國立臺灣大學生命科學院生態學與演化生物學研究所

碩士論文

Institute of Ecology and Evolutionary Biology

College of Life Science

National Taiwan University

Master Thesis

間歇性遷入與染色體倒位共同影響尼泊爾埋葬蟲種內遺傳多樣性及環境適應

Intermittent immigration and chromosomal inversions jointly influence intraspecific genetic diversity and adaptation in burying beetles

王鵬

Peng Wang

指導教授：沈聖峰 博士

Advisor: Sheng-Feng Shen, Ph.D.

中華民國 112 年 7 月

July 2023

doi:10.6342/NTU202302705

Acknowledgments

兩年的時光說長不長、說短不短，在寫下這篇誌謝時，我已經決定未來不會繼續走在學術研究的道路上。雖然有些可惜，但想到當初本就是為了探究自己是否真的想要、適合從事學術研究，才決定來到這裡，也就感到這段日子是無比充實，且難能可貴的。

首先要感謝沈聖峰老師，願意收留我這個沒有參與過任何研究室的人，使完全沒有經驗的我能有一個實際體驗學術研究生活的機會，並且在這過程中給予我許多的指教與鼓勵，以及適時的引導我論文寫作的方向，讓我能順利的完成碩士學業。同時也很感謝沈老師總是不吝於分享研究上的構想及對學術生涯的看法，讓我得以窺見對學術研究充滿熱忱的學者的思考方式。此外也很感謝李承叡老師、蔡怡陞老師、李壽先老師及王忠信老師為我的論文提供了許多建議與指點，讓我的論文能夠更加完善。

接著是研究室的夥伴們，很感謝伯飛帶著我學習分生實驗的操作以及教導我許多基因體分析的知識，謝謝彥成和百佑時常協助我處理研究生活中的大小事，謝謝袁子能對野外實驗的各種支援，以及時常關心我的研究生活、給予我鼓勵，也感謝所有的學長姐，特別是陳澔、勳承、詩蘋、崇凡，帶著我一起做實驗、聊天、做各種活動，不僅引領我更快的進入狀況，也為我兩年的研究生活增添了許多色彩。也謝謝宜軒在口試當天幫忙布置場地及準備茶水、點心，還有詩蘋及慈蔓協助我紀錄口試。

此外，也謝謝幸芸給予我陪伴、支持、鼓勵及許多的正能量，也讓我寫論文時的各種甘苦有分享的對象。

最後，謝謝我的父母從來不限制我探索自己的興趣，總是讓我衣食無缺、心無旁騖的做自己想做的事，我由衷地感到幸福，也很幸運。

doi:10.6342/NTU202302705

摘要

生物如何適應環境是研究生物多樣性形成的核心問題之一。適應性遺傳變異來源可以是新形成的突變或是既有的遺傳變異（standing genetic variation），我們對於兩者的重要性及他們如何共同作用的理解仍然不足。同時，我們也需要更多的實證研究來釐清生物如何在基因流下達成地區性適應（local adaptation）。了解這兩個生物適應的問題能增加我們對生物多樣性形成的基本原理。本研究我們以在臺灣表現出生殖光週期的地區性適應現象的尼泊爾埋葬蟲為對象，透過比較全年繁殖及季節性繁殖個體間的遺傳差異，我們發現不同族群透過不同的遺傳機制決定生殖光週期的表現型。而重建親緣關係樹的結果顯示，臺灣現今的尼泊爾埋葬蟲族群是由多次的遷入事件所構成。族群遺傳結構分析的結果則顯示染色體倒位幫助了維持族群間的遺傳差異。我們認為多次遷入與染色體倒位共同導致了多種遺傳調控機制的現象，並且也使得新突變與既有遺傳變異共同幫助了基因流下的本地適應。本篇研究展示了複雜的族群形成歷史與染色體倒位在地區性適應中扮演的角色，為生物多樣性的起源機制提供了新的觀點。

