests in understanding the genetic architecture in adaptation. A population may adapt through advantageous alleles that are new mutations (NM) subsequently sweeping through the population (1, 2). On the other hand, the adaptive course can be modulated by standing variations (SV) in quick response to an environmental change (3). The major driver of adaptation has long been emphasized on the significance of NM in theory (4) and empirical studies (5); however, recent genomic researches provide evidence for the role of SV in adaptation (6, 7). Both of NM and SV can propel adaptation with major strong effects. NM can spread rapidly to fixation and purge variation at linked sites, resulting in hard sweeps (8). SV, on the other hand, result in soft sweeps which have weaker effects on linked sites and may therefore be more difficult to detect (9, 10).

Nevertheless, not all genetic variants underlying adaptation have dramatic effects to be detectable as Mendelian genes. Adaptation can take place at many loci simultaneously, and the response to selection is generated just by modest allele frequency shifts (11, 12). Over past decades, discovery of one large-effect quantitative trait locus after another has refuted the infinitesimal theory (13, 14), leading to a consensus that views alleles having detectably large effects as the norm (15). However, such observations could not explain



that many adaptive traits are highly quantitative rather than Mendelian, and field ecologists have been using the polygenic framework to study real-world adaptation with great success. Indeed, the overwhelming focuses on Mendelian genes can just be an exception that researchers successfully sample the outlier in the tail of effect-size distribution where a whole constellation of infinitesimals may govern the evolution with effects beneath detection limit (16-18).

Ever since Fisher, population geneticists are constantly debating the effect size of NM for adaptation to novel environments. In Fisher's geometric model (19), the organism is considered a set of phenotypic characters, each having an optimal value in the present environment as a point in the high-dimensional space. After an environmental change resulting in a new adaptive landscape, adaptation, the return to the new optimum, has been predicted by Fisher to mostly have beneficial mutations of small effect. Motoo Kimura (20) later pointed out that small-effect mutations are effectively neutral and could be lost by genetic drift, and mutations of intermediate effect size are the most likely to contribute to adaptation. Orr (21) then derived the distribution of effect sizes for an entire bout of adaptation. The effect-size distribution is proposed to have few large-effect and many small-effect mutations, and the large-effect substitutions typically happened earlier when a population encountered novel adaptive optimum, resulting in a pattern of

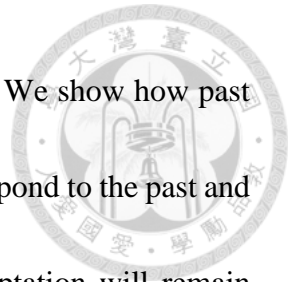


diminishing returns. This relationship is supported by empirical evidences: Microbial experimental evolution has corroborated (22-24) that adaptation involves larger-effect mutations when populations are farther from the optimum. However, studies on natural populations (25) are scarce. In this study, we provide an example from natural populations.

Taiwan is a perfect place to test the relative importance of SV and NM on environmental adaptation: Unlike oceanic islands such as Hawaii, Taiwan is a continental island where most species originated from the East Asian continent (26). The land bridge between Taiwan and China during the glacial maximum (27, 28) allows exchange of SV, and the isolation during interglacial periods enables the development of NM. One could therefore ascertain the origin of a genetic variant SV or NM by comparing the Chinese and Taiwanese populations. Moreover, the land bridge continually connected different types of environments where some are homogeneous across mainland China and Taiwan while some are not. How genetic variants (SV vs. NM) respond to such heterogeneity and homogeneity can further be explored.

Here we aim to investigate the integrative effects of the number and effect size of SV and NM as well as their responses to different types of environments. We focus on a wild banana, *Musa itinerans*, whose habitats in Taiwan are considered peripheries from ancestral area reconstructions (29), providing a good opportunity to distinguish SV from

NM as well as their effect on existing or novel adaptive landscapes. We show how past events (SV vs. NM) influence present adaptation and how they correspond to the past and present environmental range. We further assess that if present adaptation will remain under anthropogenic climate change. Combining genetics and species distribution modeling, we lastly infer the fate of *M. itinerans* by introducing a novel idea called “extinction risk”.



## Materials and 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 (Figure 1a and Table 1). Fresh leaves were harvested from nine to fifteen individuals at each site. Total genomic DNA was extracted using the standard CTAB extraction method (30). Since other commercial *Musa* species were also grown in Taiwan, we developed an indel marker for species delimitation. From previous studies (31-34), 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*

SSR primer sequences used in this study were originally developed for the genus *Musa* (35, 36), which were then applied on *Musa itinerans* (37). Previously documented primer sequences were first searched against the *Musa acuminata* DH-Pahang genome version 2 (38) 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 2). We modified each pair of primers by capping the 5' end of forward primers with M13 sequences (CACGACGTTGTAAAACGAC) and inflorescent molecules (39). SSR amplicons were run through capillary electrophoresis and the length of each allele was recorded.

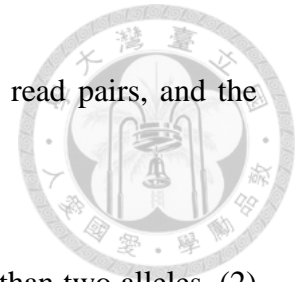
### ***Library Construction and SNP Identification***

We conducted whole genome pooled-sequencing (40) on our site-wise samples (Table 1), resulting in 24 pooled-sequencing libraries. Equal amount of DNA from ten individuals at each site were pooled, except for the PTWT population where only nine individuals were available. A library with 300 - 400 bp insert size for each pooled sample 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 (41), followed by the removal of adaptor sequences with cutadapt (42), subsequently mapped to the *Musa itinerans* reference genome assembly ASM164941v1 (43) with BWA 0.7.15 (44). 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 (45).



For the 24 pooled samples, we filtered out sites with (1) more than two alleles, (2) indels, (3) quality (QUAL) < 30, (4) quality by depth (QD) < 2, (5) call rate < 0.74, and (6) depth (DP) > genome-wide average depth plus three standard deviations, resulting in 4,200,177 SNPs. SNPs with (1) minor allele frequency (MAF) < 0.05, (2) missing data in any of the pooled-seq sample, and (3) DP per sample < 20 were further filtered out, resulting in 1,256,894 SNPs.

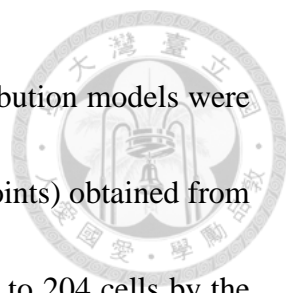
To investigate the relationship between Taiwanese and Chinese *M. itinerans*, we downloaded public data from 24 Chinese accessions (SRR6382516 - SRR6382539) (46). SNPs were called using all 24 Chinese accessions and the 24 Taiwanese pooled samples together following the pipeline described above. We did not perform any site filtering since the main objective for this combined dataset is to investigate whether specific SNPs in Taiwan also existed in China as SV. This dataset has 18,442,853 SNPs. SRR6382532 was excluded due to high missing rate. Only when evaluating the averaged expected heterozygosity between Taiwanese and Chinese populations did we filter out sites with (1) indels and (2) QUAL < 30, resulting in 15,591,923 SNPs.




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

### ***Data Analysis***

***Population Structure*** Population structure of 20 populations (Table 1) was analyzed with 14 SSR markers (Table 2). Lowland populations (C35H, WFL, THNL, PTWT, P199H, MLLYT, HDPG, TTL, NAJY, HLCN, NXIR, and DFR), east transect populations (TPS300, TPS500, TPS700, and TPS900), and west transect populations (XT400, XT700, XT1200, and XT1500) were used in the analysis. We inferred the ancestry of 244 individuals with STRUCTURE 2.3.4 (47, 48), parameterizing a run to have (1) run length of burnin and after-burnin period of 100,000, (2) admixture ancestry model, and (3) independent allele frequency model, further setting 20 runs for each *K* value.



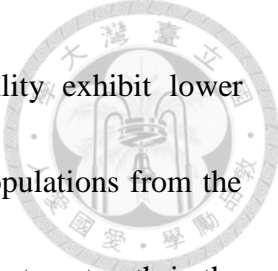
**Species Distribution Modeling** Current and future species distribution models were built for *Musa itinerans* using presence-only data (483 occurrence points) obtained from Google Street view (Table 3). Occurrence points were then reduced to 204 cells by the removal of co-occurring presence data within the same  $1 \times 1$  km grid. MaxEnt 3.4.1 (49, 50), implemented with the maximum entropy modeling approach, reports an overall niche suitability and the importance of predictors by analyzing the presence-only data as well as background (psuedo-absence) data distribution. We downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) spatial layers of 19 present-day bioclimatic variables based on high-resolution monthly temperature and rainfall data (51). Layers were selected at spatial resolution of 30 arc-second and with a mask that ranges 119.25 - 122.47 °E and 21.76 - 25.49 °N covering Taiwan. Variables showing high dependence (Pearson's correlation coefficient  $> 0.89$  calculated from ENMTools (52)) from each other were removed, resulting in nine final variables: BIO1 – mean annual temperature, BIO2 – mean diurnal range, BIO3 – isothermality, BIO7 – temperature annual range, BIO12 – annual precipitation, BIO15 – precipitation seasonality, BIO16 – precipitation of wettest quarter, BIO17 – precipitation of driest quarter, and BIO19 – precipitation of coldest quarter (Table 4).



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


To predict the species distribution under different scenarios of future climate change, we projected the present-day model onto eight future climatic conditions combining two periods (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5). Future climatic layers were obtained from the WorldClim database at spatial resolution of 30 arc-second and were developed based on two general circulation models: the Community Climate System Model (53), CCSM, and the Model for Interdisciplinary Research on Climate (54), MIROC. Species distribution models for the future were carried out using the same settings described above.

To estimate the least cost path between populations, we first generated the resistance surface by taking the reciprocal of suitability. Resistance and suitability is simply a



monotonic transformation in which locations with higher suitability exhibit lower resistance. Pairwise least cost path was then measured among 20 populations from the resistance surface, performed by SDM Toolbox v2.3 (55). While least cost path is the single line with least overall cost, we also constructed the least cost corridor between populations, allowing 1%, 2%, or 5% higher cost than the least cost value. In essence, the least cost corridors represent the realized dispersal routes of organisms along suitable habitats.

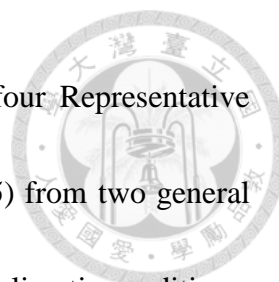
***Isolation Pattern and Adaptive SNP Identification*** To investigate the association among genetic, geographical, and environmental distance, we generated these distance matrices. Genetic distance was calculated by GenAlEx 6.503 (56, 57) from 14 SSR markers; straight geographical distance (the fly-over distance) was generated by ArcGIS 10.5 (<http://desktop.arcgis.com/en/>); environmental distance was measured as Mahalanobis distance to address the correlation among bioclimatic variables. 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. Matrix association was examined under Mantel and Partial Mantel tests. Statistical significance was examined with 1,000 permutations. We performed Mantel



tests on (1) genetic distance *vs.* fly-over distance, (2) genetic distance *vs.* resistance distance, (3) genetic distance *vs.* Mahalanobis environmental distance, and Partial Mantel tests on (4) genetic distance *vs.* Mahalanobis environmental distance while controlling for resistance distance.

We used Bayenv 2.0 (58, 59) to search for SNPs highly associated with environmental variables. Bayenv estimates the relationship between SNPs and environments while controlling the whole-genome population structure from a subset of loose linkage-disequilibrium SNPs. Loose linkage-disequilibrium SNPs were formed by sampling (1) one SNP from scaffolds more than 10 kb and less than 100 kb, (2) two SNPs from scaffolds more than 100 kb and less than 500 kb, (3) three SNPs from scaffolds more than 500 kb and less than 1000 kb, and (4) four SNPs from scaffolds more than 1000 kb. We then, for each bioclimatic variable, defined as the adaptive SNPs ones exhibiting top 1 % Bayes factor and top 5 % rho value (a nonparametric correlation coefficient capable to reduce outlier effects).

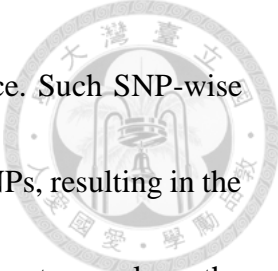
***Adaptive SNP Retention and Disruption under Climate Change*** We further investigated the fate of currently adaptive SNPs under anthropogenic climate change, performing the same Bayenv analyses of currently adaptive SNPs using future climatic



conditions. We included two time periods (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) from two general circulation models, CCSM (53) and MIROC (54), forming 16 future climatic conditions.

If a currently adaptive SNP remains strongly associated with environments, it should exhibit Bayes factor above the current adaptive threshold. We then defined as “retention” a currently adaptive SNP constantly exhibiting Bayes factor above the current adaptive threshold in all future scenarios, and defined as “disruption” a currently adaptive SNP exhibiting Bayes factor above the current adaptive threshold in none of the future scenarios.

***Genetic Offset*** We used a novel method, Gradient Forest (60, 61), to estimate the effect of environmental gradients on allele frequency differences among populations. Gradient Forest is a regression-tree based machine-learning algorithm using environmental variables to partition SNP allele frequencies. The analysis was done separately for each SNP. The “importance” measures how much of the variation in allele frequency was explained by partitioning the populations based on a specific value in an environmental variable. By making multiple regression trees (thus generating a random forest) for a SNP, the goodness-of-fit  $r^2$  of a random forest is measured as the proportion of variance explained by this random forest, which is then partitioned among

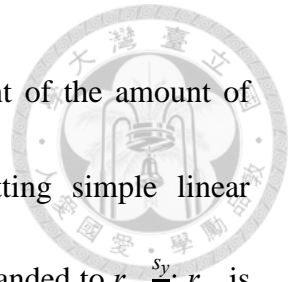


environmental variables in proportion to their conditional importance. Such SNP-wise importance of each environmental variable is then averaged across SNPs, resulting in the overall importance (of each environmental variable). SNP-wise importance along the environmental gradient can also be averaged across SNPs and summed cumulatively, resulting in the cumulative importance. After obtaining the relationship between environmental gradient and cumulative importance, one can then measure how much of the importance would change following the environmental change between current and future climates. Based on the cumulative importance results from each bioclimatic variable, the “genetic offset” (61) could then be calculated as the Euclidean distance between contemporary importance corresponding to the contemporary environmental value and future importance corresponding to the future environmental value, considering all bioclimatic variables together. Genetic offset can then be considered to be the magnitude of genetic change needed for a population to be still adaptive in the face of climate change.

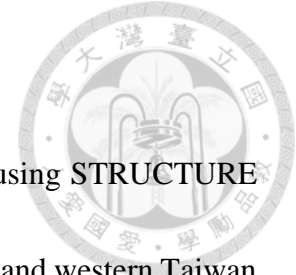
***Regression Slope*** The regression slope is not given by Gradient Forest, since it only reports the  $r^2$  importance estimate. Thus, we introduced the simple linear regression  $y=\alpha+\beta x$  to measure the regression slope. We took  $y$  as the allele frequency,  $x$  as the



standardized bioclimatic variable, and  $\beta$  (slope) as the measurement of the amount of allele frequency changes along environmental gradients. By fitting simple linear regression with the general least-square approach,  $\beta$  can then be expanded to  $r_{xy} \frac{s_y}{s_x}$ :  $r_{xy}$  is the correlation coefficient between  $x$  and  $y$ , and  $s_x$  and  $s_y$  are the standard deviation of  $x$  and  $y$ .



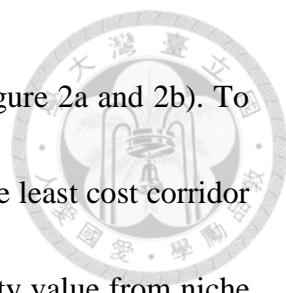
## Results



### *Population Structure and Evidences of Local Adaptation* By using STRUCTURE

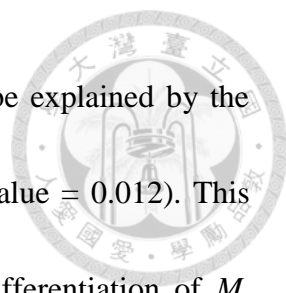
(47, 48), we observed genetic differentiation mostly between eastern and western Taiwan under  $K = 2$  (Figure 1), which is supported by the previous study that had shown the east populations being the most isolated (37). Populations in northwestern Taiwan showed admixture, implying this region could be a hybrid zone of east and west populations. However, population MLLYT stood out as having the most incongruent ancestry. The high and uniform east ancestry of MLLYT resulted from the monomorphic SSR loci exhibiting most east-specific alleles. Population MLLYT might have first been dispersed by seed vectors, such as squirrels, monkeys, and birds, followed by the massive clonal growth during the early succession (37, 62).

The niche reported (Figure 2a) was in line with the previous statement that *Musa itinerans* inhabits sunny valleys, watersheds, and hillsides with gentle slope (37). Annual mean temperature (BIO1) and precipitation of driest quarter (BIO17) turned out to be the critical bioclimatic variables determining the species range (Table 5). We then constructed resistance surface (Figure 2b) by transforming suitability into its reciprocal. Here, we observed the northeastern hillsides and Lanyang Plain exhibited the most suitable niche, mainly due to high values of BIO17. The most unsuitable environments



lay within Central Mountain Range and the southwestern plains (Figure 2a and 2b). To estimate possible connectivity between populations, we generated the least cost corridor landscape based on the resistance surface (reciprocal of the suitability value from niche modeling above) (Figure 2c). Pairwise costs of least cost paths show that northwestern Taiwan acted as a hub of connectivity (Figure 2c).

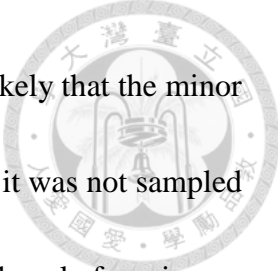
To explore the possible adaptation, we dissected the effects of geography and environment on genetic differentiation. The “fly-over” geographical distance, calculated as the straight distance between locations, could not explain the variation of genetic differentiation (Mantel’s  $r = 0.146$  and  $p$ -value = 0.062). However, if we considered that Central Mountain Range hardly possesses passages for *M. itinerans* to disperse (Figure 2c with sparse routes across Central Mountain Range), this fly-over geographical distance could be too unrealistic. We therefore used resistance distance to represent the “realized” geographical distance and found that genetic differentiation was significantly associated with resistance (Mantel’s  $r = 0.2257$  and  $p$ -value = 0.006). The environmental Mahalanobis distance of nine bioclimatic variables also showed strong association with genetic differentiation (Mantel’s  $r = 0.298$  and  $p$ -value = 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 the effect of realized



geography (resistance distance), genetic differentiation could still be explained by the Mahalanobis environmental distance (Mantel's  $r = 0.2505$  and  $p\text{-value} = 0.012$ ). This strong association implied long-term adaptation shaped genetic differentiation of *M. itinerans*.

***The Genetic Architecture of Local Adaptation*** What shaped the adaptation and what was the genetic architecture have yet to be answered. To determine the genetic architecture, we partitioned SNPs into two groups: (1) standing variations (SV) whose two alleles were found both in Taiwan and China and (2) new mutations (NM) where the polymorphism was only found in Taiwan. Bayenv (58, 59) was first employed to search for candidate SNPs underlying possible adaptation across nine bioclimatic variables. Adaptive SNPs were then defined as ones possessing top 1 % Bayes factor and 5 % rho value (Figure 3).

We found SV outnumbered NM in both adaptive and non-adaptive SNPs, and SV were even more enriched in adaptive ones (Table 6). We note, however, since adaptive SNPs also tend to have higher minor allele frequency (MAF) (Figure 4), this pattern could be confounded by allele frequency for two reasons. First, SV are more likely to have higher MAF than NM, and SNPs with higher MAF may be more likely to be detected as adaptive due to higher statistical power. Second, the designation of SV depends on



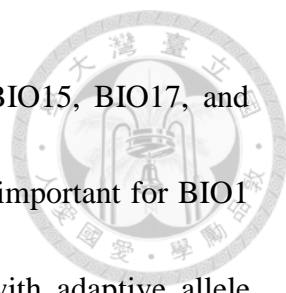
whether one could observe it in China. If a SNP has low MAF, it is likely that the minor allele is a SV in China but mis-identified as a NM in Taiwan because it was not sampled in the Chinese populations. We therefore performed the same test while only focusing on a subset of SNPs where the adaptive and non-adaptive SNPs have similar allele frequencies (ranging from the first quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each bioclimatic variable) (Figure 4). SV still disproportionately abound when allele frequency was controlled (Table 7). Here, we emphasize that SV prevail over NM in number in adaptation.

In addition to the relative number, does the difference between SV and NM depend on the environments? We therefore tested whether newly mutated alleles from NM conferring environmental adaptation are more often observed in novel environments in Taiwan relative to the ancestral Chinese environmental range experienced by SV. Indeed, when we separated the Taiwanese populations into those within the Chinese environmental range and those with novel environments, allele frequencies of newly mutated adaptive alleles are higher in the latter set of populations, especially for the precipitation-related variables (Figure 5 and Table 8). Among these variables, precipitation of driest quarter (BIO17) and precipitation of coldest quarter (BIO19)

manifest the allele frequency difference where newly mutated alleles are extremely enriched in novel environments.



However, similar to standard genetic mapping where a trait could be controlled by few Mendelian genes or numerous polygenes each with minor effect, the number of candidate SNPs may not necessarily reflect the overall importance of SV and NM. To investigate the effect size of SV and NM in environmental adaptation, we compared their Bayes factors directly from Bayenv. Mean diurnal range (BIO2), precipitation seasonality (BIO15), precipitation of driest quarter (BIO17), and precipitation of coldest quarter (BIO19) showed the most significant differences between SV and NM (Figure 6 and Table 9), where NM consistently had higher Bayes factor and therefore stronger effect size than SV. On the other hand, SV had significantly stronger effect size than NM for annual mean temperature (BIO1) and temperature annual range (BIO7) (Figure 6 and Table 9). We further employed Gradient Forest to identify influential bioclimatic variables as well as the proportion of adaptive allele frequency changes explained by these variables. Precipitation of driest quarter (BIO17) turned out to be the most important determinant for adaptive SNP allele frequency turnover (Figure 7 and 8), suggesting differential local adaptation among the Taiwanese populations exists in response to this (BIO17) and other correlated climatic variables (BIO15 and BIO19). Interestingly, for



climatic variables with strong effect on local adaptation (BIO2, BIO15, BIO17, and BIO19), NM have stronger response than SV. Again, SV are more important for BIO1 and BIO7, the two climatic variables with weakest association with adaptive allele frequency changes (Figure 8). We note, it is of no necessity to further partition the adaptive SNPs to control for allele frequency, since the minor allele frequencies of adaptive NM and SV SNPs show no obvious differences (Figure 9).

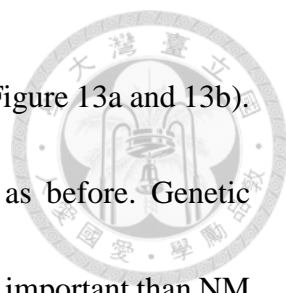
The “importance” estimated by Gradient Forest is essentially equivalent to  $r^2$ , representing the amount of allele frequency variation explained by environmental gradients. Assuming a simple linear relationship between allele frequency and environment, the value of  $r^2$  only represents how well each data point (a population) fits along the regression line. We are, however, also interested in the regression slope: the amount of allele frequency changes along environmental gradients (Figure 10). SNPs pertaining to precipitation seasonality (BIO15), precipitation of driest quarter (BIO17), and precipitation of the coldest quarter (BIO19) display the largest overall slope, and NM exhibit larger slope than SV (Figure 11). SNPs relevant to annual mean temperature (BIO1) and temperature annual range (BIO7) show the smallest slope with SV being more important than NM (Figure 11), again consistent with previous comparisons from Bayenv Bayes Factor (Figure 6) and Gradient Forest importance (Figure 8). Therefore at least in

our case, NM with larger effect size per SNP are associated with the adaptation to novel environments outside of the ancestral niche range, consistent with previous population genetics modeling results (19-21).



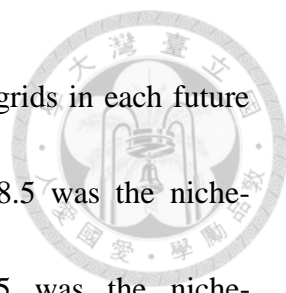
In summary, the observed patterns could be integrated with the unique climate of Taiwan. Northern Taiwan experienced northeast monsoon during winter and had higher precipitation during the typical winter dry season elsewhere of the species range (Figure 12f, 12h, and 12i), and such a pattern has at least been maintained since the last glacial maximum (Figure 13c, 13d, and 13e). Such novel environments imposed strong selection pressure and novel adaptive optimum to the ancestral immigrant population from China. Adaptation to such environmental gradient is strong (with highest Bayes factor,  $r^2$ , and slope among these bioclimatic variables) (Figure 6, 8, and 11; Table 9 and 10) especially for NM, where newly mutated alleles are strongly associated with novel environments (Figure 5 and Table 8). More importantly, for the major driver of adaptation (BIO17), the greatest increment of importance lies in between 200 mm and 300 mm (Figure 7) which are the very separating points to distinguish the novel Taiwan environment from the ancestral China environment (Figure 12h). Annual mean temperature (BIO1) and temperature annual range (BIO7) are the other extreme: the environmental gradient within Taiwan is well within the ancestral Chinese environmental range (Figure 12a and






12d), which can be traced back to the last glacial maximum as well (Figure 13a and 13b). The immigrant population experienced similar adaptive optimum as before. Genetic response to the selection pressure is lower, and therefore SV are more important than NM (Figure 6, 8, and 11; Table 9 and 10). Finally, Taiwan has higher precipitation of wettest quarter (BIO16) and therefore higher annual precipitation (BIO12) than China (Figure 12e and 12g). The per SNP contribution to adaptation is similar between SV and NM (Figure 6, 8, and 11; Table 9 and 10), suggesting that NM may not be associated with novel environments in every case. This may be related to the fact that wet-season precipitation did not impose as strong selection as dry-season precipitation did. In summary, adaptation happened through the assortment of SV for a new territory with similar adaptive landscape and optimum. For adaptation to novel environments and a new adaptive landscape, NM with larger effective size are preferred.