關鍵字：尼泊爾埋葬蟲、地區性適應、染色體倒位、既有遺傳變異

doi:10.6342/NTU202302705

Abstracts

How organisms adapt to their environment is one of the core questions in studying the formation of biological diversity. The source of adaptive genetic variation can either be newly formed mutations or standing genetic variations. Our understanding of the importance of these two and how they interact is still insufficient. Simultaneously, we also need more empirical research to clarify how organisms achieve local adaptation under gene flow. Understanding these two questions about biological adaptation can enhance our basic principles of the formation of biological diversity. In this study, we focused on the *Nicrophorus nepalensis*, which exhibits locally adapted reproductive photoperiodism in Taiwan. By comparing the genetic differences between individuals who reproduce throughout the year and those who reproduce seasonally, we found that different populations regulate the phenotype of reproductive photoperiodism through different genetic mechanisms. The results of reconstructing the phylogenetic tree show that the current *N. nepalensis* populations in Taiwan are composed of multiple immigration events. The results of population genetic structure analysis reveal that chromosomal inversion helps to maintain genetic differences between populations. We believe that multiple migrations and chromosomal inversions have collectively led to the phenomenon of multiple genetic regulatory mechanisms. Additionally, they have

doi:10.6342/NTU202302705

allowed new mutations and standing genetic variations to jointly assist in local adaptation under gene flow. This research demonstrates the roles of complex population formation history and chromosomal inversion in local adaptation, providing new perspectives on the mechanisms of origin of biological diversity.

Keywords: *Nicrophorus nepalensis*, Local adaptation, Chromosomal inversion, Standing genetic variation.

doi:10.6342/NTU202302705

Content

doi:10.6342/NTU202302705

doi:10.6342/NTU202302705

Content of Figures

doi:10.6342/NTU202302705

Content of Tables

doi:10.6342/NTU202302705

1. Introduction

Why are there such diverse forms of life on Earth? This enduring question has fascinated biologists for centuries. Darwin's theory of natural selection posits that individuals possessing characteristics more adapted to their environment have a higher probability of survival and successful reproduction. Over time, these advantageous traits are likely to increase in frequency within a population. If differences among populations accumulate gradually over time, it is possible that a wider variety of species may evolve.

For adaptation to occur, there must be various characteristics within the population from which to select. While traditional considerations focused on phenotypic diversity (Bolnick et al., 2011), the field of evolutionary genetics has increasingly recognized the contribution of genetic-level diversity to adaptation. For instance, cryptic genetic variation can promote population adaptation when environmental changes occur (Paaby & Rockman, 2014). In summary, genetic variation is an indispensable factor in the generation of adaptation in populations (Lande & Shannon, 1996).

The genetic variation that facilitates adaptation may arise from new mutations or from pre-existing variation within the population, known as standing genetic variations. Increasing research points out that standing genetic variations can enable rapid

doi:10.6342/NTU202302705

adaptation in populations (Bitter et al., 2019; Chaturvedi et al., 2021; Marques et al., 2019), many studies also suggest that most adaptations arise from standing genetic variations rather than new mutations (Barrett & Schluter, 2008; Lai et al., 2019). Therefore, standing genetic variations to some extent reflect the adaptability and vulnerability of organisms in response to changing environments (Jump et al., 2009; Prentis et al., 2008), they are also considered to play a significant role in the early stages of speciation (Fuhrmann et al., 2023; Jones et al., 2012). Despite the growing body of research elucidating the importance of standing genetic variations for adaptation, many aspects remain unclear regarding the mechanisms by which standing genetic variations arise, and how new mutations and standing genetic variations interact to influence adaptation.

In addition, how populations with gene flow achieve local adaptation is also a critical question in the study of adaptation and speciation. Gene flow is considered a fundamental force in adaptation and speciation because it can propagate genetic variations among populations, providing more genetic diversity for selection (Gomulkiewicz et al., 1999; Holt & Gomulkiewich, 1997; Kawecki, 2000). However, on the other hand, gene flow is also thought to hinder the formation of local adaptation (Bridle & Vines, 2007; Lenormand, 2002; Sanford et al., 2006; Smadja & Butlin, 2011)

doi:10.6342/NTU202302705

because it counteracts the genetic differentiation brought about by selection, and it might introduce detrimental alleles to the local populations, thus reducing their fitness.