***Local Adaptation in the Face of Future Climate Change*** In addition to understanding how past events (SV vs. NM) affected present adaptation, we are also concerned with how these factors affect the future of this species under anthropogenic climate change. We therefore predicted 16 future outcomes according to the current one, including two time periods (2050 and 2070) and four Representative Concentration Pathways (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) from two general circulation models,




CCSM (53) and MIROC (54). We summed the suitability over all grids in each future prediction, generating two extreme scenarios: CCSM4-2070-RCP8.5 was the niche-expanding extreme (Figure 14), whereas MIROC-2070-RCP4.5 was the niche-contracting extreme (Figure 15). Under the expanding extreme, Central Mountain Range could not impose strong resistance as the present (Figure 14a and 14b), possibly due to the ascending temperature overlapping with the suitable temperature range. The contracting extreme, however, showed exacerbating isolation of the eastern populations (Figure 15a and 15b). Under both extremes, the northeast remains as the suitable environment (Figure 14 and 15).

For perennial and sessile *M. itinerans*, the contemporary adaptation may no longer hold promise in the face of climate change. We used Bayenv to detect whether the currently adaptive SNP is still strongly associated with the future climatic conditions. Currently adaptive SNPs remaining high association with environment in the future are defined as “retention”, while adaptive ones no longer highly associated with environment are defined as “disruption”. Different from the present pattern that SV disproportionately abound in adaptive SNPs, we saw no clear tendency for any set of adaptive SNPs towards retention or disruption (Table 11) in terms of SNP number. For effect size, we used the genetic offset value calculated from the Gradient Forest results to estimate genetic



mismatch, which is related to the magnitude of allele frequency turnover perturbed by the future climatic conditions (61). We estimated the overall genetic offset in the 16 future climatic scenarios, especially focusing on two extremes: CCSM4-2070-RCP8.5 (expanding extreme) and MIROC-2070-RCP4.5 (contracting extreme). Adaptive SNPs as a whole encounter differing genetic offset in both extremes: Northwestern and mid-eastern Taiwan experience large genetic offset in the expanding extreme (Figure 16a), while southwestern Taiwan experiences large genetic offset in the contracting extreme (Figure 16b).

Finally, we investigated the fate of *M. itinerans*, developing a novel concept called “extinction risk” to bridge the knowledge gaps between genomics and biogeographical modeling. Traditional species distribution model like MaxEnt assumes a species reacts identically to the environment, overlooking local adaptation from within-species polymorphism. Hence, we took extinction risk, integrating species-level and population-level responses, as the division of genetic offset (estimate from Gradient Forest) by the suitability (estimate from MaxEnt). Northeastern Taiwan exhibits the lowest extinction risk as a climate change refugium, for it possesses the highest suitability and lowest genetic offset. Surprisingly, western Taiwan displays very high extinction risk (Figure 17) despite the high suitability from MaxEnt (Figure 14a and 15a). Here we emphasize that



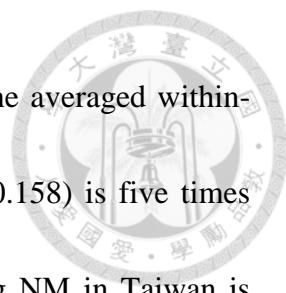
MaxEnt suitability is a measure of species-wide average, assuming all members of a species equally suitable to all environments of the species range. With differential local adaptation, our extinction risk therefore shows that despite some other population might be suitable for western Taiwan in the future, the current locally adaptive population may mismatch with future climates. Nevertheless, if we assume the dispersal distance and gene flow are high enough, local populations may obtain well-matched alleles to rescue their adaptive mismatch and therefore reduce their extinction risk. With higher connectivity, northwestern Taiwan may take advantage of alleles matching its environment from some other population against genetic offset (Figure 14c and 15c). However, southwestern Taiwan remains relatively isolated, exposing it in high risk (Figure 14c and 15c). Above all, we underscore the importance of introducing extinction risk to investigate the fate of species under climate change.

## Discussion



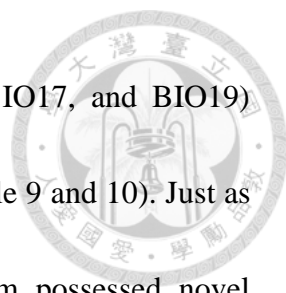
Assessing the within-species variation in climate association is the crucial first step to understand species susceptibility to fluctuating environment. Here, we investigate the relationship between genetic, geographical, and environmental distance in *Musa itinerans* and find evidences for environmental adaptation. We detect adaptive SNPs highly correlated with nine bioclimatic variables, and further probe into the past, present, and future genetic architecture of adaptive substitutions.

When tracing the present adaptive course back to the past events (SV vs. NM), we were aware of the problem of wrong classification of NM. SV can be designated as NM due to the limited samples in China which only possess a part of variation that do not contain Taiwanese alleles. Therefore, we evaluated the possibility of mis-categorizing SV into NM in three ways. First, we assessed the phylogeny of our Taiwanese populations, Chinese accessions, and one outgroup species *Musa acuminata*. All Taiwanese populations are clustered into a clade with the shortest branch length, being a sister group to one of Chinese populations (Figure 18). This demonstrates that Chinese populations capture much more variations than our Taiwanese populations which are more recently emerging. Thus, we are unlikely to falsely identify NM due to insufficient variation of Chinese samples. Second, we assessed the averaged expected heterozygosity of

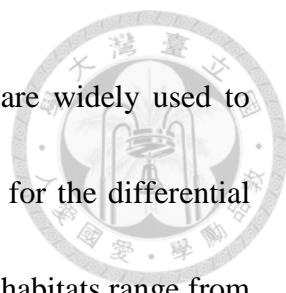


Taiwanese and Chinese populations from the integrated dataset. The averaged within-population heterozygosity among five Chinese populations ( $H_e = 0.158$ ) is five times higher than that of Taiwanese populations ( $H_e = 0.033$ ), indicating NM in Taiwan is unlikely to be mis-identified due to the insufficient sample size in China. Third, if an adaptive SNP belongs to SV, the allele which is a minor one for Taiwanese populations tends to exhibit higher allele frequency in China (Figure 9). Additionally, our previous result has revealed that adaptive SV and NM show no obvious differences (Figure 9) in minor allele frequency (MAF). Therefore, with higher minor allele frequency for SV in China, it is much more unlikely for us to designate adaptive SV as adaptive NM due to low allele frequency. We further selected the adaptive SNPs whose MAF ranks top 50 % in all adaptive ones to perform Gradient Forest analysis. In this way, we have less chance to miss such an allele in China with higher MAF. The significance test for  $r^2$  importance and slope distribution from selected adaptive SNPs is quite similar as before (Table 12). All results demonstrate our ascertainment of NM is of no risk.

Our investigation into effect sizes of SV and NM with Bayenv (58, 59) and Gradient Forest (60, 61) aid in the elaboration of how past events modulated adaptation. Orr's model (21) has predicted the distribution of effect sizes for an entire bout of adaptation where early substitutions have larger effect than later ones. We show that NM are more



strongly associated with precipitation-related variables (BIO15, BIO17, and BIO19) where they have larger effect size than SV (Figure 6, 8, and 11; Table 9 and 10). Just as Orr's prediction, Taiwan region during the last glacial maximum possessed novel environments regarding these variables (Figure 13c, 13d, and 13e), and thus more distant adaptive optima for the immigrant population. The selection pressure imposed by these climatic factors has never stopped, since Taiwan is still experiencing northeast monsoon now, resulting in maintenance of larger effect size for NM. We too inspect two bioclimatic variables (BIO1 and BIO7) where SV have larger effect size than NM (Figure 6, 8, and 11; Table 9 and 10) during the last glacial maximum. Both climatic factors of Taiwan region have been well within those of mainland China (Figure 13a and 13b), forming a similar adaptive landscape for the immigrant population, suggesting a similar adaptive optimum and landscape has last from the last glacial maximum to the present where a potpourri of SV contributes to adaptation. Here we provide another perspective to recent research showing that SV contributes to adaptation (63, 64). We show that SV indeed dominate over NM in number (63). However, the effect size of SV and NM hinges on experienced environments: SV have larger effect size when encompassed by ancestral environments, while NM have larger one encountering novel environments.



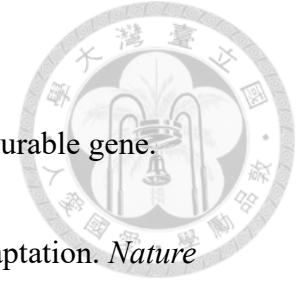
Conventional ecological niche models like MaxEnt (49, 50) are widely used to predict future species distribution, but such models do not account for the differential adaptation below species level. In the case of *Musa itinerans*, whose habitats range from 100 m to 1,500 m (Table 1) and distributing along a latitudinal gradient in Taiwan, differentially local adaptation should not be neglected. The difference between the major determinant of species range (BIO1) and driver of population-wise adaptation (BIO17) demonstrates the need to consider local adaptation. Comparison between suitability and the newly introduced extinction risk manifests the modeling limit where MaxEnt assumes homogeneity within the species. With the incorporation of genetic offset, we can therefore recognize the potential risk among western populations which possess ostensibly high suitability (Figure 14a, 15a, and 17). Methods for assessing impacts of climate change which hinge on the single species distribution model may disregard the differential adaptation, probably biasing and misplacing the conservational efforts. Our results show that the integration of genetics-environment association and species distribution model can improve predictions by revealing the mismatch between present adaptation and future climates. Such amelioration can aid in the protection of *M. itinerans* by effectively conserving the adjacent populations with relatively lower extinction risk. In particular, populations near Alisan National Scenic Area (in southwestern Taiwan) and Hsinchu-




Yilan area (in northeastern Taiwan) are proposed to be the best choice for *in situ* conservation (37).



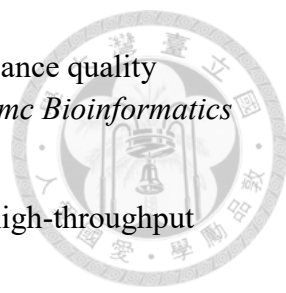
## References

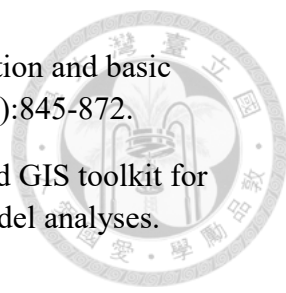


1. Smith JM & Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical research* 23(1):23-35.
2. Barton N (1998) Evolutionary biology - The geometry of adaptation. *Nature* 395(6704):751-752.
3. Barrett RD & Schluter D (2008) Adaptation from standing genetic variation. *Trends Ecol Evol* 23(1):38-44.
4. Barton NH (2000) Genetic hitchhiking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 355(1403):1553-1562.
5. Linnen CR, Kingsley EP, Jensen JD, & Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science* 325(5944):1095-1098.
6. Jones FC, *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55-61.
7. Reid NM, *et al.* (2016) The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* 354(6317):1305-1308.
8. Hermisson J & Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169(4):2335-2352.
9. Przeworski M, Coop G, & Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution* 59(11):2312-2323.
10. Teshima KM, Coop G, & Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research* 16(6):702-712.
11. Pritchard JK, Pickrell JK, & Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology : CB* 20(4):R208-R215.
12. Hancock AM, Alkorta-Aranburu G, Witonsky DB, & Di Rienzo A (2010) Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365(1552):2459-2468.
13. Orr HA (1999) The evolutionary genetics of adaptation: a simulation study. *Genetical research* 74(3):207-214.
14. Orr HA (2005) The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* 6(2):119-127.

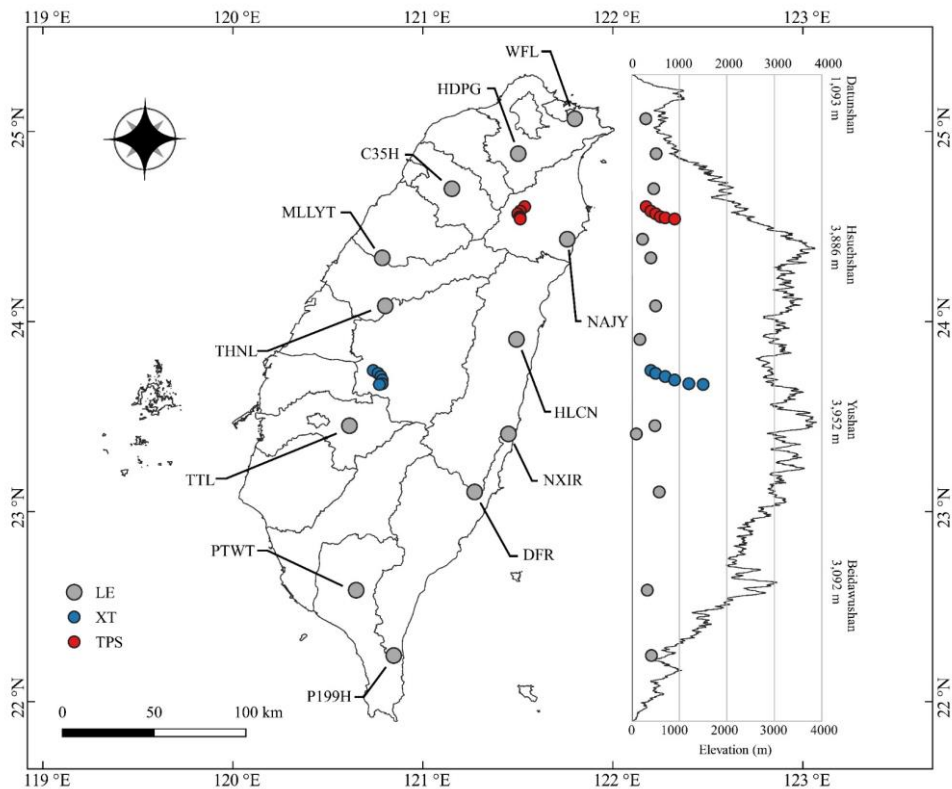
- 
15. Bell G (2009) The oligogenic view of adaptation. *Cold Spring Harb Symp Quant Biol* 74:139-144.
  16. Fisher RA (1918) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* 52(2):399-433.
  17. Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66(1):1-17.
  18. Lynch M (2007) *The origins of genome architecture* (Sinauer Associates, Sunderland, Mass.) pp xvi, 494 p.
  19. Fisher RA (1930) *The genetical theory of natural selection* (Clarendon Press, Oxford, England) pp xiv, 272-xiv, 272.
  20. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge).
  21. Orr HA (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* 52(4):935-949.
  22. McGee LW, *et al.* (2016) Synergistic Pleiotropy Overrides the Costs of Complexity in Viral Adaptation. *Genetics* 202(1):285-295.
  23. Sousa A, Magalhaes S, & Gordo I (2012) Cost of Antibiotic Resistance and the Geometry of Adaptation. *Molecular Biology and Evolution* 29(5):1417-1428.
  24. Gifford DR, Schoustra SE, & Kassen R (2011) The Length of Adaptive Walks Is Insensitive to Starting Fitness in *Aspergillus Nidulans*. *Evolution* 65(11):3070-3078.
  25. Rogers SM, *et al.* (2012) Genetic Signature of Adaptive Peak Shift in Threespine Stickleback. *Evolution* 66(8):2439-2450.
  26. Bramwell D & Caujapé-Castells J (2011) *The Biology of Island Floras* (Cambridge University Press, Cambridge).
  27. Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* 27(5):1153-1167.
  28. Liew PM & Hsieh ML (2000) Late Holocene (2 ka) sea level, river discharge and climate interrelationship in the Taiwan region. *J Asian Earth Sci* 18(4):499-505.
  29. Janssens SB, *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(4):1453-1465.
30. Doyle JJ & Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19(1):11-15.
31. White TJ, Bruns T, Lee S, & Taylor J (1990) 38 - AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. *PCR Protocols*, eds Innis MA, Gelfand DH, Sninsky JJ, & White TJ (Academic Press, San Diego), pp 315-322.
32. Taberlet P, Gielly L, Pautou G, & Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant molecular biology* 17(5):1105-1109.
33. Oxelman B, Lidén M, & Berglund D (1997) Chloroplast rps16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206(1/4):393-410.
34. Hurr KA, Lockhart PJ, Heenan PB, & Penny D (1999) Evidence for the recent dispersal of *Sophora* (Leguminosae) around the Southern Oceans: molecular data. *Journal of Biogeography* 26(3):565-577.
35. Lagoda PJ, *et al.* (1998) Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Mol Ecol* 7(5):659-663.
36. Crouch HK, Crouch JH, Jarret RL, Cregan PB, & Ortiz R (1998) Segregation at microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Sci* 38(1):211-217.
37. Chiu HL (2004) The collection, evaluation and analysis of genetic diversity of *Musa formosana* (Warb.) Hayata native in Taiwan. Doctoral dissertation (National Taiwan University).
38. Martin G, *et al.* (2016) Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* 17:243.
39. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18(2):233-234.
40. 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(11):749-763.

- 
41. Cox MP, Peterson DA, & Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *Bmc Bioinformatics* 11.
  42. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10-12.
  43. Wu W, *et al.* (2016) Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Sci Rep* 6:31586.
  44. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
  45. McKenna A, *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20(9):1297-1303.
  46. Wu W, Ng WL, Yang JX, Li WM, & Ge XJ (2018) High cryptic species diversity is revealed by genome-wide polymorphisms in a wild relative of banana, *Musa itinerans*, and implications for its conservation in subtropical China. *BMC Plant Biol* 18(1):194.
  47. Pritchard JK, Stephens M, & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945-959.
  48. Falush D, Stephens M, & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.
  49. Phillips SJ, Dud M, & Schapire RE (2004) A maximum entropy approach to species distribution modeling. in *Proceedings of the twenty-first international conference on Machine learning* (ACM, Banff, Alberta, Canada), p 83.
  50. Phillips SJ, Anderson RP, & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3-4):231-259.
  51. 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(15):1965-1978.
  52. Warren DL, Glor RE, & Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*.
  53. Gent PR, *et al.* (2011) The Community Climate System Model Version 4. *J Climate* 24(19):4973-4991.

- 
54. Watanabe S, *et al.* (2011) MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci Model Dev* 4(4):845-872.
55. 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(7):694-700.
56. Peakall ROD & Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6(1):288-295.
57. Peakall R & Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28(19):2537-2539.
58. Coop G, Witonsky D, Di Rienzo A, & Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411-1423.
59. Gunther T & Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics* 195(1):205-220.
60. Ellis N, Smith SJ, & Pitcher CR (2012) Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93(1):156-168.
61. Fitzpatrick MC & Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* 18(1):1-16.
62. Liu A-Z, Li D-Z, Wang H, & Kress WJ (2002) Ornithophilous and Chiropterophilous Pollination in *Musa itinerans* (Musaceae), a Pioneer Species in Tropical Rain Forests of Yunnan, Southwestern China. *Biotropica* 34(2):254-260.
63. Lai YT, *et al.* (2019) Standing genetic variation as the predominant source for adaptation of a songbird. *Proc Natl Acad Sci U S A* 116(6):2152-2157.
64. Ahrens CW, Byrne M, & Rymer PD (2019) Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. *Molecular Ecology* 28(10):2502-2516.

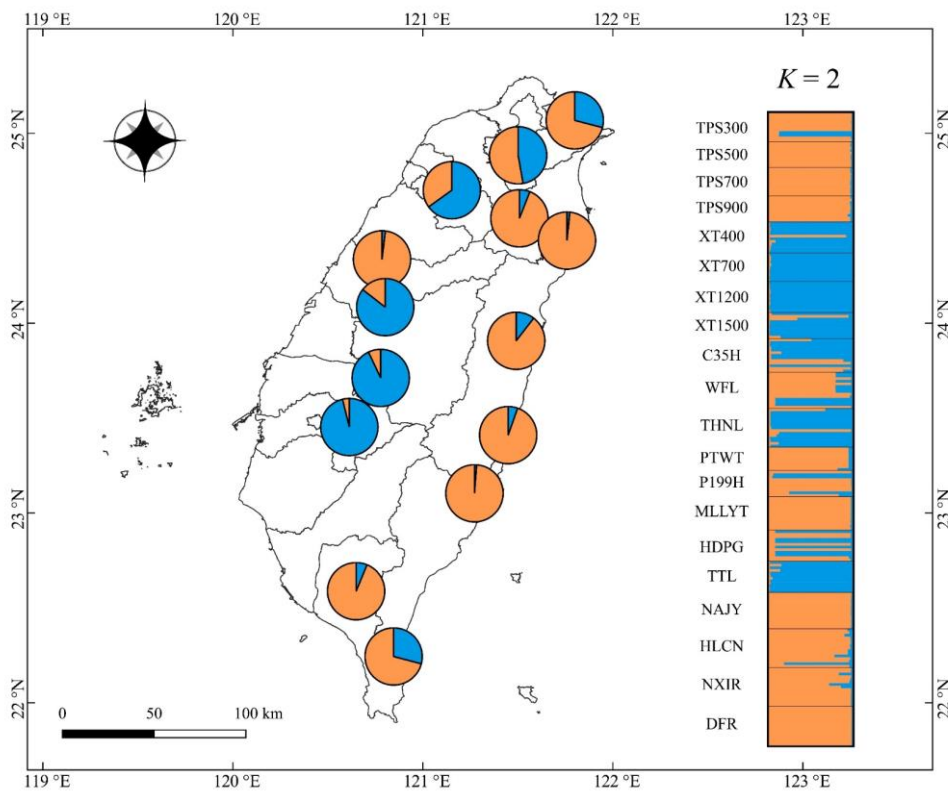
## Figures



**Figure 1a. Collection site.**

Solid circles represent collection locations corresponding to their coordinates and elevation: Gray circles indicate populations of low elevation (LE); blue circles indicate populations of Xitou transect (XT); red circles indicate populations of Taipingshan transect (TPS). Map template is provided by \*Cheng-Tao Lin.

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

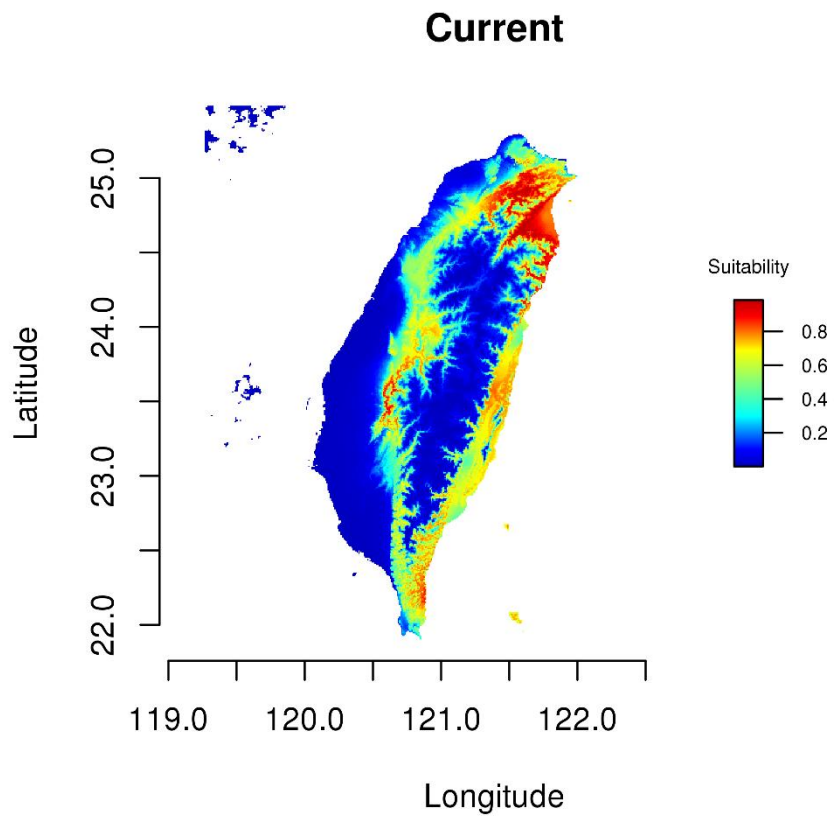


**Figure 1b. Population structure under  $K = 2$ .**

Individual ancestry is plotted on the right side, while population ancestry is plotted on map with a pie chart. Map template is provided by \*Cheng-Tao Lin.

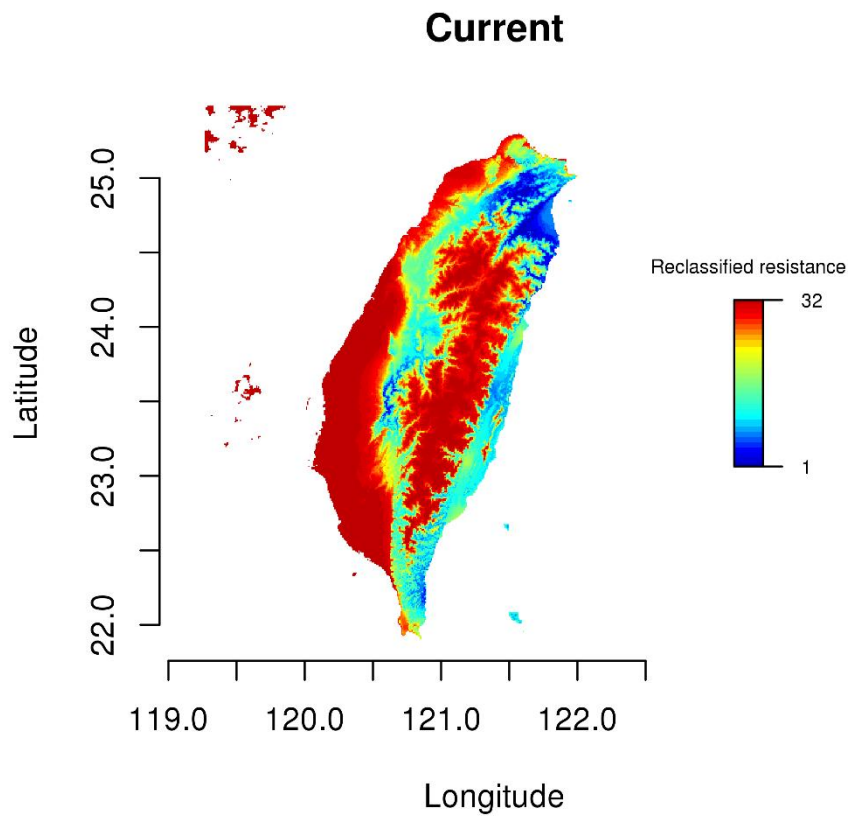
\*Cheng-Tao Lin (2018) QGIS template for displaying species distribution by horizontal and vertical view in Taiwan. URL: [https://github.com/mutolisp/distrmap\\_tw.qgis](https://github.com/mutolisp/distrmap_tw.qgis). DOI: 10.5281/zenodo.1493690





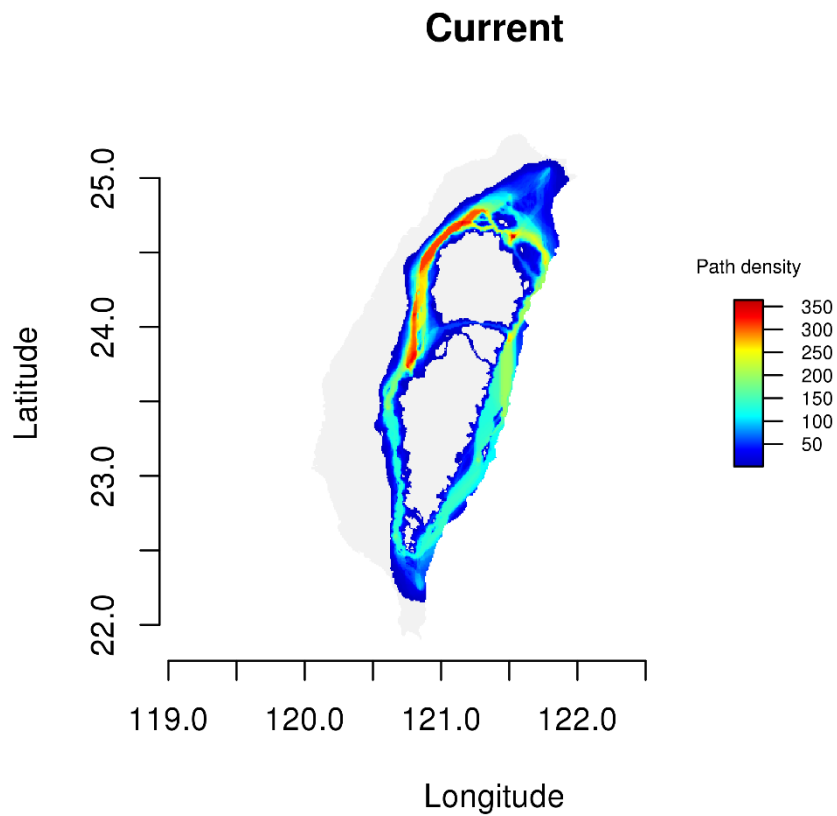
**Figure 2a. Current suitability of *Musa itinerans*.**

Suitability is provided by MaxEnt in cloglog output.



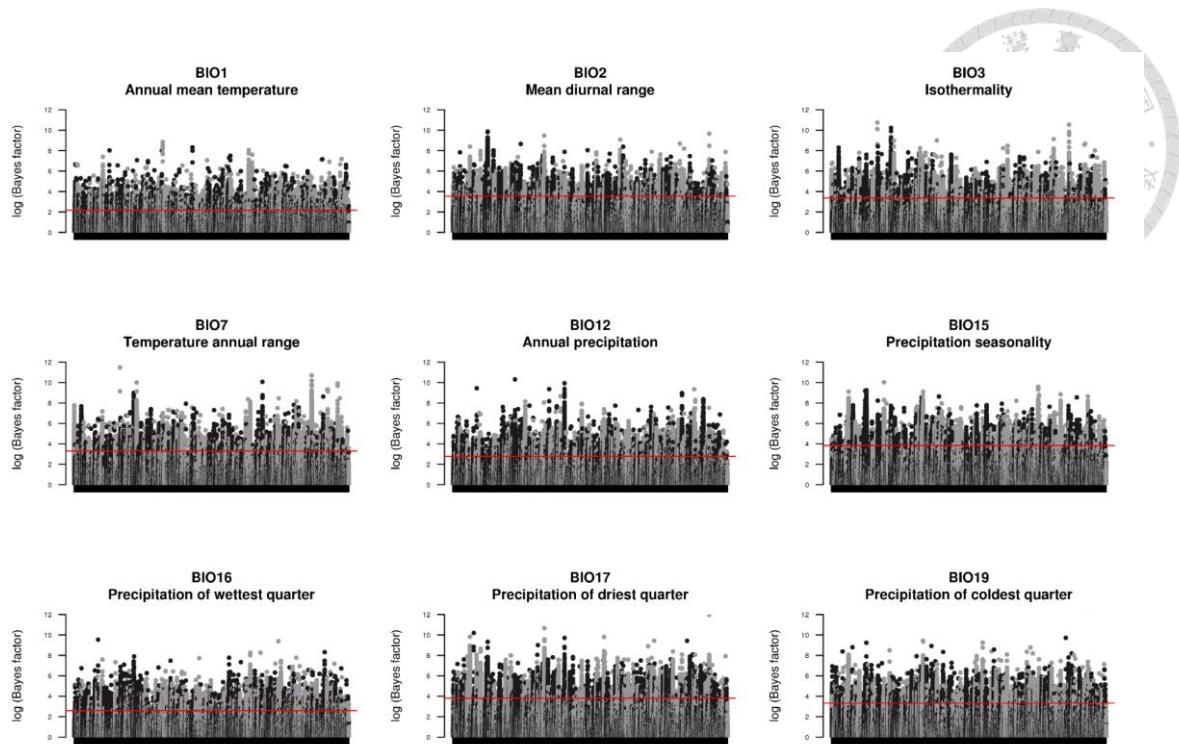
**Figure 2b. Current resistance of *Musa itinerans*.**

Current resistance is the reciprocal of current suitability from MaxEnt, and is then reclassified into 32 color blocks.



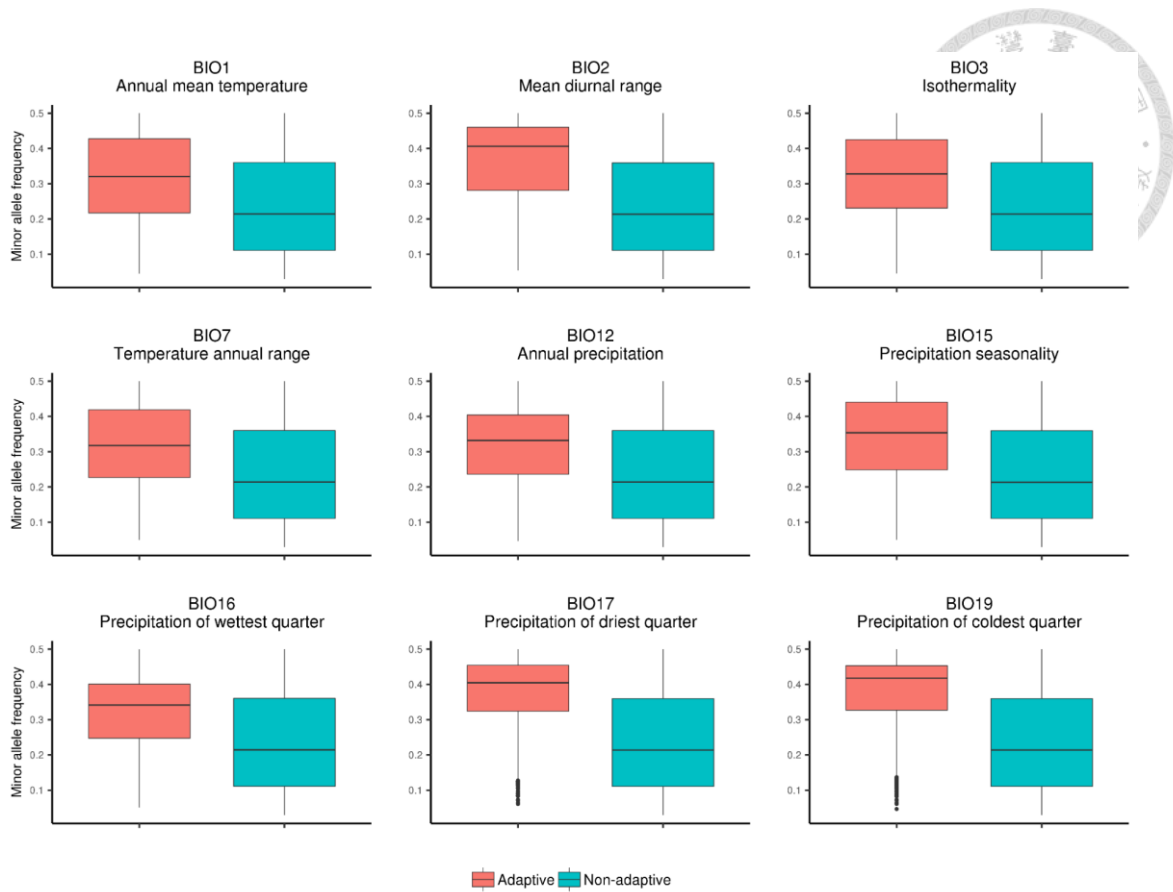
**Figure 2c. Current least cost corridor landscape.**

Least cost corridor landscape integrates all least cost corridors which allow paths with 1%, 2%, or 5% higher cost than the least cost value. Path density is colored according to the amount of overlapping paths.



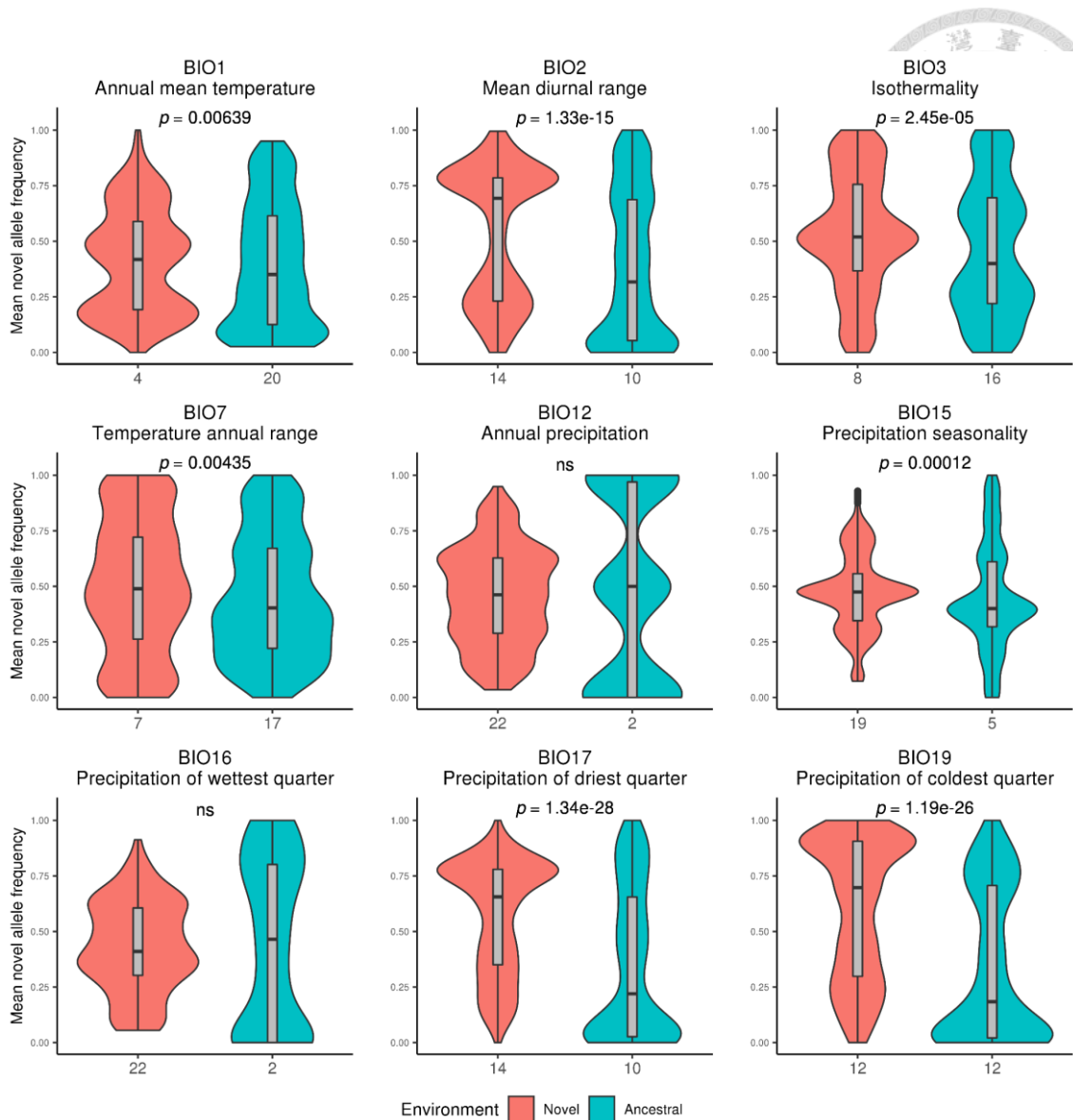
**Figure 3. Adaptive candidate SNPs.**

Bayenv detects adaptive candidates in total SNPs (1,256,894 SNPs) across nine bioclimatic variables. Red solid line shows the top 1 % Bayes factor value as the threshold.



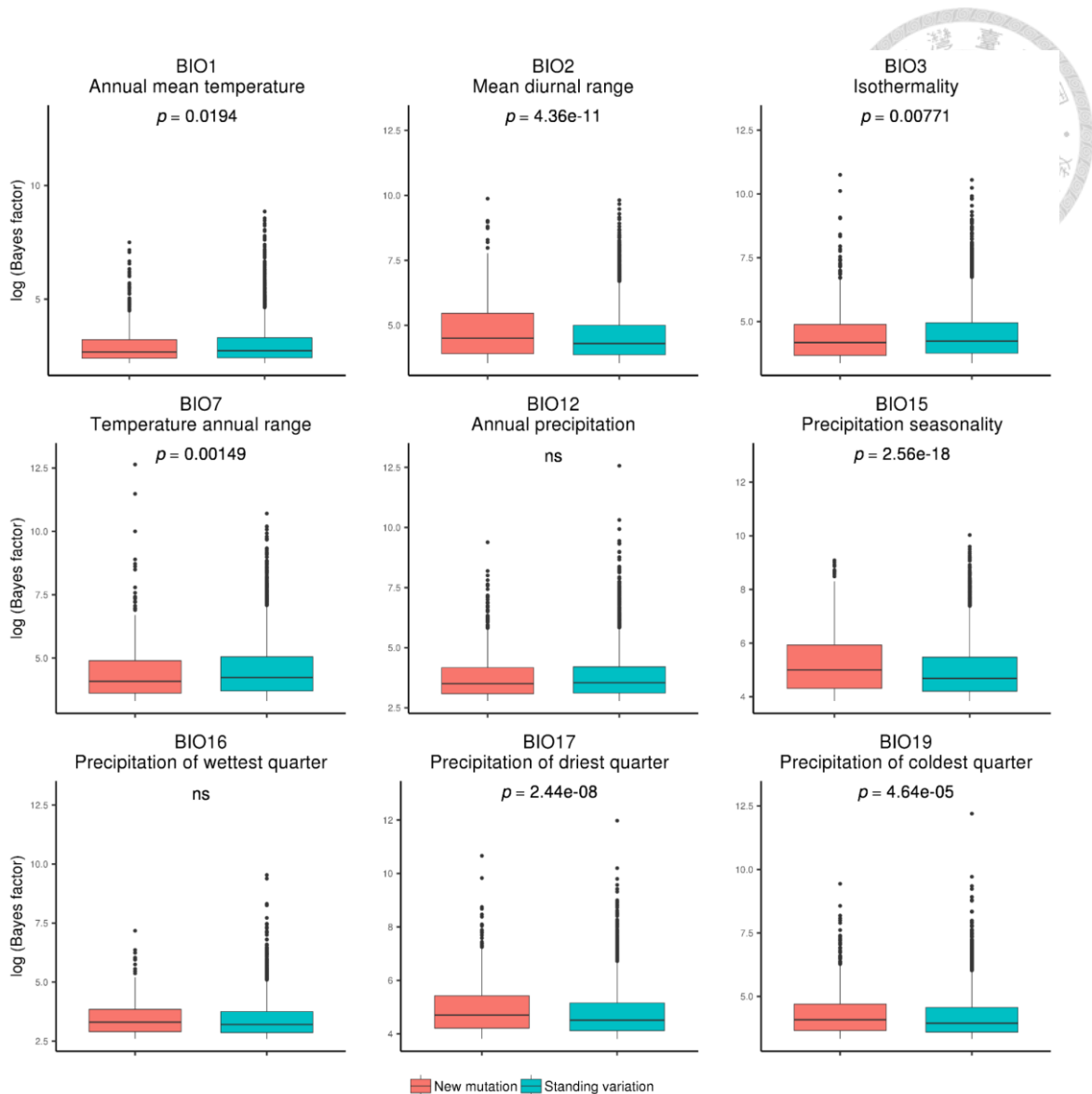
**Figure 4. Minor allele frequency distribution in adaptive and non-adaptive SNPs.**

Minor allele frequency is plotted for adaptive and non-adaptive SNPs separately for each bioclimatic variable.



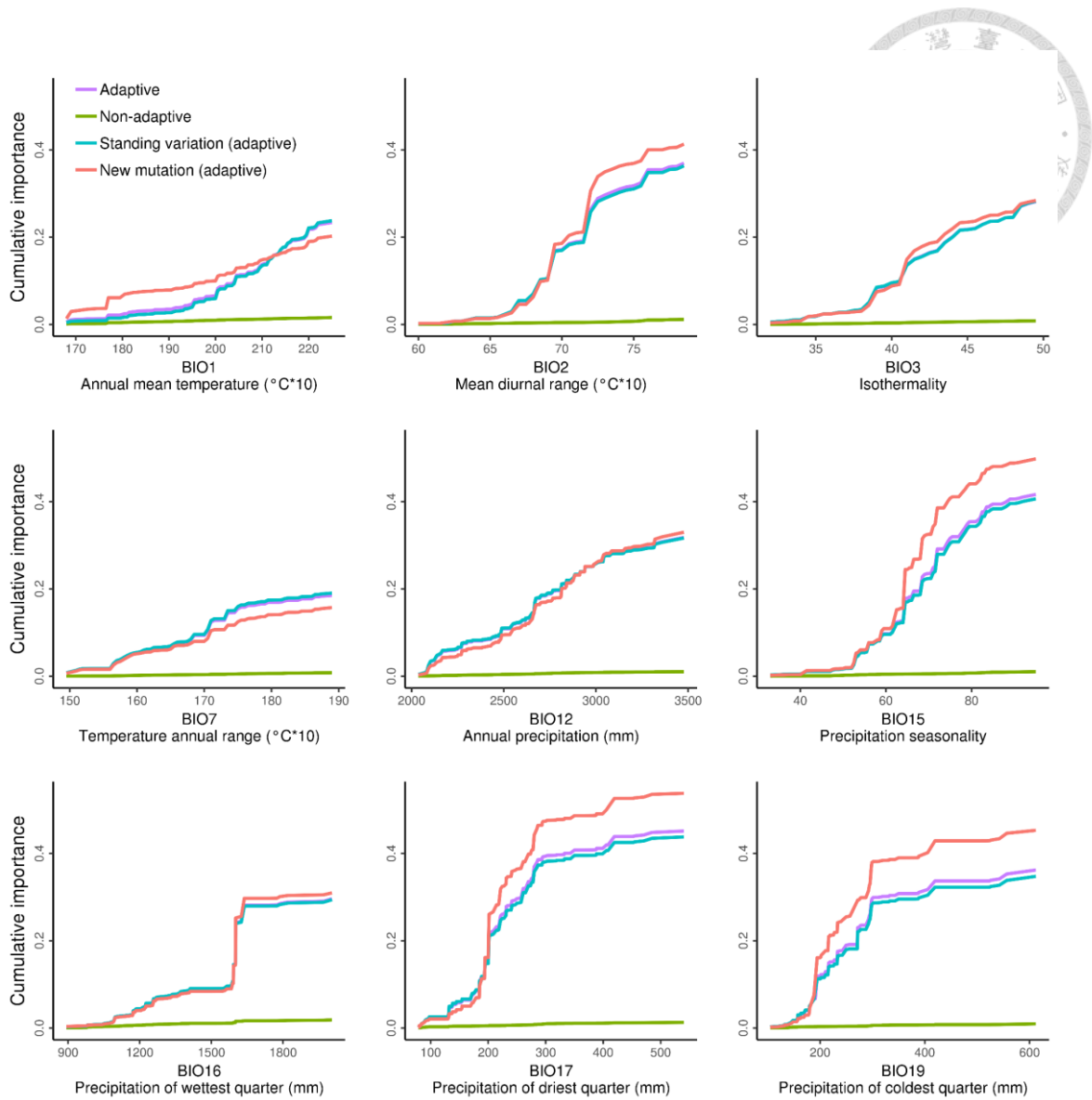
**Figure 5. Allele frequency of newly mutated alleles within the ancestral and novel environmental range.**

Averaged allele frequency of newly mutated alleles which are within the ancestral or novel environmental range in Taiwan is plotted. Statistical significance from paired *t*-test between the novel and ancestral environmental range for each bioclimatic variable is shown. The value on x-axis denotes the number of Taiwanese populations within the ancestral or novel environmental range.



**Figure 6. Bayenv Bayes factor distribution.**

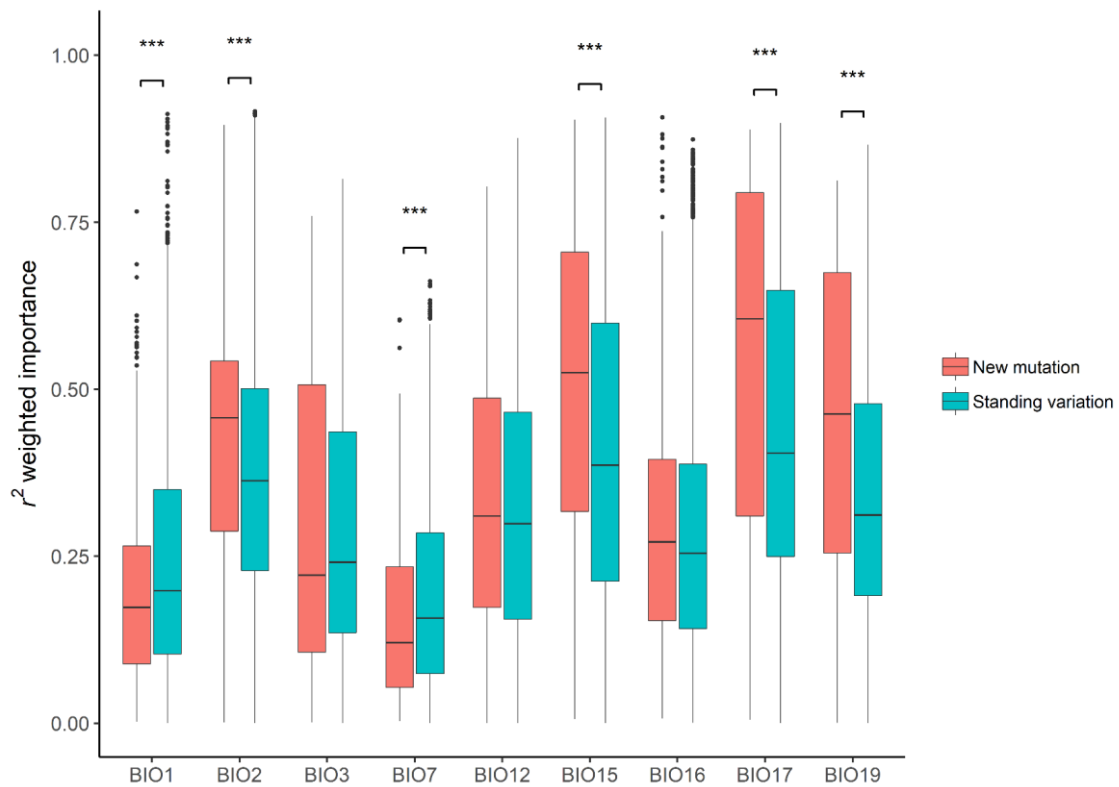
Bayes factor is plotted for adaptive new mutation and standing variation. Statistical significance from Wilcoxon rank sum test between new mutation and standing variation for each bioclimatic variable is shown.