In order to explore the source of adaptive genetic variation and how species maintain local adaptation under gene flow conditions, in this study, we used the *Nicrophorus nepalensis*, which is widely distributed in Asia and shows a local adaptation phenomenon in Taiwanese populations, as research material. We aim to identify the genetic variations related to adaptation in this population and trace their origin and formation mechanisms.

N. nepalensis is a stenothermic insect (Hwang & Shiao, 2011). In many populations inhabiting different latitudes, both field population density surveys and laboratory thermal tolerance experiments reveal that these populations possess similar thermal niches (with an optimal temperature of approximately 16°C~18°C, Tsai et al., 2020). However, due the to different latitudes and elevational ranges of their habitats, some populations are unable to rely solely on migration to higher elevations to endure the relatively hot summer. Therefore, some populations undergo reproductive diapause to survive the warm season (Tsai et al., 2020).

The key factor triggering the diapause response in *N. nepalensis* is photoperiod (Hwang & Shiao, 2011). As mentioned, different habitats have different latitudinal and elevational ranges, thus, for different populations, the photoperiod corresponding to the

doi:10.6342/NTU202302705

suitable reproductive season also varies. Accordingly, *N. nepalensis* in Taiwan demonstrates locally-adapted reproductive photoperiodism (Tsai et al., 2020): the population in Mt. Hehuan does not enter diapause under long-day stimulation due to the mountain's year-round provision of habitats with suitable temperatures for this population, in contrast, the upper limit of Wulai mountain's elevation is lower, and it cannot provide suitable temperature habitats in summer, some individuals here enter diapause under short-day stimulation. Transplant experiments confirm that such differentiation in reproductive photoperiodism is the result of local adaptation (Tsai et al., 2020).

In our previous research, following the method of common garden experiment by Tsai et al. (2020), we found that *N. nepalensis* from Mt. Yangming, when reared under long-day conditions (14L:10D light conditions), failed to reproduce successfully. On the contrary, when reared under short-day conditions (10L:14D light conditions), almost all individuals were able to reproduce successfully. Concurrently, our field population density surveys showed that there was hardly any *N. nepalensis* found in Mt. Yangming during the warmer months from June to September. Therefore, even without the evidence from transplant experiments, we believe that the Mt. Yangming population likely exhibits local adaptation in reproductive photoperiodism.

doi:10.6342/NTU202302705

In addition, we have observed a chromosomal inversion occurred in the *N. nepalensis* populations in Taiwan and Fujian among all Asian populations. In some individuals, an inverted segment of approximately 13MB is present on the third chromosome (Figure 1). Moreover, in Taiwan, similar to the distribution of the reproductive photoperiod phenotype, the frequency of chromosomal inversions also shows a gradual difference among three populations (Mt.Yangming, Wulai, and Mt. Hehuan) (Figure 2). That is, no inversion occurs in the populations south of Mt. Lala (Mt. Lala, Mt. Hehuan, and Dahan forest road), while it exists at a certain frequency in the Wulai population, and is fixed in the Mt.Yangming population.

However, the genotype of the chromosomal inversion does not completely correspond with the reproductive photoperiod phenotype. In the Wulai population, we found that a small portion of chromosomal inversion homozygotes can still successfully reproduce under long-day conditions. This suggests that we might not fully understand the genetic mechanisms that determine the reproductive photoperiodism of *N. nepalensis* in Taiwan.