**Figure 7. Cumulative importance distribution.**

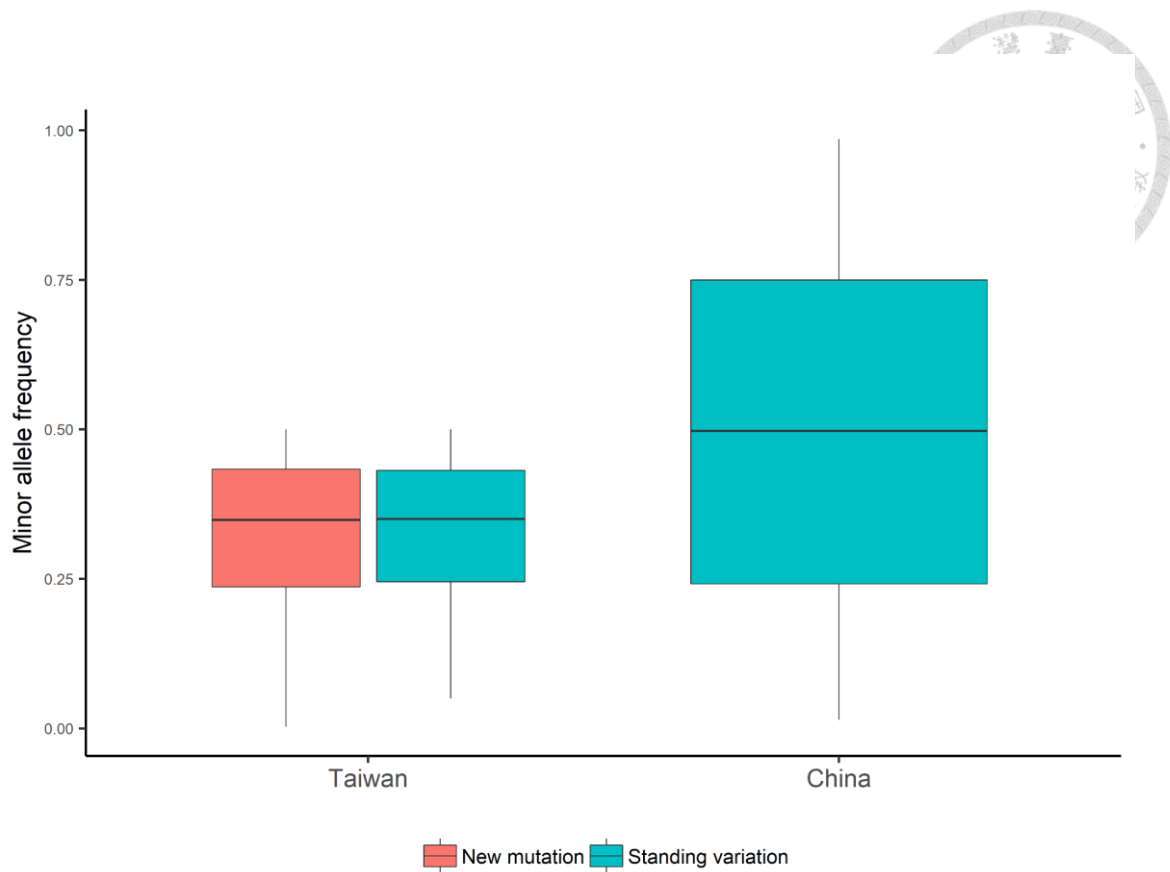
Gradient Forest cumulative importance is plotted for four sets of SNPs.





**Figure 8. Gradient Forest  $r^2$  importance distribution.**

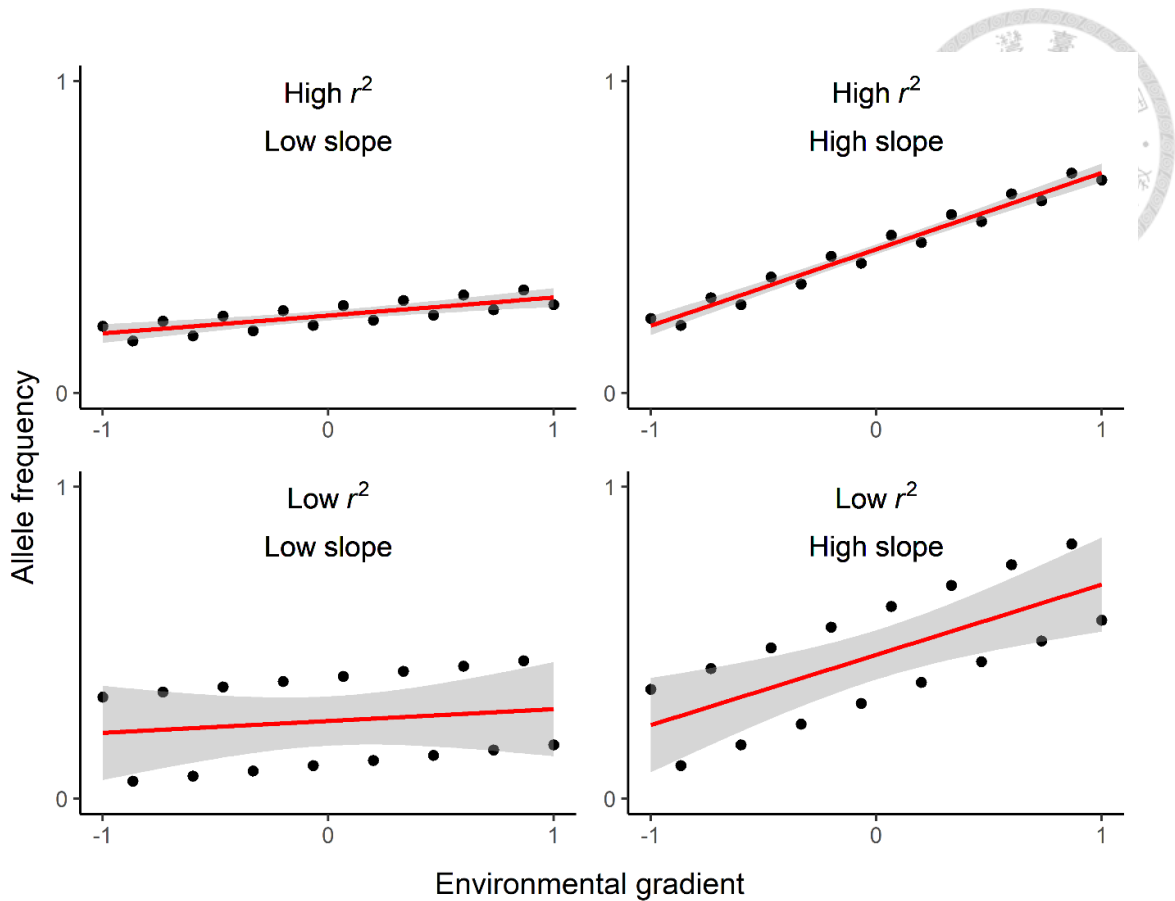
$r^2$  importance is plotted for adaptive new mutation and standing variation separately for each bioclimatic variable. Statistical significance from Welch two sample  $t$ -test between new mutation and standing variation for each bioclimatic variable is shown (\*\*\*) ( $p$ -value  $< 0.001$ ).



**Figure 9. Minor allele frequency distribution of adaptive SNPs.**

Minor allele frequency for adaptive new mutation and standing variation is shown.

Standing variation allele frequency of the minor allele in Taiwan is plotted for both Taiwanese and Chinese populations.



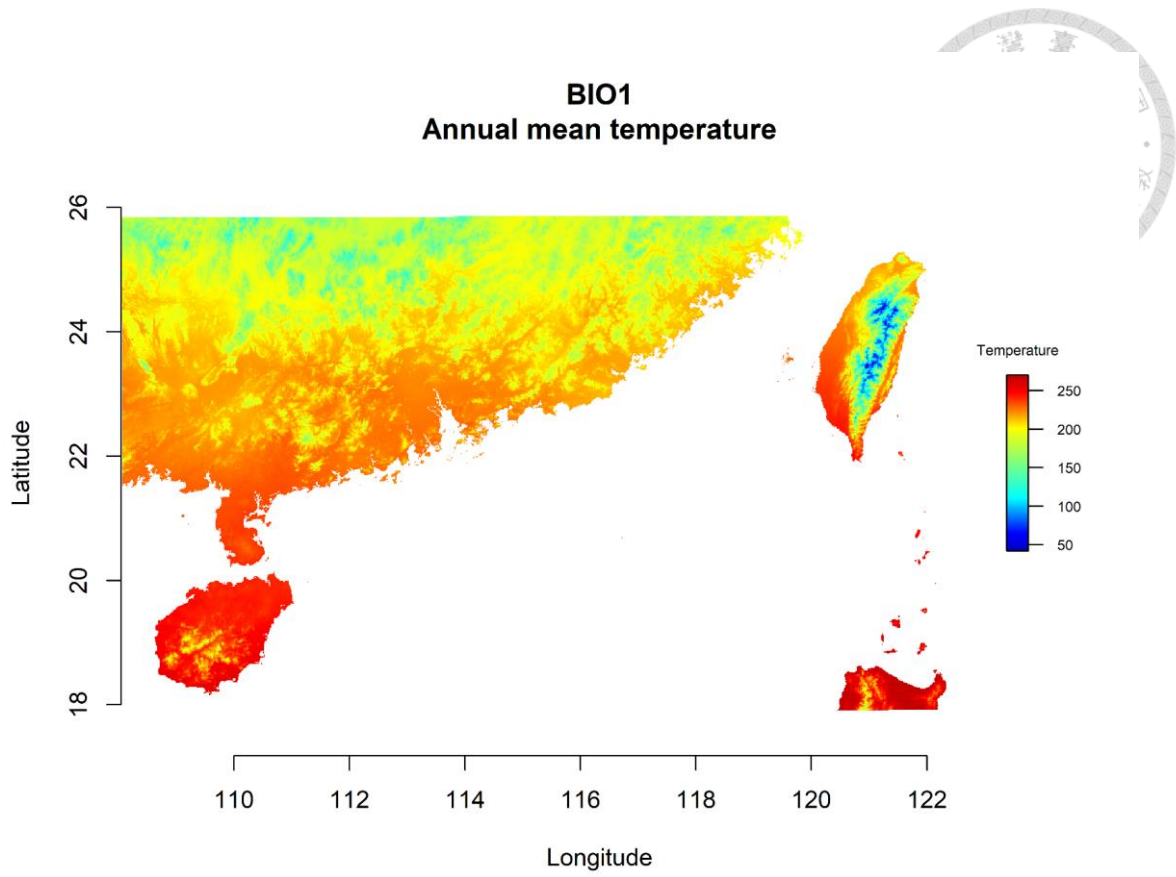
**Figure 10. The relationship between  $r^2$  importance and slope.**

$r^2$  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  $r^2$ . The value on x-axis shows the range of standardized environmental variables.



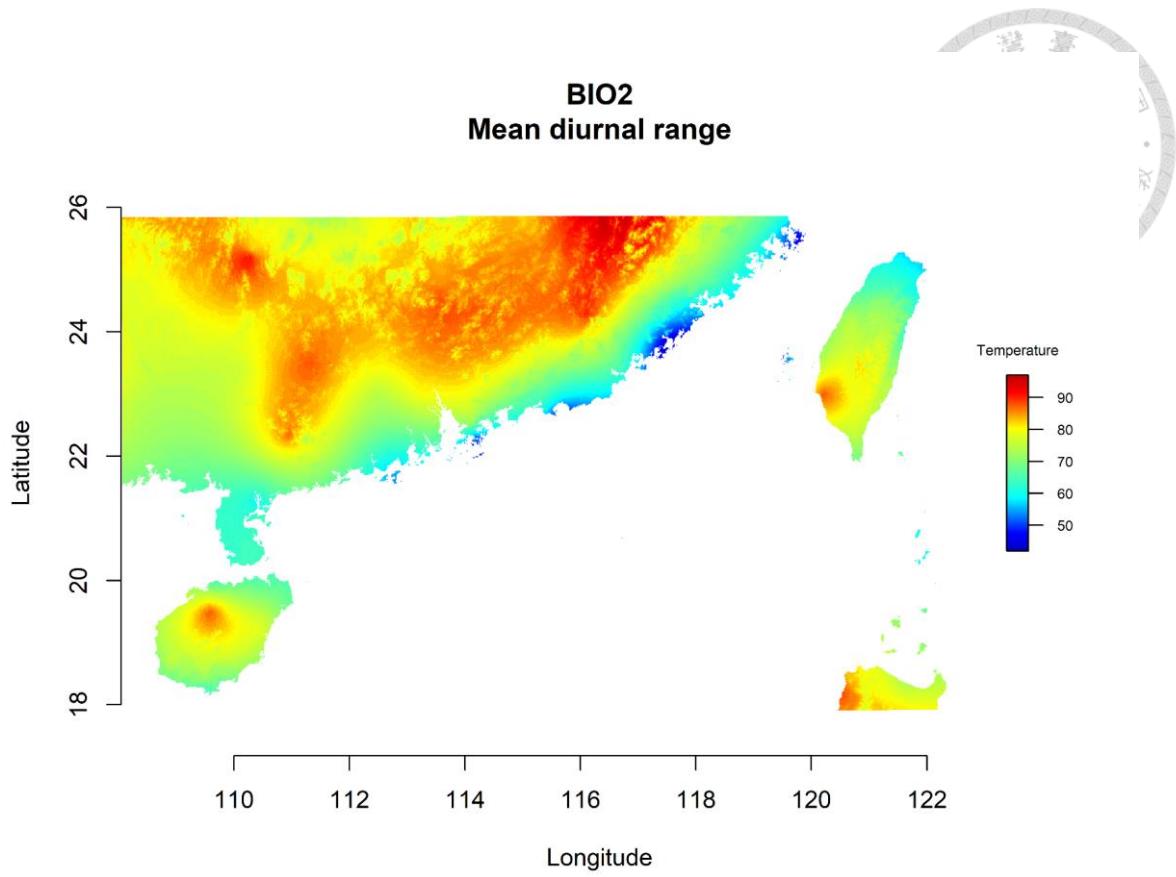
**Figure 11. Slope distribution.**

Slope is plotted for adaptive new mutation and standing variation. Statistical significance from Welch two sample  $t$ -test between new mutation and standing variation for each bioclimatic variable is shown (\*\*\*)  $p$ -value  $< 0.001$ ).



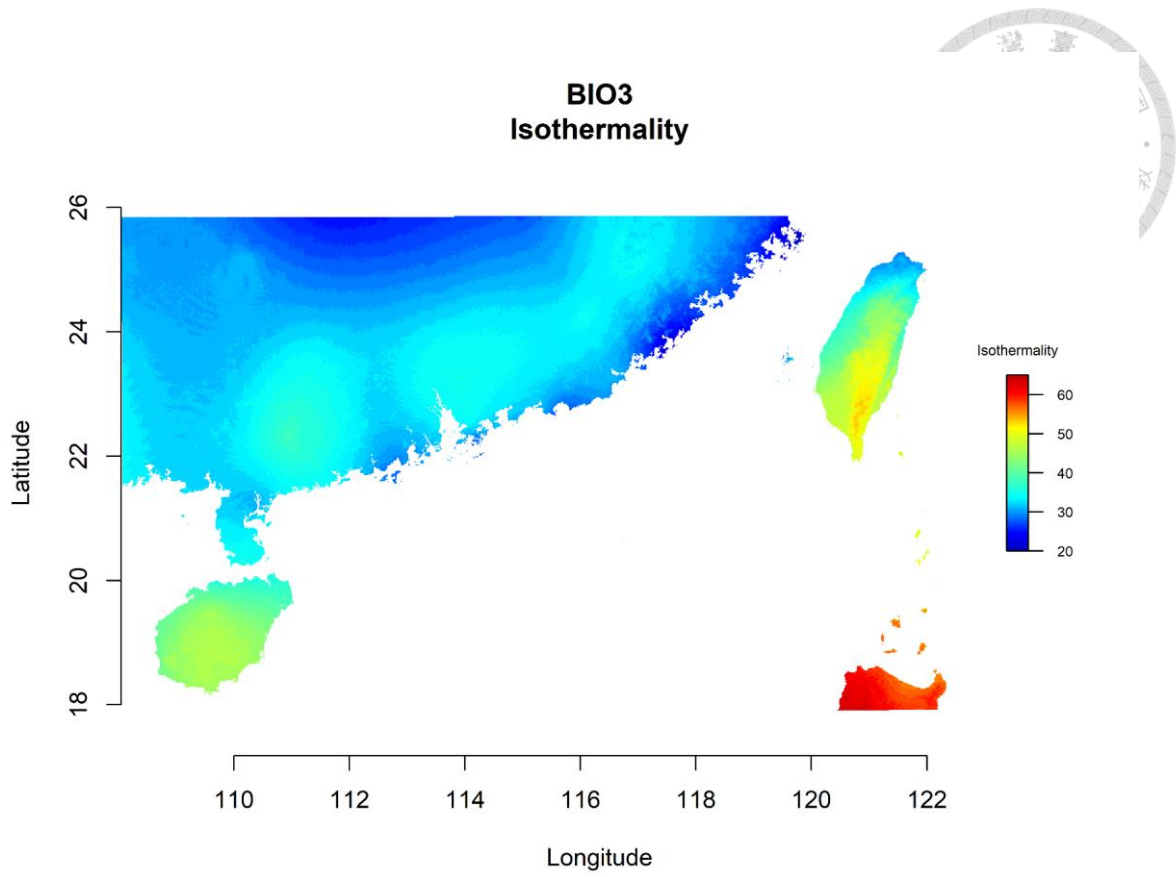
**Figure 12a. BIO1 environmental map.**

Current annual mean temperature (BIO1) is shown for Taiwan and its adjacent areas.



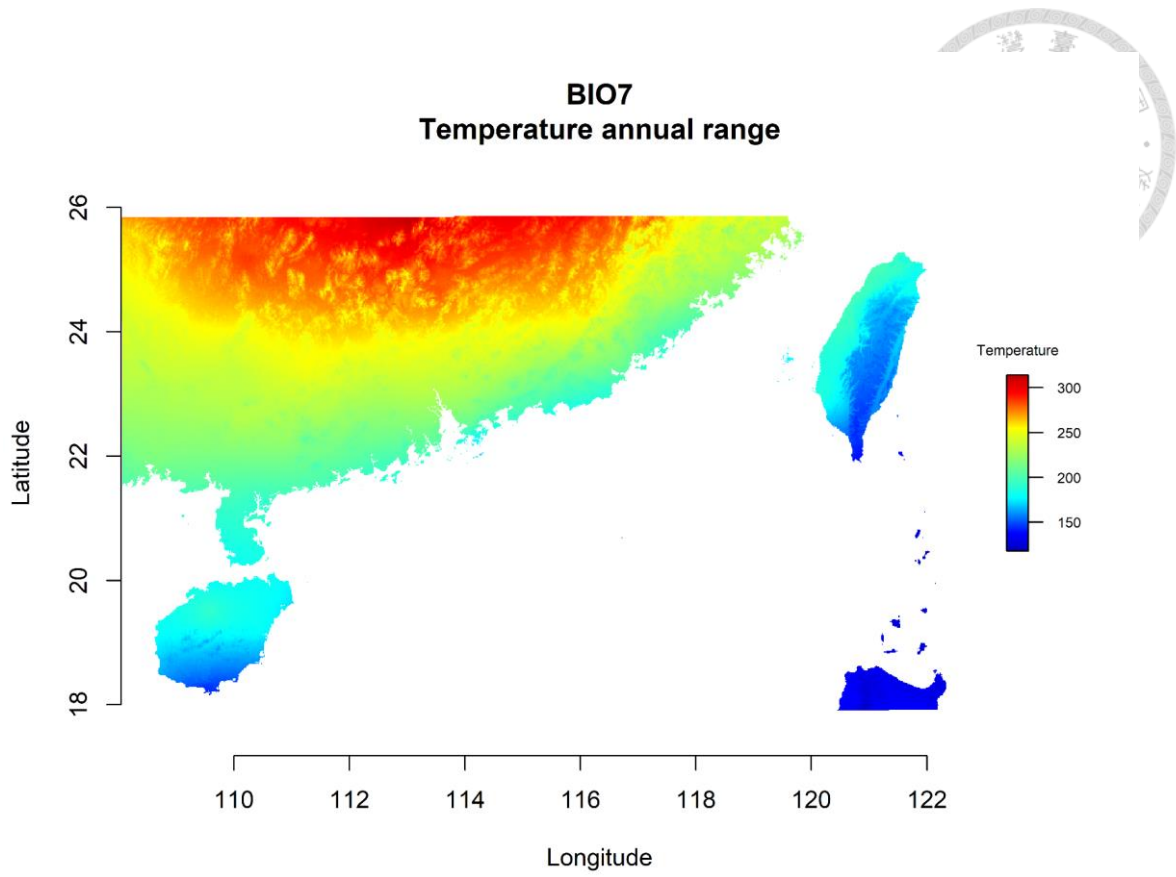
**Figure 12b. BIO2 environmental map.**

Current mean diurnal range (BIO2) is shown for Taiwan and its adjacent areas.



**Figure 12c. BIO3 environmental map.**

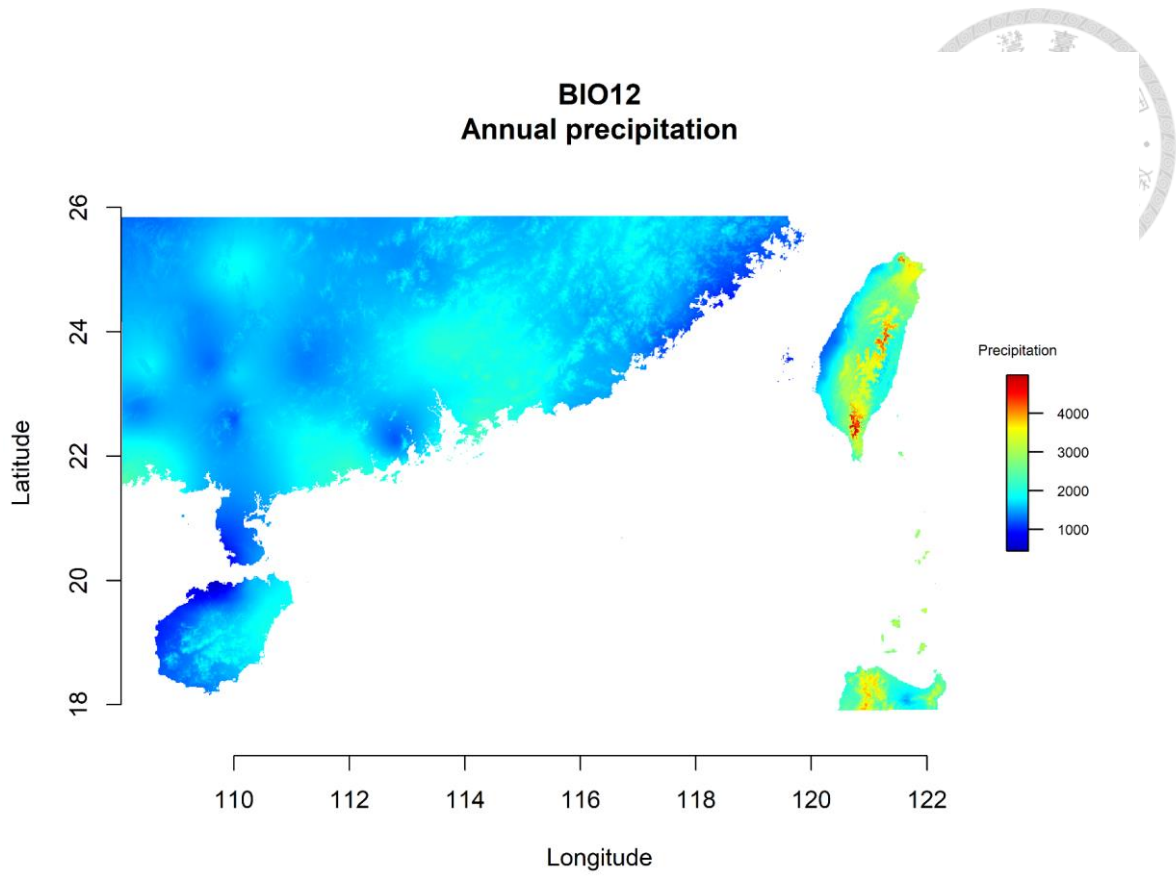
Current isothermality (BIO3) is shown for Taiwan and its adjacent areas.



**Figure 12d. BIO7 environmental map.**

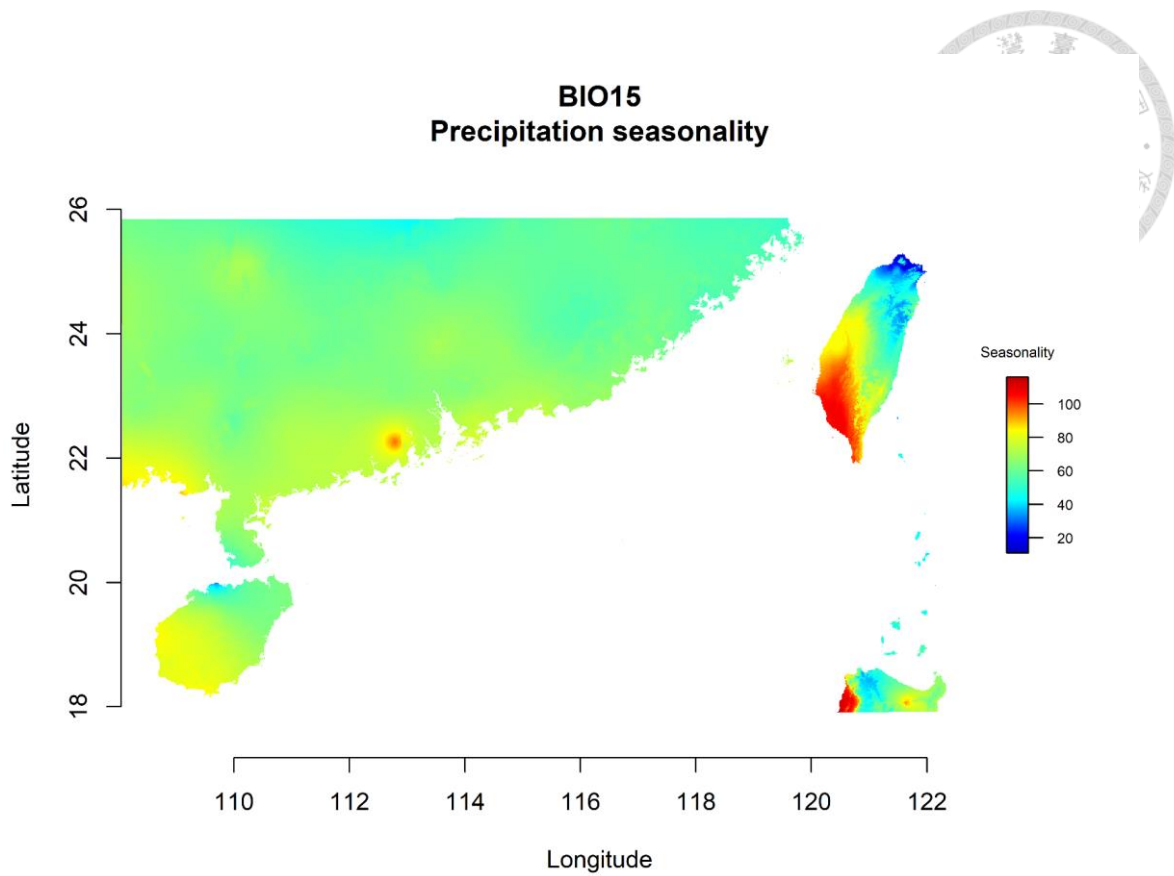
Current temperature annual range (BIO7) is shown for Taiwan and its adjacent areas.