Here, we compared the fixation index (F_{ST}) of different reproduction photoperiodism phenotypes in the Mt. Yangming, Wulai, and Mt. Hehuan populations to find genetic variations related to photoperiodic reproduction. We used a random forest classifier to examine the actual effects of these variant genotypes on the

doi:10.6342/NTU202302705

photoperiodic reproduction trait. At the same time, we used population genetic structure analysis and phylogenetic trees to clarify the evolutionary history of Taiwan's *N. nepalensis* and the gene flow between populations. With this combined evidence, we hope this study can provide a possibility for the sources of adaptive genetic variation and the reasons for the formation of local adaptation, thus helping us to better understand the principles of biodiversity formation.

doi:10.6342/NTU202302705

2. Materials and Methods

2.1 Establishment and maintenance of buried insect populations in laboratories

To conduct breeding experiments indoors, we established laboratory populations for the various field populations of the *N. nepalensis*, including those from Mt. Yangming, Wulai, and Mt. Hehuan. We trapped wild *N. nepalensis* individuals from the aforementioned locations using hanging traps filled with rotten pork mince and brought them back to the lab for breeding to establish the first generation of laboratory populations, which we referred to as wild-types. Subsequent generations were called F1, F2, and so forth. All the populations were kept at a constant temperature of 17℃ and humidity of 83-100%. However, the Mt. Yangming and Wulai populations were raised in a 10L:14D light condition, while the Mt. Hehuan population was raised in a 14L:10D light condition. Each individual was separately housed in a transparent 320ml plastic box filled with fresh potting soil, and the soil humidity was checked weekly. As food, we provided superworms (*Zophobas morio*, Fabricius, 1776).

When carrying out routine breeding to maintain the populations, we randomly selected beetles from different parental generations for pairing. Each pair of beetles were placed in a transparent plastic breeding box (21cm x 13cm x 13cm) filled with

doi:10.6342/NTU202302705

10cm of potting soil. A fresh rat carcass weighing 75±7.5 grams was placed on the soil surface as a breeding resource. Approximately 14 days later, when the larvae finish off the corpse and leave the original position, we moved them into separate transparent plastic boxes filled with potting soil for pupation. Once they had completed metamorphosis, we recorded their sex and individually placed each adult into a 320ml transparent plastic box with potting soil.

2.2 Distinguishing reproductive photoperiodism phenotypes via breeding experiments

In order to compare genetic differences between individuals of different reproductive photoperiodism phenotypes, we use the reproductive performance of individuals under long-day (14L:10D) stimulus as the basis for distinguishing the reproductive photoperiodism phenotype of the individuals. In many species exhibiting reproductive diapause, the critical period for initiating reproductive diapause (sensitive period for diapause) typically falls within a few days after emergence. For instance, in *Drosophila montana*, the sensitive period is approximately within 3 to 9 days after emergence (Salminen & Hoikkala, 2013)、while in *Leptinotarsa decemlineata*, the sensitive period is within 5 days after emergence (de Kort, 1990). Therefore, we

doi:10.6342/NTU202302705

immediately transferred the larvae to a photoperiod environment of 14L:10D after they left the nest. We ensured that the duration of exposure to long-day conditions for our experimental individuals adequately covered this sensitive period. This precaution was adopted to ensure that our experimental conditions could effectively determine the diapause performance of these individuals.

In the breeding experiment, we randomly paired a male and female individual from the same population that came from different parents. Each pair was placed in a transparent plastic breeding box (21cm × 13cm × 13cm) filled with 10cm thick culture soil. We provided a fresh rat corpse weighing 75±7.5 grams placed on the soil surface as a breeding resource and observed the breeding status of each group after 14 days. In order to avoid misjudgment of the diapause state of the individual due to a single individual in the pair not being sexually mature leading to breeding failure, or the individual being physiologically mature but not breeding, we only defined the individuals from the experimental pairs where the corpse was not buried and there were no signs of hair removal or biting as seasonal breeders. We defined the experimental individuals who successfully bred and had larvae hatched as year-round breeders for further analysis.

doi:10.6342/NTU202302705

2.3 DNA extraction & whole–genome sequencing

After all individuals completed the breeding experiment, the culture soil attached to the individuals was removed with a clean brush. While still alive, they were placed in clean 1.5 ml centrifuge tubes and stored in a -80°C freezer. We used the MagPurix Tissue DNA Extraction Kit (ZP02004, Zinexts Life Science Corp., Taipei, Taiwan) and MagPurix 12 EVO (Zinexts Life Science Corp., Taipei, Taiwan) for DNA extraction, and the Novaseq 6000 system (Illumina, Inc., CA, USA) was used for paired-end sequencing with a read length of 150 bp.