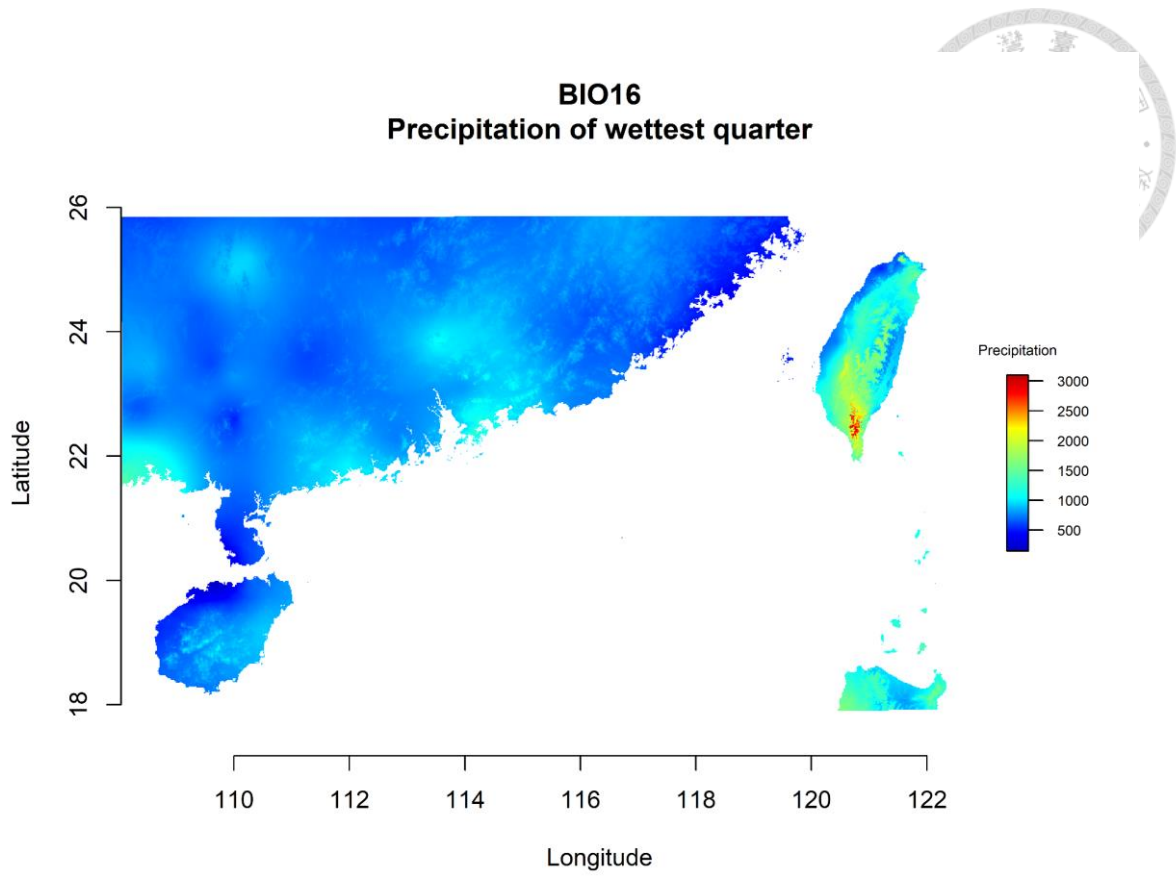
**Figure 12c. BIO12 environmental map.**

Current annual precipitation (BIO12) is shown for Taiwan and its adjacent areas.



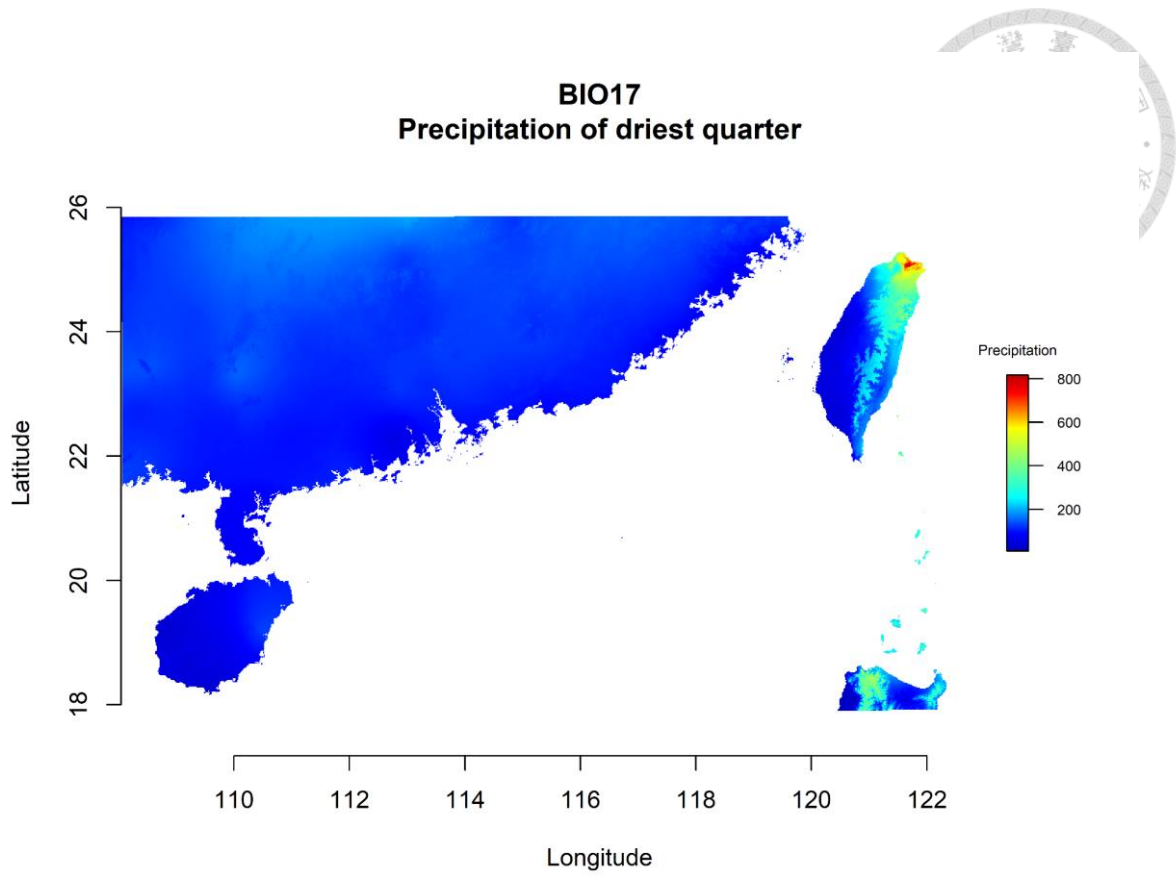
**Figure 12f. BIO15 environmental map.**

Current precipitation seasonality (BIO15) is shown for Taiwan and its adjacent areas.



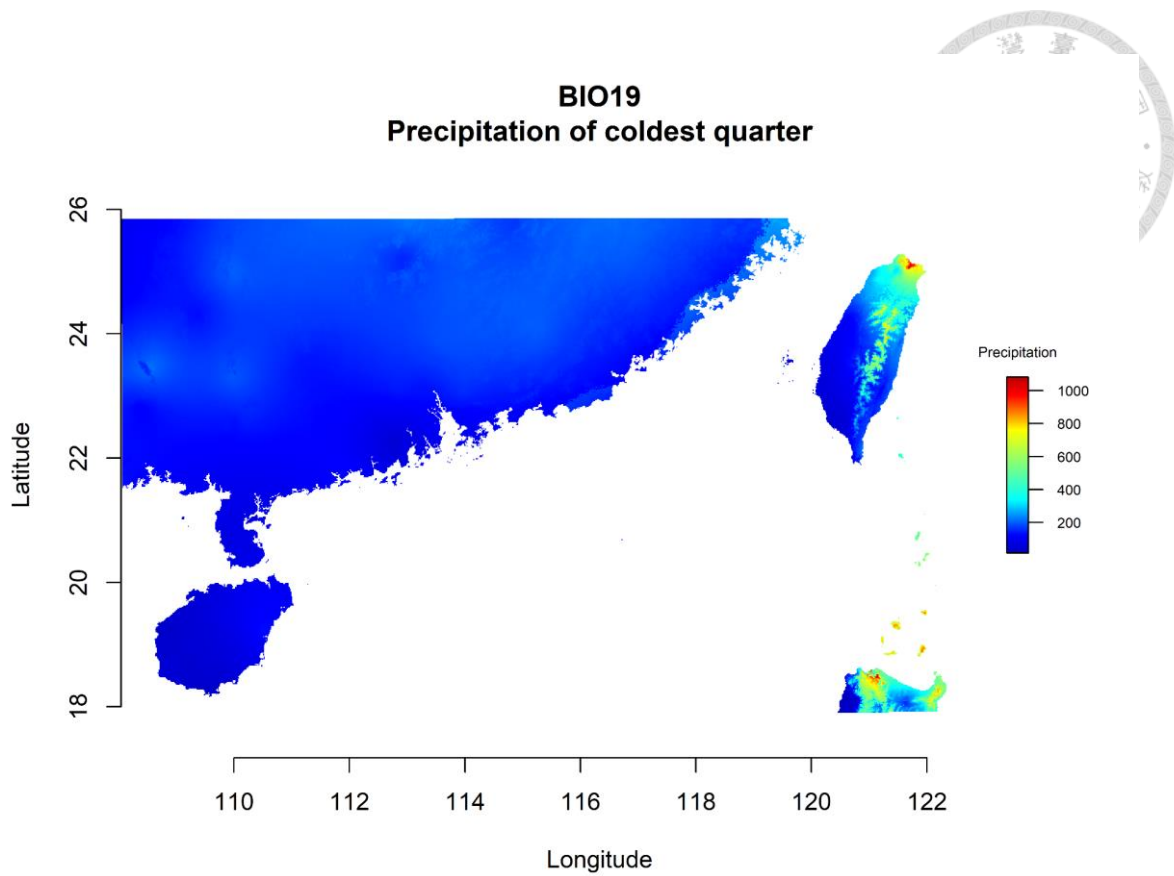
**Figure 12g. BIO16 environmental map.**

Current precipitation of wettest quarter (BIO16) is shown for Taiwan and its adjacent areas.



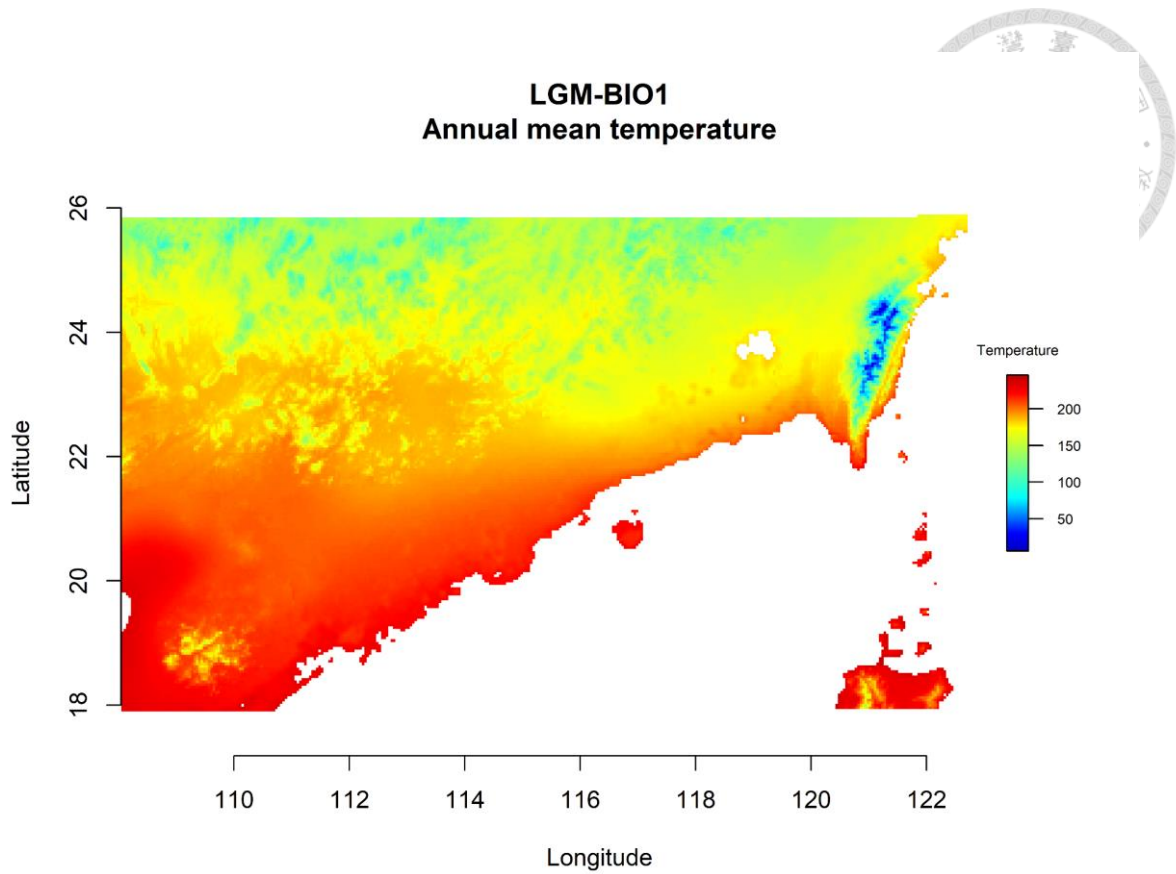
**Figure 12h. BIO17 environmental map.**

Current precipitation of driest quarter (BIO17) is shown for Taiwan and its adjacent areas.



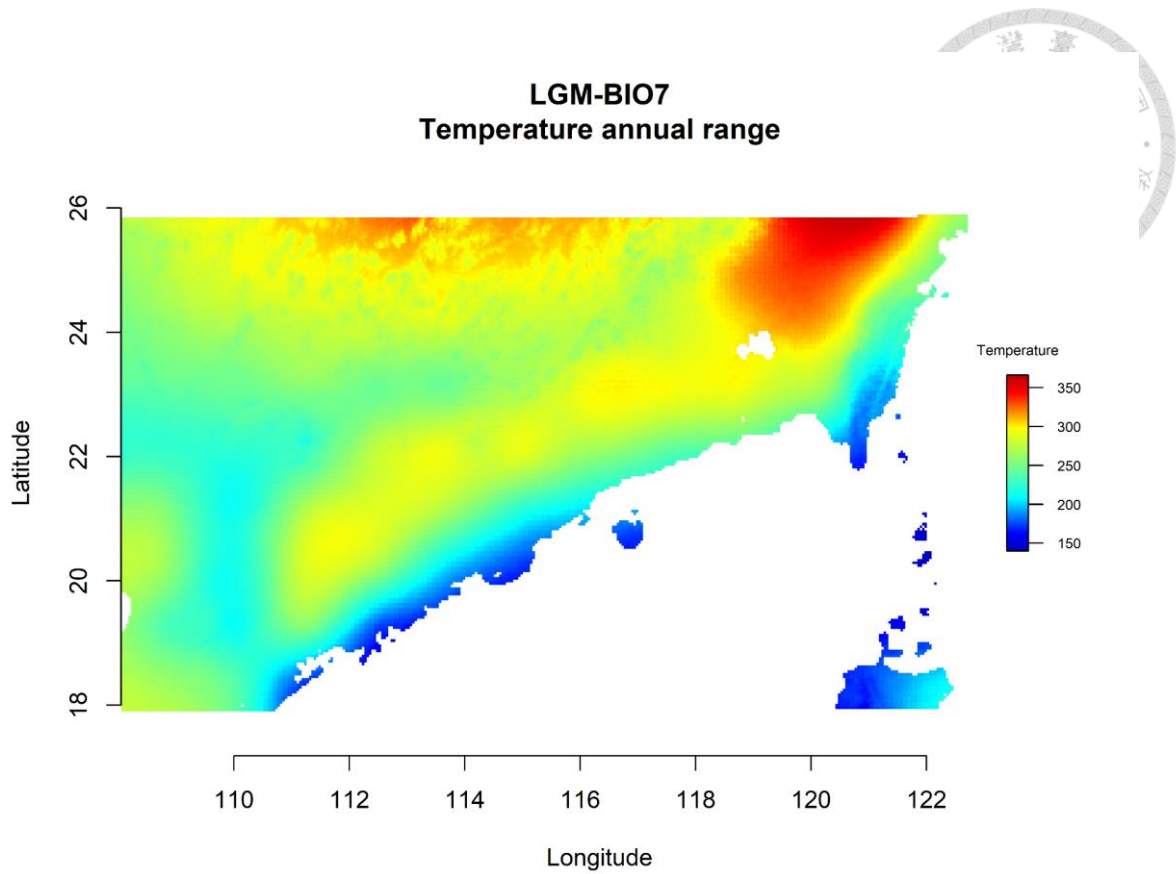
**Figure 12i. BIO19 environmental map.**

Current precipitation of coldest quarter (BIO19) is shown for Taiwan and its adjacent areas.



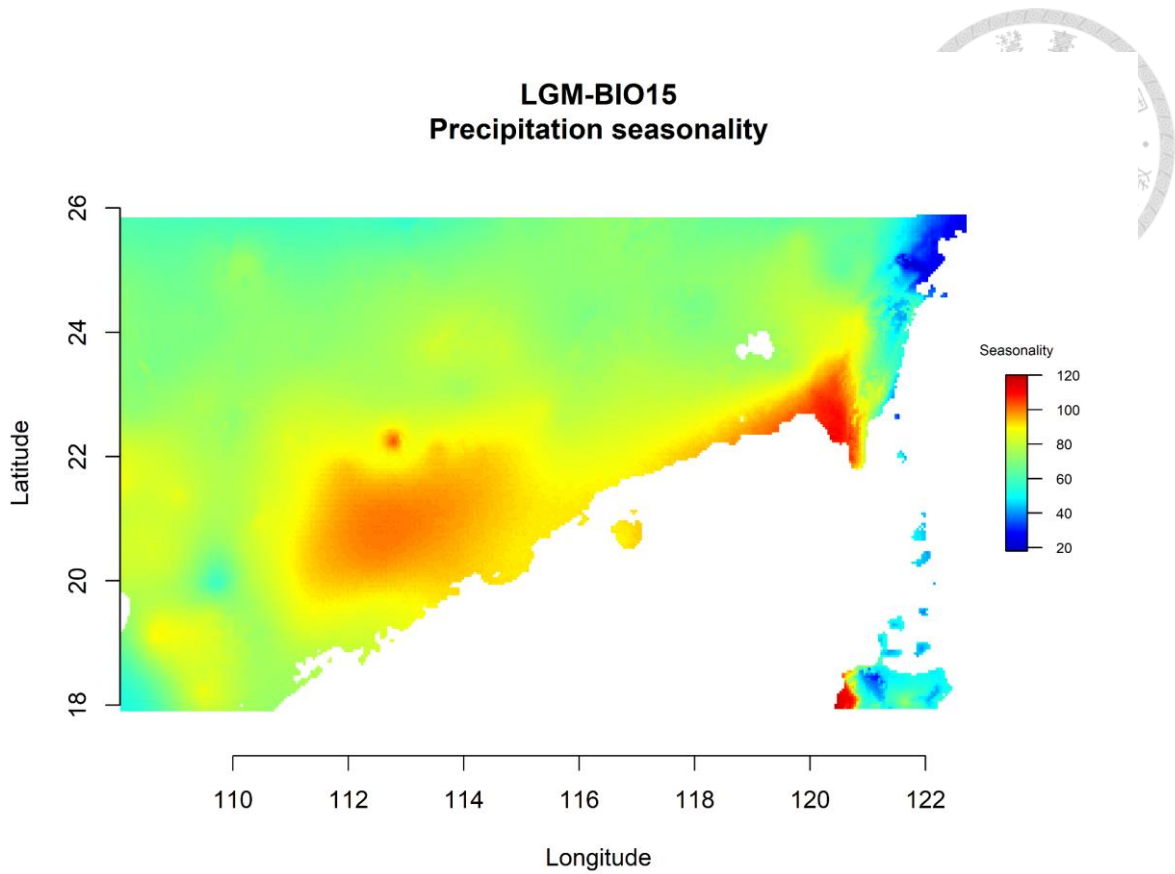
**Figure 13a. BIO1 environmental map during the last glacial maximum.**

Annual mean temperature (BIO1) is shown for Taiwan and its adjacent areas during the last glacial maximum. The climatic layer is downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) at spatial resolution of 2.5 arc-minute.



**Figure 13b. BIO7 environmental map during the last glacial maximum.**

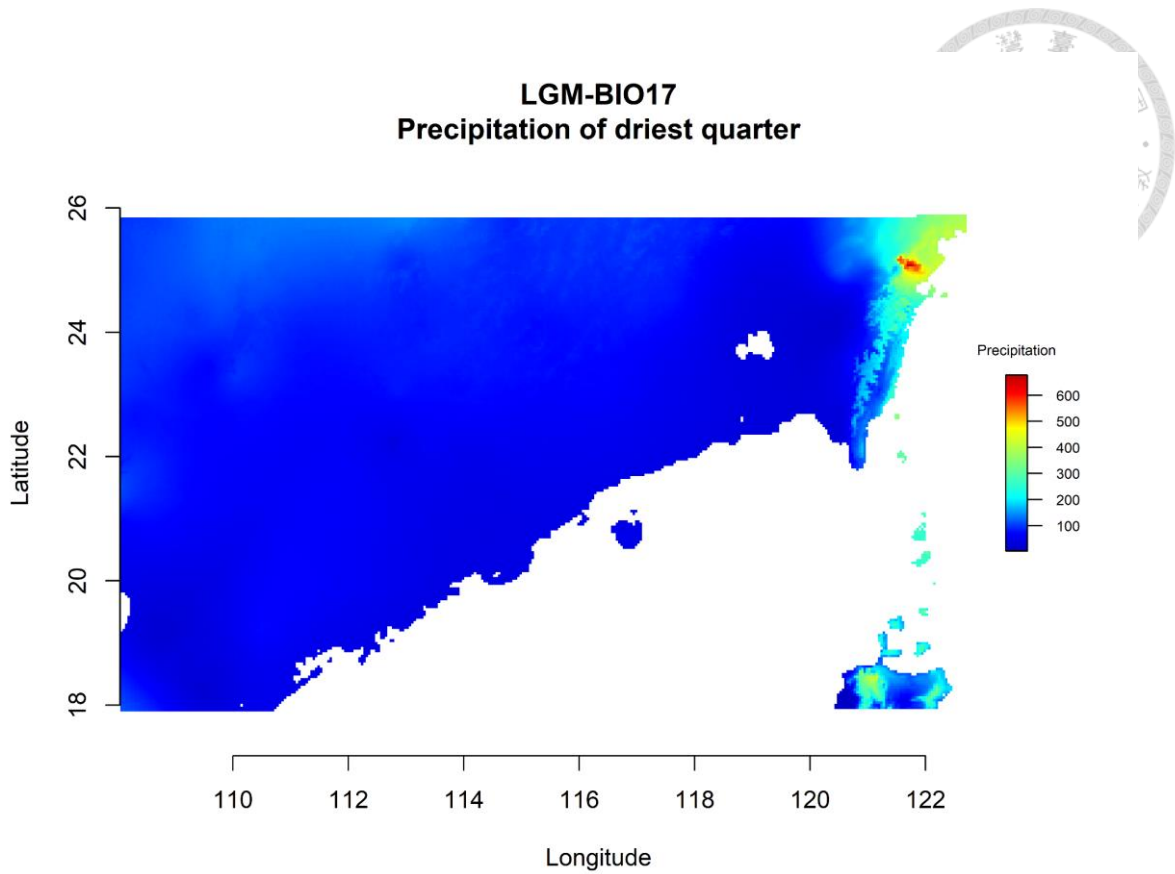
Temperature annual range (BIO7) is shown for Taiwan and its adjacent areas during the last glacial maximum. The climatic layer is downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) at spatial resolution of 2.5 arc-minute.



**Figure 13c. BIO15 environmental map during the last glacial maximum.**

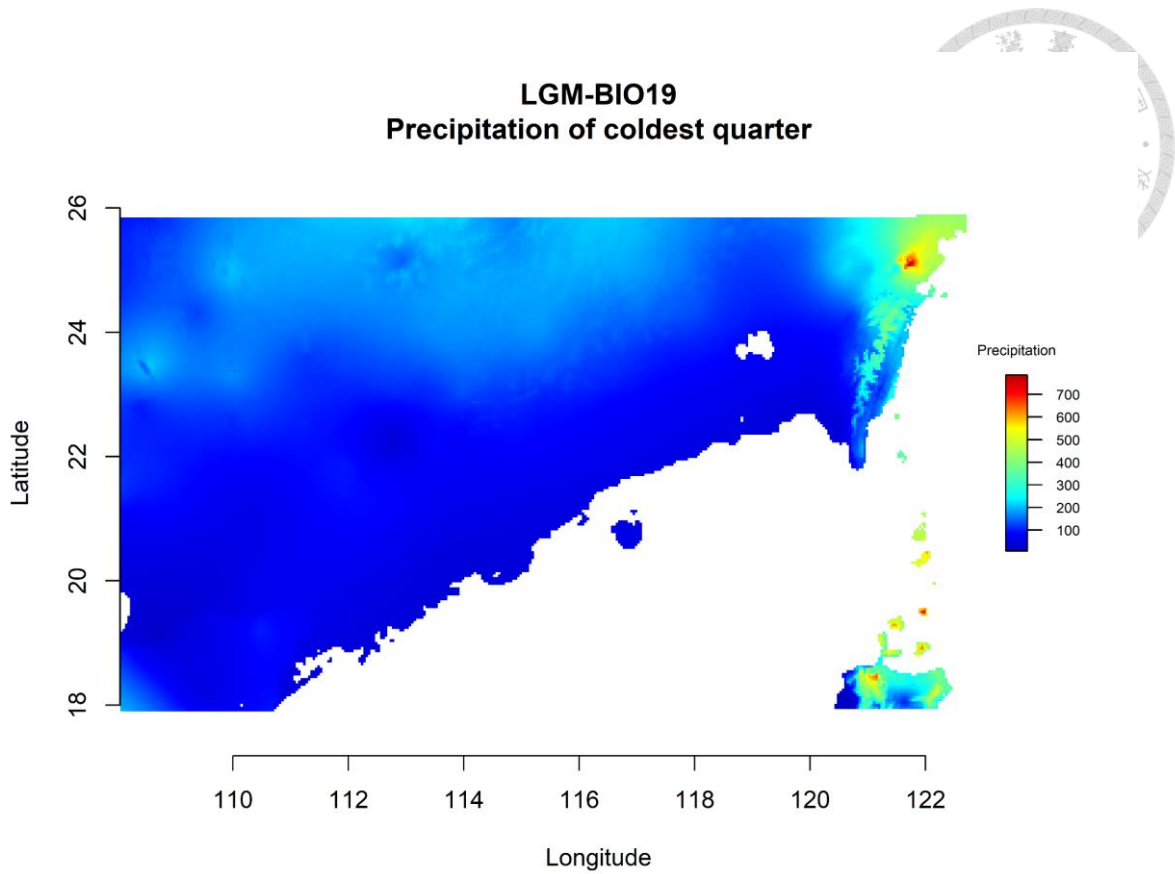
Precipitation seasonality (BIO15) is shown for Taiwan and its adjacent areas during the last glacial maximum. The climatic layer is downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) at spatial resolution of 2.5 arc-minute.





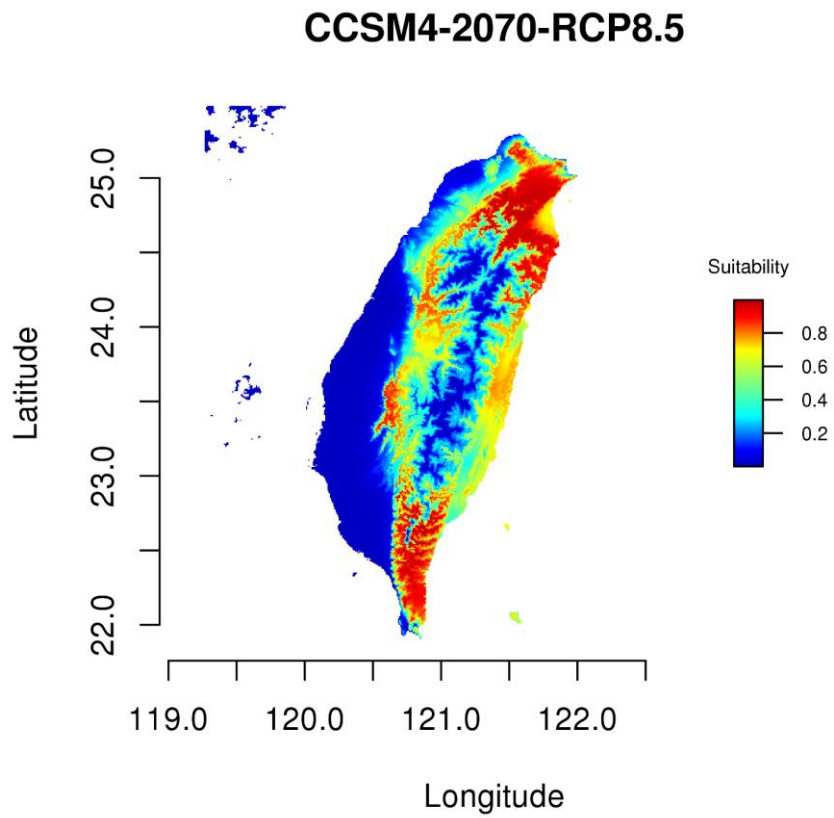
**Figure 13d. BIO17 environmental map during the last glacial maximum.**

Precipitation of driest quarter (BIO17) is shown for Taiwan and its adjacent areas during the last glacial maximum. The climatic layer is downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) at spatial resolution of 2.5 arc-minute.



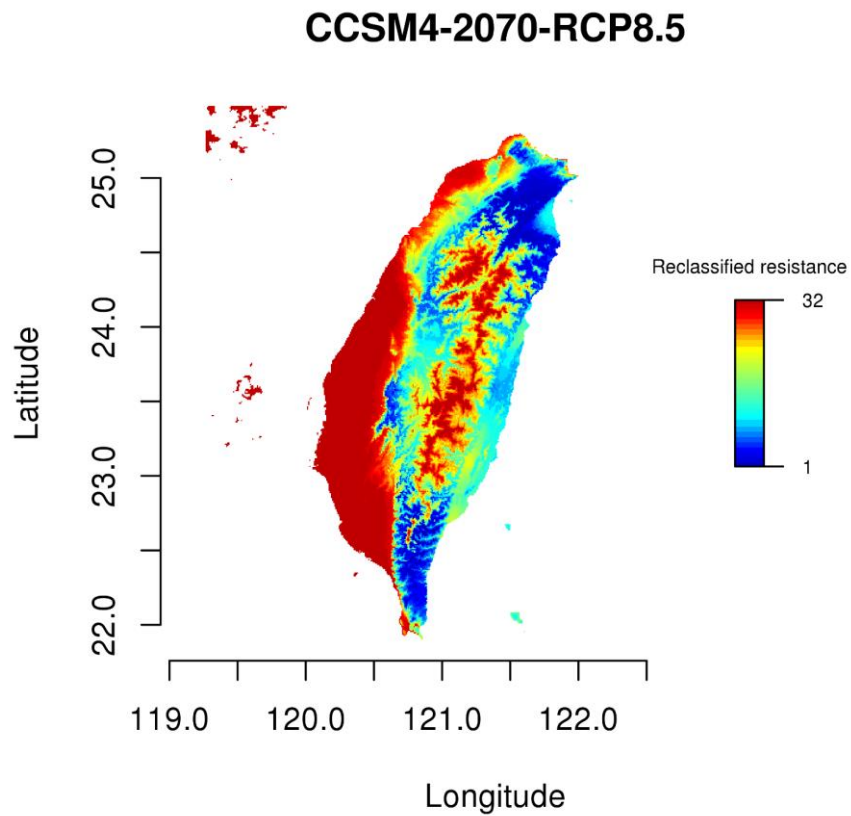
**Figure 13e. BIO19 environmental map during the last glacial maximum.**

Precipitation of coldest quarter (BIO19) is shown for Taiwan and its adjacent areas during the last glacial maximum. The climatic layer is downloaded from WorldClim database version 1.4 (<http://worldclim.org/>) at spatial resolution of 2.5 arc-minute.



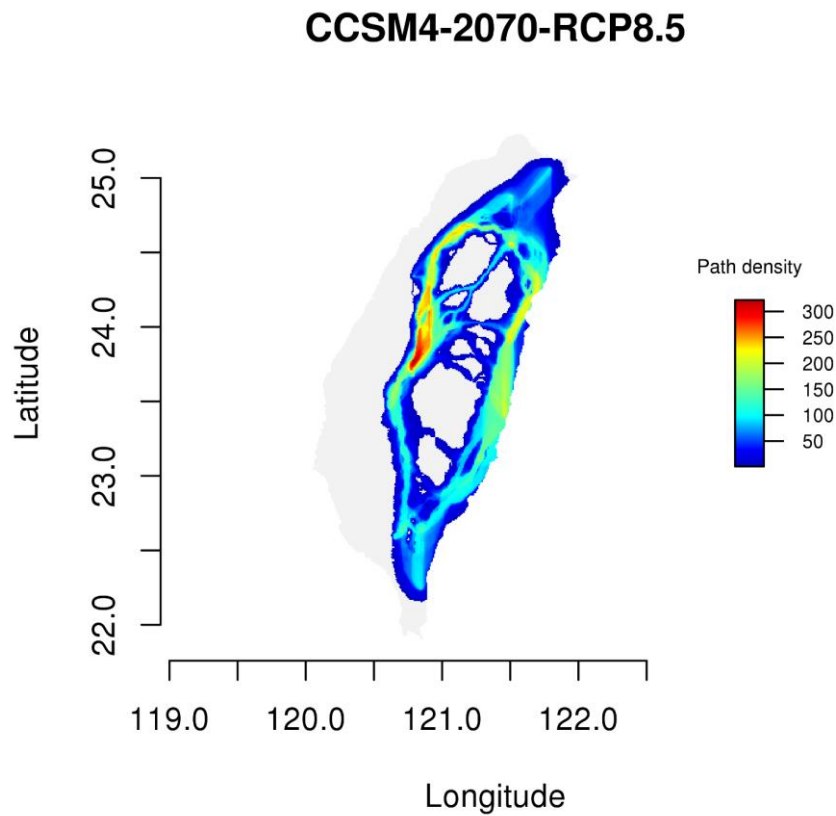
**Figure 14a. Expanding-extreme suitability of *Musa itinerans*.**

Suitability is provided by MaxEnt in cloglog output.



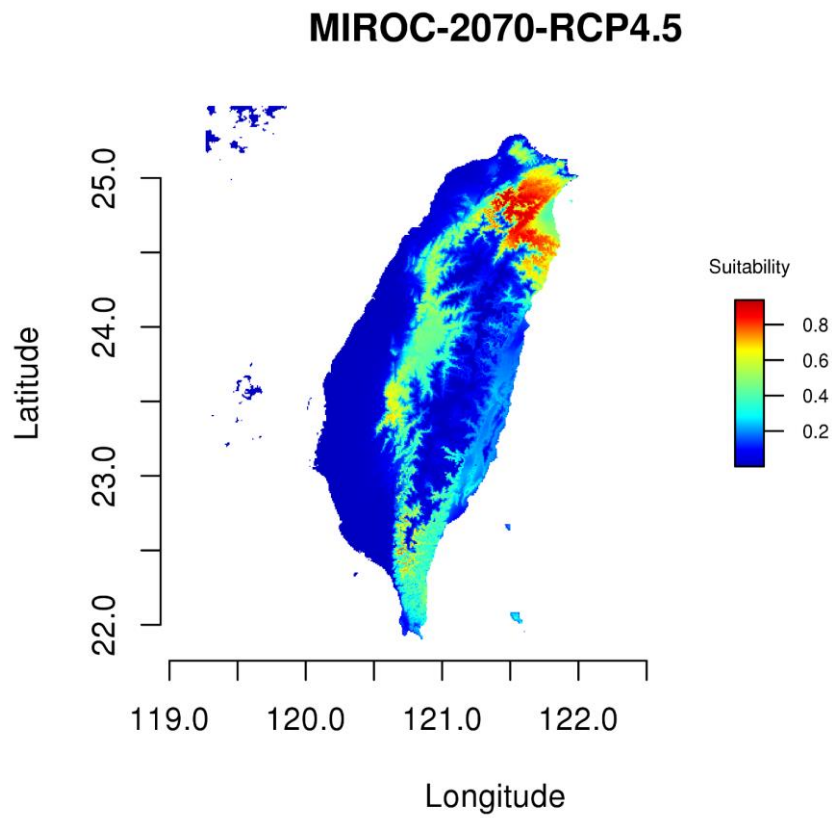
**Figure 14b. Expanding-extreme resistance of *Musa itinerans*.**

Resistance is the reciprocal of suitability from MaxEnt under the expanding extreme (CCSM4-2070-RCP8.5), and is then reclassified into 32 color blocks.



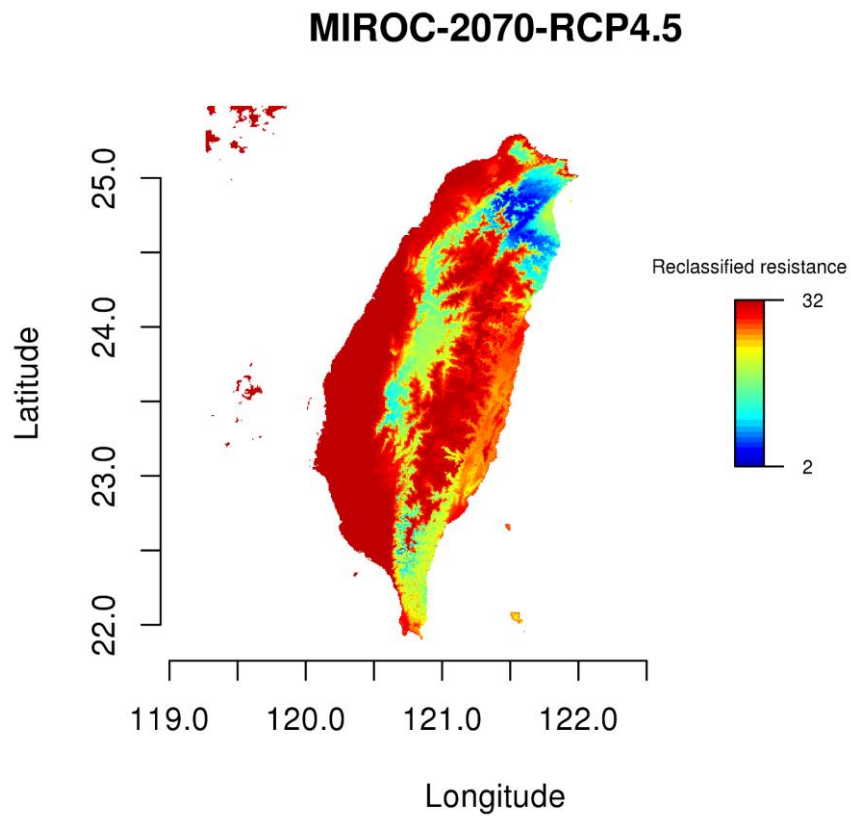
**Figure 14c. Least cost corridor landscape for the expanding extreme.**

Least cost corridor landscape integrates all least cost corridors which allow paths with 1%, 2%, or 5% higher cost than the least cost value. Path density is colored according to the amount of overlapping paths.



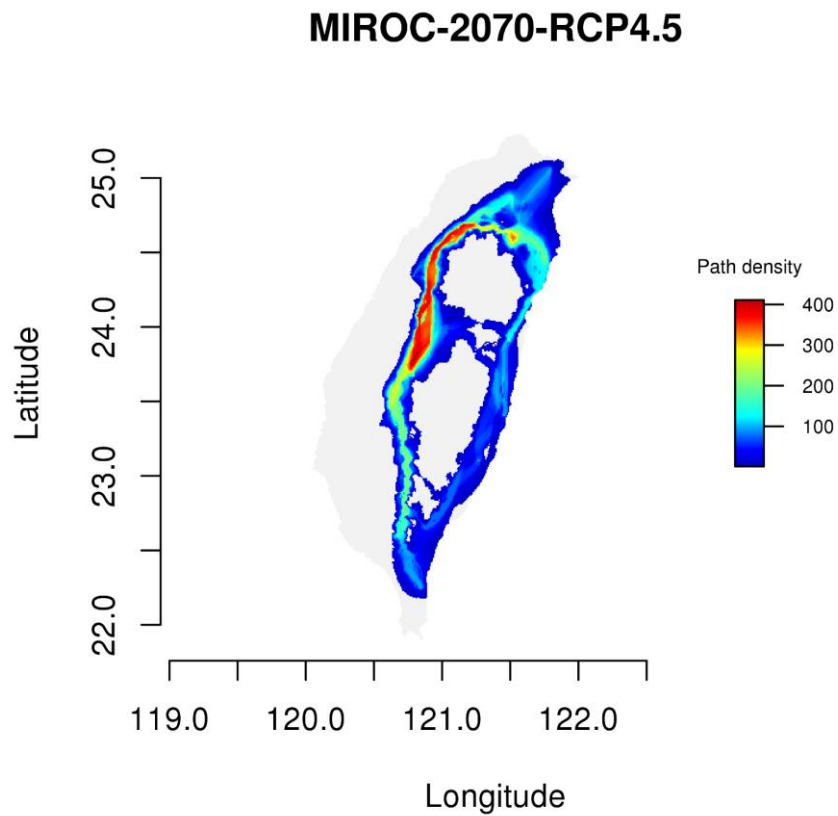
**Figure 15a. Contracting-extreme suitability of *Musa itinerans*.**

Suitability is provided by MaxEnt in cloglog output.



**Figure 15b. Contracting-extreme resistance of *Musa itinerans*.**

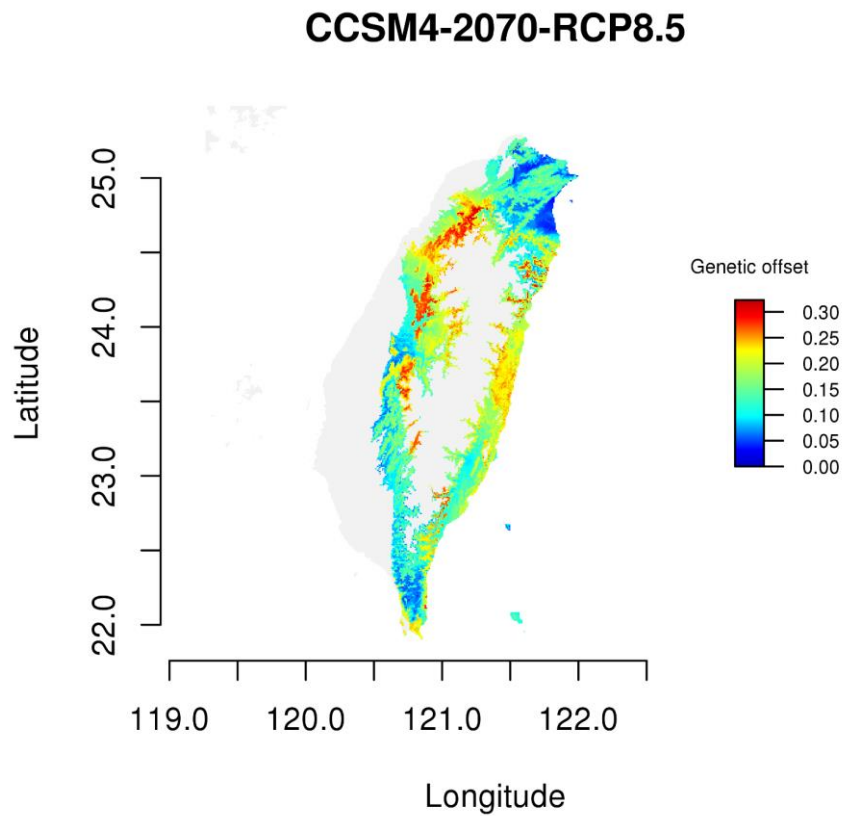
Resistance is the reciprocal of suitability from MaxEnt under the contracting extreme (MIROC-2070-RCP4.5), and is then reclassified into 32 color blocks.



**Figure 15c. Least cost corridor landscape for the contracting extreme.**

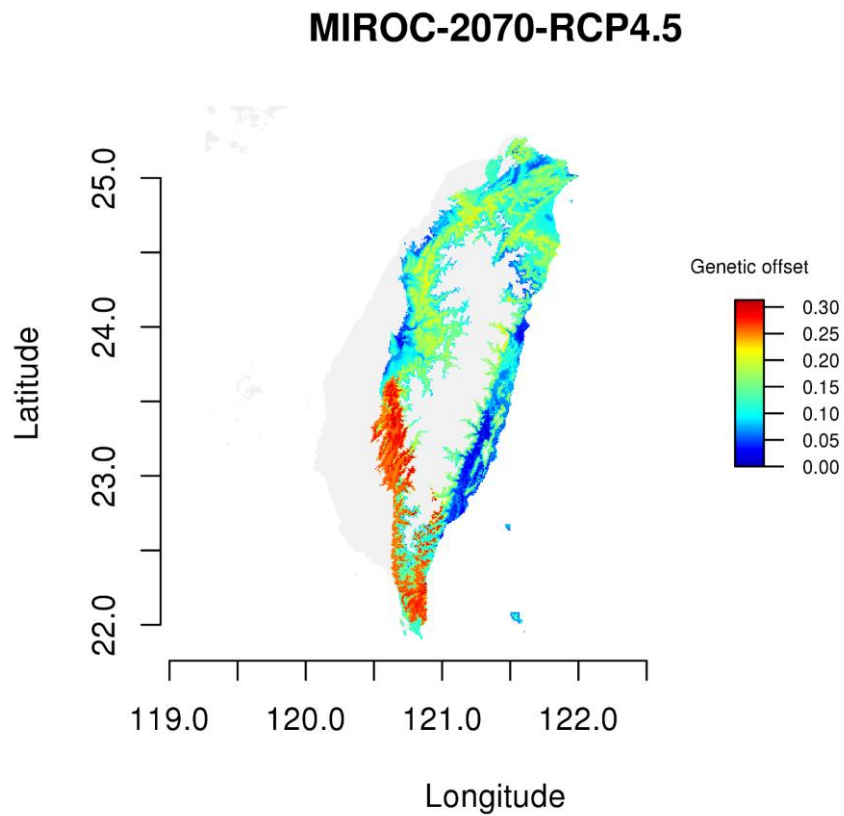
Least cost corridor landscape integrates all least cost corridors which allow paths with 1%, 2%, or 5% higher cost than the least cost value. Path density is colored according to the amount of overlapping paths.





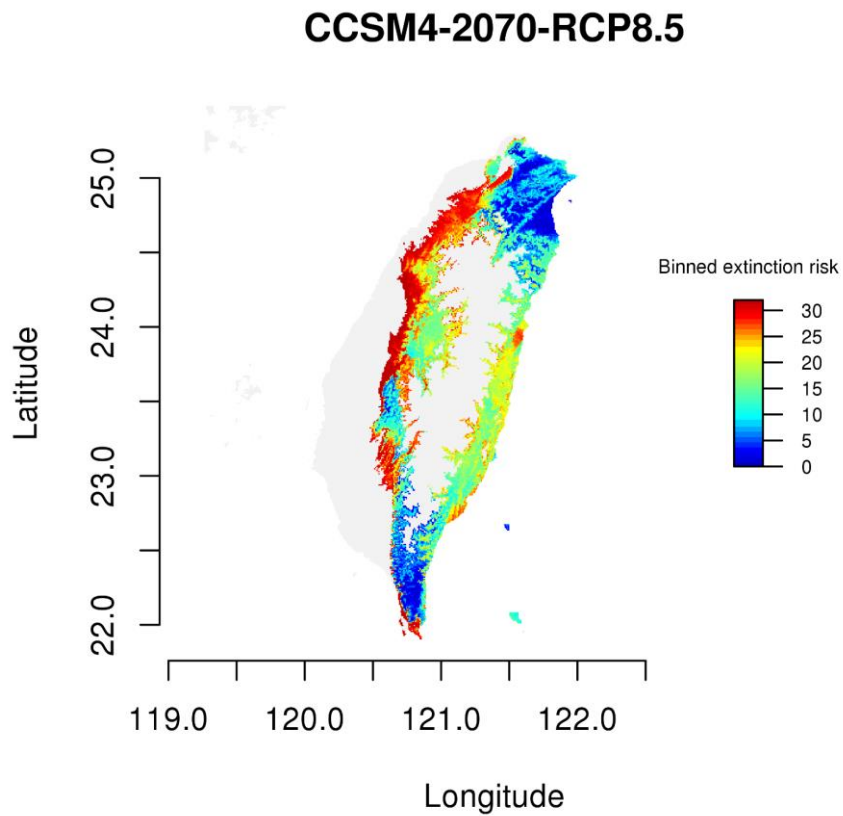
**Figure 16a. Genetic offset for the expanding extreme.**

Overall genetic offset of nine bioclimatic variables under the expanding extreme (CCSM4-2070-RCP8.5) is shown. Grids with current suitability  $< 0.2$  are excluded.



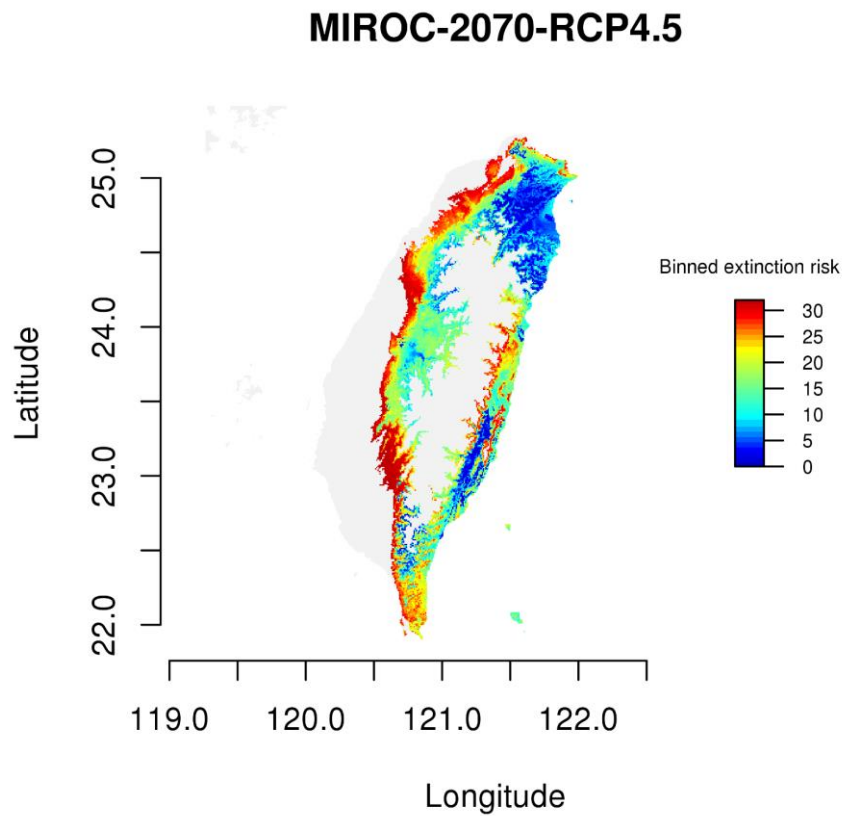
**Figure 16b. Genetic offset for the contracting extreme.**

Overall genetic offset of nine bioclimatic variables under the contracting extreme (MIROC-2070-RCP4.5) is shown. Grids with current suitability  $< 0.2$  are excluded.



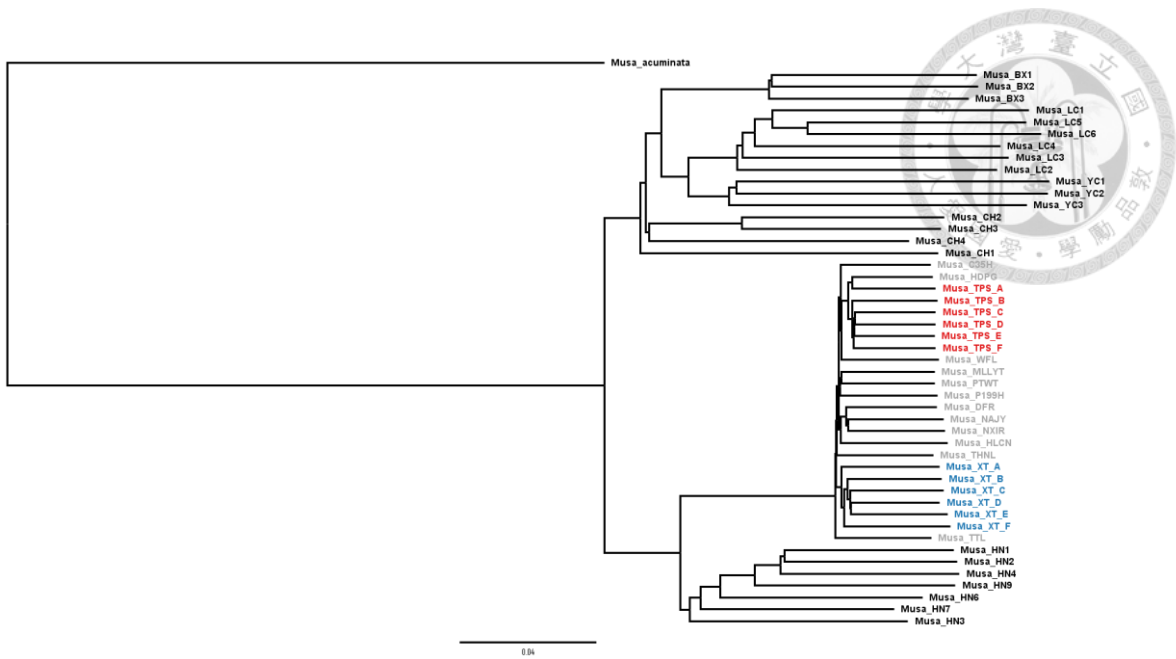
**Figure 17a. Extinction risk for the expanding extreme.**

Extinction risk under the expanding extreme (CCSM4-2070-RCP8.5) is partitioned into 32 bins with equal amount of data points, and then reclassified into 32 color blocks. Grids with current suitability  $< 0.2$  are excluded.



**Figure 17b. Extinction risk for the contracting extreme.**

Extinction risk under the contracting extreme (MIROC-2070-RCP4.5) is partitioned into 32 bins with equal amount of data points, and then reclassified into 32 color blocks. Grids with current suitability  $< 0.2$  are excluded.



**Figure 18. Phylogeny of Taiwanese and Chinese *Musa itinerans*.**

The gray-colored indicates Taiwanese lowland populations; the blue-colored indicates Taiwanese populations of Xitou transect; the red-colored indicates Taiwanese populations of Taipingshan transect; the black-colored indicates Chinese accessions and one outgroup *Musa acuminata*.

## Tables



**Table 1. *Musa itinerans* collection information.**

Site	Code	Latitude	Longitude	Altitude	Used for SSR	Used for SNP
TPS300	CLM0031	24.60428	121.5336	331	v	v
TPS300	CLM0032	24.60427	121.5344	334	v	v
TPS300	CLM0033	24.60436	121.5348	332	v	v
TPS300	CLM0034	24.60446	121.5351	329	v	v
TPS300	CLM0035	24.60459	121.5355	324	v	
TPS300	CLM0036	24.60463	121.5356	323	v	v
TPS300	CLM0037	24.60472	121.5357	320	v	v
TPS300	CLM0038	24.60475	121.536	323	v	v
TPS300	CLM0039	24.60499	121.5366	318	v	v
TPS300	CLM0040	24.60514	121.5373	318	v	v
TPS300	CLM0041	24.60526	121.5374	312	v	v
TPS400	CLM0042	24.58588	121.5148	394		v
TPS400	CLM0043	24.58469	121.5138	400		v
TPS400	CLM0044	24.58422	121.5132	390		
TPS400	CLM0045	24.58401	121.5131	397		v
TPS400	CLM0046	24.58374	121.5129	397		v
TPS400	CLM0047	24.58356	121.5129	405		
TPS400	CLM0048	24.58336	121.5128	411		v
TPS400	CLM0049	24.58227	121.5124	411		v
TPS400	CLM0050	24.58163	121.5112	397		v
TPS400	CLM0051	24.5814	121.511	404		v
TPS400	CLM0052	24.58085	121.5106	416		v
TPS400	CLM0053	24.5802	121.5102	448		v
TPS500	CLM0086	24.57036	121.4993	429	v	v
TPS500	CLM0087	24.57036	121.4993	429	v	v
TPS500	CLM0088	24.56993	121.4996	443	v	v
TPS500	CLM0089	24.5691	121.4996	441	v	v
TPS500	CLM0090	24.56884	121.4995	439	v	v
TPS500	CLM0091	24.56832	121.4994	441	v	v
TPS500	CLM0092	24.56705	121.4993	449	v	v
TPS500	CLM0093	24.56662	121.4992	447	v	v
TPS500	CLM0094	24.56639	121.499	441	v	v
TPS500	CLM0095	24.56558	121.4989	443	v	v
TPS600	CLM0054	24.55402	121.5079	596		v

TPS600	CLM0055	24.55415	121.5082	610		
TPS600	CLM0056	24.55403	121.5089	619		
TPS600	CLM0057	24.55385	121.5091	616		
TPS600	CLM0058	24.55383	121.5091	614		
TPS600	CLM0059	24.553	121.5095	613		
TPS600	CLM0060	24.55218	121.5093	600		v
TPS600	CLM0061	24.55218	121.5093	598		v
TPS600	CLM0062	24.55209	121.5093	603		v
TPS600	CLM0063	24.55144	121.5091	605		v
TPS600	CLM0064	24.55002	121.5083	576		v
TPS700	CLM0065	24.54913	121.5116	698	v	v
TPS700	CLM0066	24.54861	121.5119	701	v	v
TPS700	CLM0067	24.54733	121.5124	725	v	v
TPS700	CLM0068	24.54612	121.5128	737	v	
TPS700	CLM0069	24.54621	121.5126	723	v	v
TPS700	CLM0070	24.54486	121.5126	745	v	v
TPS700	CLM0071	24.54485	121.5126	744	v	v
TPS700	CLM0072	24.54467	121.5125	749	v	v
TPS700	CLM0073	24.54452	121.512	740	v	v
TPS700	CLM0074	24.5444	121.5117	739	v	v
TPS700	CLM0075	24.54408	121.5113	738	v	v
TPS900	CLM0076	24.53878	121.5219	987	v	v
TPS900	CLM0077	24.5399	121.5126	917	v	v
TPS900	CLM0078	24.53934	121.5122	912	v	v
TPS900	CLM0079	24.53912	121.512	896	v	v
TPS900	CLM0080	24.53962	121.5119	890	v	v
TPS900	CLM0081	24.54018	121.5123	900	v	v
TPS900	CLM0082	24.54005	121.5122	901	v	v
TPS900	CLM0083	24.53983	121.5121	898	v	v
TPS900	CLM0084	24.53953	121.5119	891	v	v
TPS900	CLM0085	24.53922	121.5119	895	v	v
XT400	CLM0121	23.7413	120.7367	407	v	v
XT400	CLM0122	23.74138	120.7367	405	v	v
XT400	CLM0123	23.74167	120.7367	394	v	v
XT400	CLM0124	23.7425	120.7396	356	v	v
XT400	CLM0125	23.74263	120.7394	363	v	v
XT400	CLM0126	23.74267	120.7393	364	v	v
XT400	CLM0127	23.74279	120.7391	365	v	v



XT400	CLM0128	23.74288	120.739	365	v	
XT400	CLM0129	23.74294	120.7388	367	v	
XT400	CLM0130	23.74293	120.7379	378	v	
XT400	CLM0131	23.74281	120.7376	378	v	
XT400	CLM0132	23.74211	120.7368	384	v	
XT500	CLM0133	23.72787	120.7618	543		v
XT500	CLM0134	23.72796	120.7621	538		v
XT500	CLM0135	23.72797	120.7622	537		v
XT500	CLM0136	23.72816	120.7623	527		v
XT500	CLM0137	23.72826	120.7627	521		v
XT500	CLM0138	23.72807	120.7634	523		v
XT500	CLM0139	23.72796	120.7637	525		v
XT500	CLM0140	23.72799	120.7643	516		v
XT500	CLM0141	23.72785	120.7644	521		v
XT500	CLM0142	23.72775	120.7652	520		
XT500	CLM0143	23.72776	120.7655	516		v
XT700	CLM0144	23.71015	120.7792	681	v	v
XT700	CLM0145	23.71089	120.7789	669	v	v
XT700	CLM0146	23.71158	120.7786	659	v	
XT700	CLM0147	23.71152	120.7781	658	v	v
XT700	CLM0148	23.71119	120.7779	668	v	
XT700	CLM0149	23.71077	120.7777	681	v	
XT700	CLM0150	23.71044	120.7775	689	v	v
XT700	CLM0151	23.71032	120.7774	692	v	v
XT700	CLM0152	23.71002	120.7773	698	v	
XT700	CLM0153	23.70981	120.7771	702	v	v
XT700	CLM0154	23.70994	120.7773	699	v	v
XT700	CLM0255	23.70981	120.7767	710		v
XT700	CLM0256	23.7099	120.7768	708		v
XT700	CLM0257	23.71013	120.7766	709		v
XT900	CLM0155	23.69081	120.7853	857		v
XT900	CLM0156	23.69314	120.786	873		
XT900	CLM0157	23.69323	120.786	874		v
XT900	CLM0158	23.69354	120.7859	870		v
XT900	CLM0159	23.69366	120.7859	869		v
XT900	CLM0160	23.69405	120.786	872		v
XT900	CLM0161	23.69422	120.7859	872		v
XT900	CLM0162	23.69435	120.7859	871		v





XT900	CLM0163	23.6949	120.7858	878		
XT900	CLM0258	23.6949	120.7858	878		
XT900	CLM0164	23.69512	120.7858	882		
XT900	CLM0259	23.69275	120.7858	865		
XT900	CLM0260	23.6935	120.786	874		
XT1200	CLM0165	23.67361	120.7866	1150	v	v
XT1200	CLM0166	23.67358	120.786	1149	v	v
XT1200	CLM0167	23.67349	120.7855	1152	v	
XT1200	CLM0168	23.67338	120.7855	1158	v	
XT1200	CLM0169	23.67337	120.7854	1157	v	v
XT1200	CLM0170	23.67313	120.7851	1163	v	v
XT1200	CLM0171	23.67279	120.7849	1175	v	v
XT1200	CLM0172	23.67266	120.7846	1175	v	v
XT1200	CLM0173	23.67251	120.7844	1179	v	v
XT1200	CLM0174	23.6725	120.7843	1177	v	v
XT1200	CLM0175	23.6724	120.7837	1170	v	v
XT1200	CLM0176	23.67226	120.7834	1176	v	v
XT1500	CLM0177	23.66741	120.7717	1497	v	v
XT1500	CLM0178	23.66883	120.7715	1481	v	v
XT1500	CLM0179	23.66951	120.771	1491	v	v
XT1500	CLM0180	23.6714	120.7726	1453	v	v
XT1500	CLM0181	23.67186	120.7728	1450	v	v
XT1500	CLM0182	23.67161	120.7729	1446	v	v
XT1500	CLM0183	23.67278	120.7729	1444	v	v
XT1500	CLM0184	23.67284	120.7731	1441	v	v
XT1500	CLM0185	23.67229	120.7734	1432	v	v
XT1500	CLM0186	23.67239	120.7733	1440	v	v
C35H	CLM0109	24.69913	121.1537	265	v	v
C35H	CLM0110	24.69826	121.1529	273	v	v
C35H	CLM0111	24.69789	121.1528	275	v	v
C35H	CLM0112	24.69782	121.1527	275	v	v
C35H	CLM0113	24.6978	121.1527	275	v	
C35H	CLM0114	24.69766	121.1526	275	v	v
C35H	CLM0115	24.69748	121.1525	275	v	
C35H	CLM0116	24.69743	121.1525	276	v	v
C35H	CLM0117	24.69688	121.1514	285	v	
C35H	CLM0118	24.69683	121.1511	289	v	v
C35H	CLM0119	24.69682	121.1509	289	v	v



C35H	CLM0120	24.69658	121.1506	290	v	v
C35H	CLM0240	24.69311	121.1451	315	v	v
WFL	CLM0188	25.06747	121.7998	332	v	v
WFL	CLM0189	25.06671	121.8001	324	v	v
WFL	CLM0190	25.06686	121.7993	334	v	v
WFL	CLM0191	25.06638	121.7985	333	v	v
WFL	CLM0192	25.0666	121.8002	322	v	v
WFL	CLM0193	25.07073	121.7995	355	v	v
WFL	CLM0194	25.07183	121.7987	365	v	v
WFL	CLM0195	25.07218	121.7985	366	v	v
WFL	CLM0196	25.07243	121.7977	379	v	v
WFL	CLM0197	25.07242	121.7974	384	v	v
WFL	CLM0250	25.07479	121.7955	430	v	v
WFL	CLM0251	25.07419	121.7949	427	v	v
WFL	CLM0252	25.07405	121.7958	412	v	v
WFL	CLM0253	25.07177	121.7991	363	v	v
THNL	CLM0213	24.08425	120.7876	318	v	v
THNL	CLM0214	24.08309	120.7981	344	v	v
THNL	CLM0215	24.08274	120.8001	363	v	v
THNL	CLM0216	24.08266	120.8002	367	v	v
THNL	CLM0217	24.08261	120.8006	367	v	v
THNL	CLM0218	24.08269	120.8008	361	v	v
THNL	CLM0219	24.08258	120.803	358	v	v
THNL	CLM0220	24.08212	120.8043	373	v	v
THNL	CLM0221	24.0813	120.8074	391	v	v
THNL	CLM0222	24.08134	120.8074	391	v	v
THNL	CLM0223	24.08167	120.8086	410	v	v
THNL	CLM0224	24.08164	120.809	413	v	v
THNL	CLM0225	24.0817	120.8079	395	v	v
THNL	CLM0226	24.07969	120.8079	435	v	v
THNL	CLM0227	24.08017	120.8076	417	v	v
PTWT	CLM0241	22.56724	120.6454	254	v	v
PTWT	CLM0242	22.56481	120.6465	350	v	v
PTWT	CLM0243	22.5657	120.6475	363	v	v
PTWT	CLM0244	22.56364	120.6456	378	v	v
PTWT	CLM0245	22.56276	120.6452	403	v	v
PTWT	CLM0246	22.56327	120.6443	401	v	v
PTWT	CLM0247	22.56226	120.6449	420	v	v



PTWT	CLM0248	22.57957	120.6505	485	v	v
PTWT	CLM0249	22.58491	120.6484	381	v	v
P199H	CLM0001	22.24463	120.835	470	v	v
P199H	CLM0004	22.2448	120.8353	470	v	v
P199H	CLM0005	22.24513	120.8367	452	v	v
P199H	CLM0008	22.24495	120.8369	457	v	v
P199H	CLM0009	22.24407	120.8394	455	v	v
P199H	CLM0010	22.24326	120.8508	450	v	v
P199H	CLM0012	22.23958	120.8575	450	v	v
P199H	CLM0014	22.23862	120.8575	437	v	v
P199H	CLM0016	22.23777	120.8584	439	v	v
P199H	CLM0017	22.23735	120.8599	432	v	v
MLLYT	CLM0096	24.3352	120.7832	409	v	v
MLLYT	CLM0097	24.33541	120.7844	417	v	v
MLLYT	CLM0098	24.33551	120.7847	412	v	v
MLLYT	CLM0099	24.33525	120.7853	471	v	v
MLLYT	CLM0100	24.33484	120.7858	440	v	v
MLLYT	CLM0101	24.335	120.7859	432	v	v
MLLYT	CLM0102	24.33525	120.7863	419	v	v
MLLYT	CLM0103	24.33532	120.7864	416	v	v
MLLYT	CLM0104	24.33541	120.7866	414	v	v
MLLYT	CLM0105	24.33584	120.7867	404	v	v
MLLYT	CLM0106	24.33591	120.787	409	v	v
MLLYT	CLM0107	24.33642	120.787	396	v	v
MLLYT	CLM0108	24.33632	120.7869	395	v	v
HDPG	CLM0201	24.88384	121.5023	221	v	v
HDPG	CLM0202	24.88383	121.5022	221	v	v
HDPG	CLM0203	24.8836	121.5021	223	v	v
HDPG	CLM0204	24.88337	121.5019	225	v	v
HDPG	CLM0205	24.88313	121.5016	229	v	v
HDPG	CLM0206	24.88312	121.5016	230	v	v
HDPG	CLM0207	24.88293	121.5013	234	v	v
HDPG	CLM0208	24.88292	121.5013	234	v	v
HDPG	CLM0209	24.88273	121.5009	235	v	v
HDPG	CLM0210	24.88411	121.5029	223	v	v
HDPG	CLM0211	24.88415	121.503	224	v	v
HDPG	CLM0212	24.88478	121.5036	216	v	v
TTL	CLM0228	23.45159	120.6137	433	v	v



TTL	CLM0229	23.45215	120.6137	413	v	v
TTL	CLM0230	23.45241	120.6131	417	v	v
TTL	CLM0231	23.45244	120.6126	421	v	v
TTL	CLM0232	23.45226	120.6121	433	v	v
TTL	CLM0233	23.45219	120.612	438	v	v
TTL	CLM0234	23.45212	120.6118	441	v	v
TTL	CLM0235	23.45205	120.611	453	v	v
TTL	CLM0236	23.45214	120.6108	457	v	v
TTL	CLM0237	23.45285	120.6111	440	v	v
TTL	CLM0238	23.45362	120.6114	437	v	v
TTL	CLM0239	23.45357	120.6115	435	v	v
NAJY	CLM0261	24.43272	121.7606	98	v	v
NAJY	CLM0262	24.43254	121.7614	106	v	v
NAJY	CLM0263	24.43244	121.7616	110	v	v
NAJY	CLM0264	24.4321	121.7619	112	v	v
NAJY	CLM0265	24.42486	121.7634	148	v	v
NAJY	CLM0266	24.42448	121.7636	145	v	v
NAJY	CLM0267	24.42392	121.7641	147	v	v
NAJY	CLM0268	24.4237	121.7643	146	v	v
NAJY	CLM0269	24.42345	121.7643	143	v	v
NAJY	CLM0270	24.42267	121.7625	153	v	v
NAJY	CLM0271	24.42248	121.7624	155	v	v
NAJY	CLM0272	24.42167	121.7629	162	v	v
NAJY	CLM0273	24.42145	121.763	165	v	v
NAJY	CLM0274	24.42121	121.7629	167	v	v
HLCN	CLM0275	23.90632	121.493	249	v	v
HLCN	CLM0276	23.90623	121.4928	249	v	v
HLCN	CLM0277	23.90616	121.4927	249	v	v
HLCN	CLM0278	23.90581	121.4921	246	v	v
HLCN	CLM0279	23.90578	121.4916	246	v	v
HLCN	CLM0280	23.90563	121.491	248	v	v
HLCN	CLM0281	23.90447	121.4888	266	v	v
HLCN	CLM0282	23.90456	121.4885	267	v	v
HLCN	CLM0283	23.90492	121.4872	303	v	v
HLCN	CLM0284	23.90471	121.487	306	v	v
HLCN	CLM0285	23.90449	121.4869	294	v	v
HLCN	CLM0286	23.90413	121.4869	279	v	v
HLCN	CLM0287	23.90386	121.4868	279	v	v



HLCN	CLM0288	23.90352	121.4865	289	v	
HLCN	CLM0289	23.90318	121.4863	292	v	
NXIR	CLM0290	23.40875	121.4504	118	v	
NXIR	CLM0291	23.40884	121.4503	117	v	
NXIR	CLM0292	23.40958	121.4497	124	v	
NXIR	CLM0293	23.40979	121.4493	127	v	
NXIR	CLM0294	23.40993	121.4491	129	v	v
NXIR	CLM0295	23.41069	121.4481	148	v	v
NXIR	CLM0296	23.41058	121.4492	150	v	
NXIR	CLM0297	23.41058	121.4495	151	v	
NXIR	CLM0298	23.41057	121.4486	147	v	v
NXIR	CLM0299	23.40839	121.4504	127	v	v
NXIR	CLM0300	23.40854	121.4493	136	v	
NXIR	CLM0301	23.40887	121.4481	146	v	v
NXIR	CLM0302	23.40959	121.4467	155	v	v
NXIR	CLM0303	23.4097	121.4466	152	v	v
NXIR	CLM0304	23.40996	121.4461	151	v	v
DFR	CLM0305	23.10971	121.2745	536	v	
DFR	CLM0306	23.10958	121.2746	535	v	v
DFR	CLM0307	23.10758	121.2752	551	v	v
DFR	CLM0308	23.10752	121.2751	552	v	v
DFR	CLM0309	23.10535	121.2724	585	v	v
DFR	CLM0310	23.10518	121.2723	587	v	v
DFR	CLM0311	23.10535	121.2724	585	v	v
DFR	CLM0312	23.10502	121.2723	583	v	v
DFR	CLM0313	23.10356	121.2724	587	v	v
DFR	CLM0314	23.10245	121.2719	599	v	
DFR	CLM0315	23.10211	121.272	599	v	
DFR	CLM0316	23.10183	121.2725	589	v	
DFR	CLM0317	23.10033	121.2759	597	v	v
DFR	CLM0318	23.09964	121.2747	586	v	
DFR	CLM0319	23.09905	121.2742	581	v	v



Genomic DNA used for different analyses is marked “v” in the last two columns.

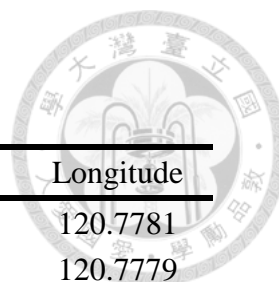
**Table 2. SSR primer information.**

Primer	Sequence (5' to 3')	Chromosome*	Position*
CIR436	ATAAGCTCATATGGGTACAGTCACA CTGCAGCAACCAAATTTATTTCT	1	4917979 - 4918003 4918057 - 4918080
CIR276	CTCCTCCATAGCCTGACTGC TGACCCACGAGAAAAGAAGC	1	13588759 - 13588778 13588847 - 13588866
CIR1113	ACTCTCGCCCATCTTCATCC ACTTATCCCCCGCACTCAA	1	28009948 - 28009967 28010173 - 28010192
Ma-3-132	TCCCTCTTCAACCAAAGCAC AACGCGAATGTGTGTTTCA	2	20365901 - 20365920 20366041 - 20366060
CIR646	AACACCGTACAGGGAGTCAC GATACATAAGGCAGTCACATTG	2	23940966 - 23940985 23941276 - 23941297
Ma-1-17	AGGCGGGGAATCGGTAGA GGCGGGAGACAGATGGAGT	2	24383391 - 24383408 24383488 - 24383506
CIR332a	ATGACCTGTCGAACATCCTTT TCCAACCCCTGCAACCACT	3	8978575 - 8978595 8978831 - 8978848
Ma-3-48	CCCGTCCCATTCTCA TTCGTTGTTTCATGGAATCA	5	32672883 - 32672898 32673018 - 32673036
CIR631a	ATTAGATCACCGAAGAACTC ATCTTTTCTTATCCTTCTAACG	6	33942751 - 33942770 33943017 - 33943038
CIR16a	TCATCTCACAAATGCTTTCATAGTT TGGTTGAGTAGATCTTCTTGTGT	8	1234609 - 1234632 1234700 - 1234722
Ma-3-103	TCGCCTCTCTTTAGCTCTG TGTTGGAGGATCTGAGATTG	8	40127782 - 40127800 40127910 - 40127929
CIR348b	ACAGAATCGCTAACCCCTAATCCTCA CCCTTTGCGTGCCCCTAA	10	27228162 - 27228186 27228325 - 27228342
Ma-3-139	ACTGCTGCTCTCCACCTCAAC GTCCCCCAAGAACCATATGATT	10	30237591 - 30237611 30237717 - 30237735
CIR550a	ACCGCACCTCCACCTCCTG TGCTGCCTTCATCGCTACTA	10	31914219 - 31914237 31914459 - 31914478

Primer sequences were searched against the *Musa acuminata* DH-Pahang version 2 (38)

on Banana Genome Hub (<https://banana-genome-hub.southgreen.fr/>).

\*The chromosome and position indicate locations on *Musa acuminata* where primer sequences were blast.



**Table 3. Integrated presence-only data.**

No.	Latitude	Longitude	No.	Latitude	Longitude
1	24.68661	121.229	243	23.71152	120.7781
2	24.62276	121.5501	244	23.71119	120.7779
3	24.61333	121.5452	245	23.71077	120.7777
4	24.60428	121.5336	246	23.71044	120.7775
5	24.60427	121.5344	247	25.06747	121.7998
6	24.60436	121.5348	248	25.07479	121.7955
7	24.60446	121.5351	249	25.07419	121.7949
8	24.60459	121.5355	250	25.07405	121.7958
9	24.60463	121.5356	251	25.07177	121.7991
10	24.60472	121.5357	252	24.91284	121.5428
11	24.60475	121.536	253	23.70981	120.7767
12	24.60499	121.5366	254	23.7099	120.7768
13	24.60514	121.5373	255	23.71013	120.7766
14	24.60526	121.5374	256	23.6949	120.7858
15	24.58588	121.5148	257	23.69275	120.7858
16	24.58469	121.5138	258	23.6935	120.786
17	24.58422	121.5132	259	24.43272	121.7606
18	24.58356	121.5129	260	24.43254	121.7614
19	24.58336	121.5128	261	24.43244	121.7616
20	24.58227	121.5124	262	24.4321	121.7619
21	24.58163	121.5112	263	24.42486	121.7634
22	24.58085	121.5106	264	24.42448	121.7636
23	24.5802	121.5102	265	24.42392	121.7641
24	24.61676	121.1546	266	24.4237	121.7643
25	23.10835	121.2918	267	24.42345	121.7643
26	24.83443	121.3905	268	24.42267	121.7625
27	22.53487	120.9658	269	24.42248	121.7624
28	24.62621	121.5546	270	24.42167	121.7629
29	24.58401	121.5131	271	24.42145	121.763
30	24.58374	121.5129	272	24.42121	121.7629
31	24.5814	121.511	273	23.90632	121.493
32	24.55402	121.5079	274	23.90623	121.4928
33	24.55415	121.5082	275	23.90616	121.4927
34	24.55403	121.5089	276	23.90581	121.4921
35	24.55385	121.5091	277	23.90578	121.4916

36	24.55383	121.5091	278	23.90563	121.491
37	24.553	121.5095	279	23.90447	121.4888
38	24.55218	121.5093	280	23.90456	121.4885
39	24.55209	121.5093	281	23.90492	121.4872
40	24.55144	121.5091	282	23.90471	121.487
41	24.55002	121.5083	283	23.90449	121.4869
42	24.54913	121.5116	284	23.90413	121.4869
43	24.54861	121.5119	285	23.90386	121.4868
44	24.54733	121.5124	286	23.90352	121.4865
45	24.54612	121.5128	287	23.90318	121.4863
46	24.54621	121.5126	288	23.40875	121.4504
47	24.54486	121.5126	289	23.40884	121.4503
48	24.54485	121.5126	290	23.40958	121.4497
49	24.54467	121.5125	291	23.40979	121.4493
50	24.54452	121.512	292	23.40993	121.4491
51	24.5444	121.5117	293	23.41069	121.4481
52	24.54408	121.5113	294	23.41058	121.4492
53	24.53878	121.5219	295	23.41058	121.4495
54	24.5399	121.5126	296	23.41057	121.4486
55	24.53934	121.5122	297	23.40839	121.4504
56	24.53912	121.512	298	23.40854	121.4493
57	24.53962	121.5119	299	23.40887	121.4481
58	24.54018	121.5123	300	23.40959	121.4467
59	24.54005	121.5122	301	23.4097	121.4466
60	24.53983	121.5121	302	23.40996	121.4461
61	24.53953	121.5119	303	23.10971	121.2745
62	24.53922	121.5119	304	23.10958	121.2746
63	24.57036	121.4993	305	23.10758	121.2752
64	24.57036	121.4993	306	23.10752	121.2751
65	24.56993	121.4996	307	23.10535	121.2724
66	24.5691	121.4996	308	23.10518	121.2723
67	24.56832	121.4994	309	23.10535	121.2724
68	24.33525	120.7853	310	23.10502	121.2723
69	24.33484	120.7858	311	23.10356	121.2724
70	24.69913	121.1537	312	23.10245	121.2719
71	24.69826	121.1529	313	23.10211	121.272
72	24.69789	121.1528	314	23.10183	121.2725
73	24.69782	121.1527	315	23.10033	121.2759





74	24.6978	121.1527	316	23.09964	121.2747
75	24.69766	121.1526	317	23.09905	121.2742
76	24.69748	121.1525	318	23.27941	121.3464
77	24.69743	121.1525	319	24.71341	121.1137
78	24.69688	121.1514	320	24.71137	121.1147
79	24.69683	121.1511	321	23.76997	120.9408
80	24.69682	121.1509	322	23.4855	120.7572
81	24.69658	121.1506	323	24.41948	121.7631
82	23.71032	120.7774	324	24.01298	121.544
83	23.71002	120.7773	325	23.82657	121.4252
84	23.70981	120.7771	326	24.00425	121.5369
85	23.70994	120.7773	327	23.97538	121.5355
86	23.69081	120.7853	328	23.99935	121.433
87	23.69314	120.786	329	23.92911	121.4947
88	23.69323	120.786	330	23.83699	121.4226
89	23.69354	120.7859	331	23.81386	121.542
90	23.69405	120.786	332	24.89776	121.5705
91	23.69422	120.7859	333	24.84946	121.579
92	23.69435	120.7859	334	24.78193	121.5047
93	23.6949	120.7858	335	24.77708	121.482
94	23.69512	120.7858	336	24.47827	121.4209
95	23.67361	120.7866	337	24.48819	121.4334
96	23.67358	120.786	338	24.4858	121.4256
97	23.67349	120.7855	339	24.60777	121.4866
98	23.67338	120.7855	340	24.60417	121.4961
99	23.67313	120.7851	341	24.60772	121.5033
100	23.67279	120.7849	342	24.67661	121.3562
101	23.67266	120.7846	343	24.79827	121.3201
102	23.67251	120.7844	344	24.27296	120.9119
103	23.6725	120.7843	345	24.24832	120.9186
104	23.6724	120.7837	346	24.25119	120.9177
105	23.67226	120.7834	347	24.24724	120.9111
106	23.66741	120.7717	348	23.89392	120.9
107	23.66883	120.7715	349	24.01224	121.0711
108	23.66951	120.771	350	24.10386	120.853
109	23.6714	120.7726	351	24.15118	120.8077
110	23.67186	120.7728	352	23.70394	120.6345
111	23.67161	120.7729	353	23.7018	120.6324



112	23.67278	120.7729	354	24.419	120.742
113	23.67284	120.7731	355	22.66975	120.6877
114	23.67229	120.7734	356	22.66933	120.6946
115	23.67239	120.7733	357	22.68218	120.6813
116	24.7813	121.3587	358	22.26229	120.8496
117	22.24463	120.835	359	22.25786	120.842
118	22.2448	120.8353	360	22.25378	120.8415
119	22.24513	120.8367	361	24.71359	121.6514
120	22.24495	120.8369	362	24.70048	121.6119
121	22.24407	120.8394	363	24.70383	121.6407
122	22.24326	120.8508	364	24.63892	121.5546
123	22.24332	120.8509	365	24.62729	121.5205
124	22.23958	120.8575	366	24.61627	121.6367
125	23.10148	120.6869	367	25.03128	121.7968
126	22.15934	120.824	368	25.25787	121.5979
127	22.1873	120.8756	369	23.87936	121.4902
128	22.23862	120.8575	370	23.7606	121.4382
129	22.23865	120.8576	371	23.75373	121.4238
130	22.23777	120.8584	372	23.70575	121.3924
131	22.23735	120.8599	373	23.66266	121.4065
132	22.22688	120.8679	374	23.57077	121.3575
133	22.20977	120.8643	375	23.48988	121.3267
134	22.19893	120.8514	376	23.4399	121.3355
135	22.18587	120.8609	377	23.37335	121.3177
136	24.56884	121.4995	378	23.40048	121.3092
137	24.56705	121.4993	379	23.2742	121.2543
138	24.56662	121.4992	380	23.22564	121.2604
139	24.56639	121.499	381	25.12625	121.6534
140	24.56558	121.4989	382	25.14488	121.6787
141	24.3352	120.7832	383	22.66763	120.6996
142	24.33541	120.7844	384	22.66913	120.6968
143	24.33551	120.7847	385	24.45524	120.8581
144	24.335	120.7859	386	22.94361	121.1051
145	24.33525	120.7863	387	22.94061	121.1172
146	24.33532	120.7864	388	22.71047	120.996
147	24.33541	120.7866	389	22.58201	120.9731
148	24.33584	120.7867	390	22.48514	120.9172
149	24.33591	120.787	391	22.3718	120.8889



150	24.33642	120.787	392	22.41534	120.8489
151	24.33632	120.7869	393	23.12748	121.3481
152	22.19621	120.8793	394	23.25711	121.3837
153	24.55218	121.5093	395	23.40548	121.4338
154	23.69366	120.7859	396	23.40858	121.4493
155	23.67337	120.7854	397	24.9355	121.6568
156	25.06671	121.8001	398	23.47265	121.4809
157	25.06686	121.7993	399	23.48579	121.4741
158	25.06638	121.7985	400	23.49765	121.4228
159	25.0666	121.8002	401	23.6107	121.5071
160	25.07073	121.7995	402	23.62768	121.4993
161	25.07183	121.7987	403	23.72981	121.508
162	25.07218	121.7985	404	24.7813	121.3587
163	25.07243	121.7977	405	24.61676	121.1546
164	25.07242	121.7974	406	24.83443	121.3905
165	24.92046	121.5115	407	22.53487	120.9658
166	24.91049	121.511	408	24.62621	121.5546
167	24.90145	121.4978	409	24.89618	121.3949
168	24.88384	121.5023	410	25.02319	121.7706
169	24.88383	121.5022	411	25.08722	121.793
170	24.8836	121.5021	412	25.0837	121.7861
171	24.88337	121.5019	413	24.41998	121.732
172	24.88313	121.5016	414	24.04696	120.8066
173	24.88312	121.5016	415	24.03439	120.7803
174	24.88293	121.5013	416	24.03537	120.7765
175	24.88292	121.5013	417	24.03548	120.7741
176	24.88273	121.5009	418	24.03764	120.7646
177	24.88411	121.5029	419	24.03696	120.7625
178	24.88415	121.503	420	24.03749	120.7623
179	24.88478	121.5036	421	24.05166	120.7794
180	24.08425	120.7876	422	24.05164	120.7654
181	24.08309	120.7981	423	23.31425	120.5017
182	24.08274	120.8001	424	23.37682	120.5904
183	24.08266	120.8002	425	23.34793	120.6195
184	24.08261	120.8006	426	24.0345	120.7797
185	24.08269	120.8008	427	23.36246	120.6193
186	24.08258	120.803	428	23.36383	120.6184
187	24.08212	120.8043	429	23.36494	120.6177

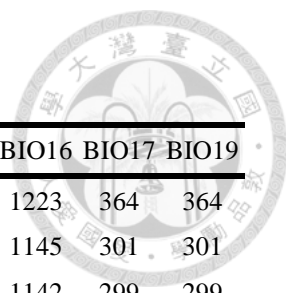


188	24.0813	120.8074	430	23.36514	120.6166
189	24.08134	120.8074	431	23.36345	120.6143
190	24.08167	120.8086	432	23.3645	120.6149
191	24.08164	120.809	433	23.36513	120.6155
192	24.0817	120.8079	434	23.36287	120.6108
193	24.07969	120.8079	435	23.36187	120.6107
194	24.08017	120.8076	436	24.05237	120.981
195	23.45159	120.6137	437	24.15331	120.8126
196	23.45215	120.6137	438	24.09865	120.8119
197	23.45241	120.6131	439	24.09013	120.8081
198	23.45244	120.6126	440	22.57301	120.6583
199	23.45226	120.6121	441	23.06605	121.3017
200	23.45219	120.612	442	23.07139	121.3015
201	23.45212	120.6118	443	23.36184	120.5806
202	23.45205	120.611	444	24.65611	121.5847
203	23.45214	120.6108	445	24.60462	121.5079
204	23.45285	120.6111	446	24.89542	121.7266
205	23.45362	120.6114	447	24.64151	121.5608
206	23.45357	120.6115	448	24.64151	121.5608
207	24.69311	121.1451	449	24.64151	121.5608
208	22.56724	120.6454	450	25.00026	121.9923
209	22.56481	120.6465	451	23.36814	120.611
210	22.5657	120.6475	452	23.10867	121.2933
211	22.56364	120.6456	453	22.2036	120.7424
212	22.56276	120.6452	454	23.09963	121.2727
213	22.56327	120.6443	455	23.10564	121.2729
214	22.56226	120.6449	456	23.36814	120.611
215	22.57957	120.6505	457	23.43363	120.6398
216	22.58491	120.6484	458	23.43369	120.6154
217	23.7413	120.7367	459	24.5852	121.033
218	23.74138	120.7367	460	24.68691	121.2287
219	23.74167	120.7367	461	24.24274	120.9116
220	23.7425	120.7396	462	24.49218	121.429
221	23.74263	120.7394	463	24.74367	121.3465
222	23.74267	120.7393	464	24.76854	121.142
223	23.74279	120.7391	465	24.83851	121.4095
224	23.74288	120.739	466	24.84268	121.4
225	23.74294	120.7388	467	24.83341	121.4083



226	23.74293	120.7379	468	24.83184	121.4078
227	23.74281	120.7376	469	24.8344	121.3892
228	23.74211	120.7368	470	24.83456	121.39
229	23.72787	120.7618	471	24.83449	121.3907
230	23.72796	120.7621	472	24.83418	121.3908
231	23.72797	120.7622	473	25.23869	121.5571
232	23.72816	120.7623	474	25.20907	121.5902
233	23.72826	120.7627	475	25.23856	121.5569
234	23.72807	120.7634	476	24.84671	121.5447
235	23.72796	120.7637	477	24.84702	121.5508
236	23.72799	120.7643	478	24.83947	121.5331
237	23.72785	120.7644	479	24.95613	121.7872
238	23.72775	120.7652	480	24.97453	121.9266
239	23.72776	120.7655	481	25.04844	121.7878
240	23.71015	120.7792	482	24.83642	121.6528
241	23.71089	120.7789	483	24.94826	121.7925
242	23.71158	120.7786	-	-	-

Presence-only data incorporates collection data and Google Street view data, resulting in 483 occurrence points.

**Table 4. 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	2970	46	1223	364	364
TPS400	24.58135	121.5124	194	65	36	177	2769	50	1145	301	301
TPS500	24.56765	121.4992	199	65	36	179	2763	50	1142	299	299
TPS600	24.55172	121.509	193	65	36	176	2709	51	1124	289	289
TPS700	24.54638	121.5125	185	65	37	174	2699	49	1107	302	302
TPS900	24.54021	121.5126	178	66	38	173	2757	46	1110	336	336
XT400	23.74292	120.7386	217	77	42	180	2468	86	1395	87	127
XT500	23.7278	120.7641	206	77	44	173	2600	84	1420	97	141
XT700	23.71137	120.7782	199	77	45	169	2412	83	1311	94	138
XT900	23.69294	120.7857	189	77	47	162	2357	83	1287	97	145
XT1200	23.67333	120.7854	176	77	48	159	2478	82	1346	97	148
XT1500	23.66936	120.771	160	78	48	160	2637	81	1402	98	149
C35H	24.69907	121.1528	202	65	34	188	2500	55	1073	192	289
WFL	25.06752	121.7998	203	58	31	185	3440	29	1229	606	749
THNL	24.08273	120.8021	212	74	40	184	2188	83	1220	78	138
PTWT	22.58684	120.6489	227	78	49	158	2912	98	1784	79	87
P199H	22.24302	120.8465	223	73	50	144	3114	92	1856	178	178
MLLYT	24.33526	120.7864	211	71	37	190	2056	76	1084	79	148
HDPG	24.88347	121.5019	197	62	33	185	3201	37	1201	475	475
TTL	23.4523	120.6134	215	79	45	174	3515	98	2154	81	123
NAJY	24.43352	121.7602	216	62	35	174	2860	61	1328	271	296
HLCN	23.90623	121.4928	213	68	40	168	2127	46	849	251	255
NXIR	23.40887	121.4503	227	71	43	163	2147	57	938	211	211
DFR	23.10356	121.2724	206	75	48	155	2017	68	1026	166	166




**Table 5. MaxEnt permutation importance.**

<b>Variable</b>	<b>Permutation importance</b>
BIO1	36.7
BIO17	27.3
BIO12	9.4
BIO19	7.5
BIO3	6.9
BIO15	5.8
BIO16	3.8
BIO2	1.4
BIO7	1.2

Permutation importance is determined by randomly permuting the values of that variable among the training points and measuring the resulting decrease in training predictive power. A large value indicates that the model depends heavily on that variable.

**Table 6. SNP composition.**

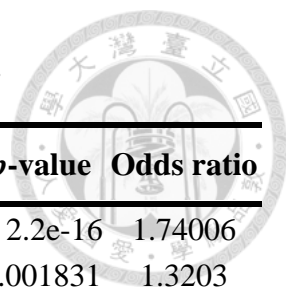


Variable	Non-adaptive NM	Non-adaptive SV	Adaptive NM	Adaptive SV	<i>p</i> -value	Odds ratio
BIO1	317,505	931,458	982	6,949	< 2.2e-16	2.41212
BIO2	317,411	930,311	1,076	8,096	< 2.2e-16	2.56716
BIO3	317,719	930,983	768	7,424	< 2.2e-16	3.29897
BIO7	317,802	931,072	685	7,335	< 2.2e-16	3.65496
BIO12	317,625	931,349	862	7,058	< 2.2e-16	2.79239
BIO15	317,242	928,623	1,245	9,784	< 2.2e-16	2.68472
BIO16	318,069	934,612	418	3,795	< 2.2e-16	3.08977
BIO17	317,614	932,066	873	6,341	< 2.2e-16	2.47512
BIO19	317,663	932,429	824	5,978	< 2.2e-16	2.47161

The number of four sets of SNPs is recorded. New mutation is abbreviated as NM; standing variation is abbreviated as SV. Statistical significance from  $\chi^2$  test is shown for each bioclimatic variable. Odds ratio is calculated as (“adaptive SV” / “adaptive NM”) / (“non-adaptive SV” / “non-adaptive NM”).



**Table 7. SNP composition (controlled for minor allele frequency).**



Variable	Non-adaptive NM	Non-adaptive SV	Adaptive NM	Adaptive SV	<i>p</i> -value	Odds ratio
BIO1	51,405	253,881	293	2,518	< 2.2e-16	1.74006
BIO2	26,893	145,492	147	1,050	0.001831	1.3203
BIO3	45,243	229,020	229	2,642	< 2.2e-16	2.27916
BIO7	47,029	236,530	227	2,750	< 2.2e-16	2.40872
BIO12	42,821	219,108	270	2,464	< 2.2e-16	1.78351
BIO15	37,778	197,032	392	2,545	6.41e-05	1.24481
BIO16	38,733	201,947	142	1,253	2.49e-09	1.69241
BIO17	13,755	70,721	75	603	0.000296	1.56375
BIO19	12,625	65,198	48	422	0.000532	1.70243

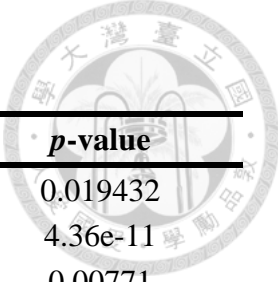
The number of four sets of SNPs whose minor allele frequency (MAF) ranges from the first quantile of adaptive MAF to the third quantile of non-adaptive MAF separately for each bioclimatic variable is recorded. New mutation is abbreviated as NM; standing variation is abbreviated as SV. Statistical significance from  $\chi^2$  test is shown for each bioclimatic variable. Odds ratio is calculated as (“adaptive SV” / “adaptive NM”) / (“non-adaptive SV” / “non-adaptive NM”).

**Table 8. Significance test for allele frequency distribution within the novel and ancestral environmental range.**

<b>Variable</b>	<b>Novel</b>	<b>Ancestral</b>	<b><i>p</i>-value</b>
BIO1	0.412705	0.391079	0.00639
BIO2	0.53879	0.384593	1.33e-15
BIO3	0.533299	0.45664	2.45e-05
BIO7	0.502003	0.450798	0.004346
BIO12	0.456482	0.482977	0.1015
BIO15	0.471812	0.449859	0.000116
BIO16	0.430376	0.45188	0.212893
BIO17	0.571548	0.343989	1.34e-28
BIO19	0.603138	0.340112	1.19e-26

The averaged frequency of newly mutated alleles within the novel and ancestral environmental range in Taiwan is recorded. Statistical significance from paired *t*-test is shown for each bioclimatic comparison between the novel and ancestral environmental range.

**Table 9. Significance test for Bayesian Bayes factor distribution.**



<b>Variable</b>	<b>NM (rank)</b>	<b>SV (rank)</b>	<b><i>p</i>-value</b>
BIO1	3,806.171	3,988.586	0.019432
BIO2	5,085.828	4,520.137	4.36e-11
BIO3	3,880.061	4,118.89	0.00771
BIO7	3,741.818	4,035.592	0.00149
BIO12	3,874.063	3,971.057	0.239722
BIO15	6,255.701	5,420.747	2.56e-18
BIO16	2,212.05	2,095.429	0.062812
BIO17	3,975.735	3,556.803	2.44e-08
BIO19	3,662.596	3,365.511	4.64e-05

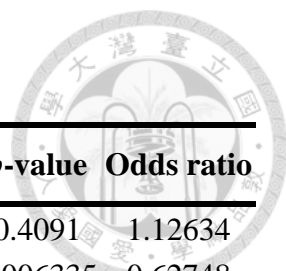
The average of ranked Bayes factor of adaptive new mutation (NM) and standing variation (SV) is recorded. Statistical significance from Wilcoxon rank sum test is shown for each bioclimatic comparison between NM and SV.

**Table 10. Significance test for Gradient Forest  $r^2$  importance and slope distribution.**

Variable	$r^2$			Slope		
	NM	SV	$p$ -value	NM	SV	$p$ -value
BIO1	0.20225	0.23771	2.66e-05	0.13242	0.16355	6.56e-14
BIO2	0.41309	0.36337	2.47e-14	0.2671	0.23895	6.38e-19
BIO3	0.28392	0.28256	0.89344	0.19906	0.19133	0.13266
BIO7	0.15727	0.19032	2.88e-05	0.1325	0.14948	1.29e-05
BIO12	0.32996	0.31662	0.12053	0.20024	0.19491	0.13391
BIO15	0.49811	0.40632	2.97e-30	0.2897	0.24879	1.60e-32
BIO16	0.30961	0.29408	0.25876	0.19231	0.189	0.51441
BIO17	0.5378	0.43781	4.84e-19	0.30059	0.26587	2.50e-13
BIO19	0.45297	0.34754	2.67e-28	0.27773	0.23629	8.23e-22

The averaged  $r^2$  importance and slope of adaptive new mutation (NM) and standing variation (SV) are recorded. Statistical significance from Welch two sample  $t$ -test is shown for each bioclimatic comparison between NM and SV.

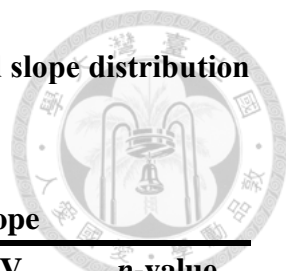
**Table 11. SNP composition for the future.**



<b>Variable</b>	<b>Disruption NM</b>	<b>Disruption SV</b>	<b>Retention NM</b>	<b>Retention SV</b>	<b><i>p</i>-value</b>	<b>Odds ratio</b>
BIO1	106	623	171	1,132	0.4091	1.12634
BIO2	45	363	389	1,969	0.006335	0.62748
BIO3	29	261	150	2,246	0.0222	1.6637
BIO7	37	243	162	2,359	4.66e-05	2.21722
BIO12	76	470	163	1,482	0.01155	1.4702
BIO15	47	355	555	2,864	0.02192	0.6832
BIO16	41	274	29	233	0.5584	1.20224
BIO17	46	345	236	1,105	0.007549	0.62429
BIO19	34	352	111	424	1.46e-06	0.36896

The number of four sets of SNPs under future scenarios is recorded. New mutation is abbreviated as NM; standing variation is abbreviated as SV. Statistical significance from  $\chi^2$  test is shown for each bioclimatic variable. Odds ratio is calculated as (“retention SV” / “retention NM”) / (“disruption SV” / “disruption NM”).

**Table 12. Significance test for Gradient Forest  $r^2$  importance and slope distribution (controlled for minor allele frequency).**



Variable	$r^2$			Slope		
	NM	SV	<i>p</i> -value	NM	SV	<i>p</i> -value
BIO1	0.19737	0.21328	0.23039	0.16593	0.17117	0.42363
BIO2	0.42462	0.38436	7.93e-10	0.28673	0.26359	2.85e-13
BIO3	0.33661	0.33365	0.8402	0.23496	0.22597	0.21744
BIO7	0.15572	0.2014	2.66e-06	0.14358	0.16534	8.91e-06
BIO12	0.3322	0.32251	0.42739	0.21776	0.2142	0.50127
BIO15	0.57536	0.49643	2.38e-18	0.33311	0.29867	1.70e-21
BIO16	0.23986	0.24695	0.61079	0.18398	0.18725	0.60515
BIO17	0.5909	0.49654	1.13e-13	0.33493	0.29796	6.90e-13
BIO19	0.48532	0.38889	2.37e-18	0.30343	0.2628	4.39e-18

The averaged  $r^2$  importance and slope of adaptive new mutation (NM) and standing variation (SV) whose minor allele frequency ranks top 50 % in all adaptive SNPs are recorded. Statistical significance from Welch two sample *t*-test is shown for each bioclimatic comparison between NM and SV.