2.3 Read processing and variants calling

After obtaining the original sequencing data, we used Fastqc v0.11.8 (Andrews. 2010) to check the quality of all samples, then used Trimmomatic v0.36 (Bolger et al., 2014) for sequence trimming (fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:36), then use bwa-mem2 v2.2.1 (Li & Durbin, 2009) to map all trimmed reads to the reference genome assembled by (Chen, 2019).

doi:10.6342/NTU202302705

For single nucleotide polymorphism (SNPs) calling, we overlap the SNPs set generated by GATK v4.4.0.0 (McKenna et al., 2010) with the set generated by Samtools v1.17 (Danecek et al., 2021) to make the detected SNPs more credible We used GATK HaplotypeCaller to generate the raw variation set and performed base quality score recalibration (BQSR) to obtain the recalibrated bam file. Then we used GATK and Bcftools mpileup in Samtools to perform SNPs calling again on this bam file. Finally, GATK was used to overlap the SNPs datasets generated by two different programs to generate the SNPs dataset for subsequent analysis. For structural variation (SV) calling, we also overlapped SV sets generated by different methods to produce a more credible SV set. We use GATK to intersect SV sets generated by DELLY v1.1.6 (Rausch et al., 2012), Manta v1.6.0 (Chen et al., 2016) and Lumpy-sv v0.3.1 (Layer et al., 2014) for subsequent analysis.

2.4 Population genetic structure and evolutionary history

Understanding the genetic structure and evolutionary history among all samples can help us better comprehend the gene flow between populations and the origin of genetic variation. Using PLINK v1.90b6.21 (Purcell et al., 2007) we calculated the linkage disequilibrium of SNPs from individuals from Malaysia, Vietnam, Sichuan,

doi:10.6342/NTU202302705

Fujian, Okinawa, Amami Oshima, Mt. Yangming, Wulai, Mt Lala, Mt. Hehuan, and Dahan Forest Road. SNPs with $r^2 < 0.3$ were provided to Admixture v1.3.0 (Alexander & Lange, 2011) for population genetic structure analysis with K=2~15. To inspect the evolutionary history of the Mt. Yangming, Wulai, and Mt. Hehuan populations, we utilized iqtree v2.2.2.3 (Nguyen et al., 2015). The whole-genome reconstructed phylogenetic tree employed the optimal model TVM+F+ASC+R4, the tree reconstructed with the inversion region utilized the optimal model TVM+F+ASC+R5. For the phylogenetic tree reconstructed with the haplotype of the inversion region, we employed whatshap v1.6 (Martin et al., 2016) for haplotype phasing, subsequently calculating the phylogenetic tree using the optimal model TVM+F+ASC+R6.

2.5 Identifying genetic variations regulating reproductive photoperiod phenotypes

To understand the sources of adaptive genetic variation, it is essential to identify genetic variants associated with adaptive traits. We utilized individuals from Mt. Yangming, Wulai, and Mt. Hehuan populations bred in the laboratory (Mt. Yangming: n = 19, Wulai seasonal breeder: n = 10, Wulai year-round breeder: n = 9, Mt. Hehuan: n = 10) in an attempt to discover related variations. As individuals with different reproductive photoperiodism phenotypes might be picked from different populations,

doi:10.6342/NTU202302705

we carried out comparisons among various population and phenotype combinations to sift out differences truly associated with phenotypes from genetic disparities between populations. Specifically, we grouped individuals with the same reproductive photoperiod phenotype from the same population (e.g., Wulai seasonal breeders as a group), and using SNPs and structural variation, set a sliding window size of 1KB to calculate the fixation index (F_{ST}) between different groups, selecting sites with $F_{ST} > 0.8$ as candidate loci. We then used BEDTools v2.31.0 (Quinlan & Hall, 2010) to intersect these candidate loci with genomic features (Chen, 2019), considering any candidate loci within 1KB of known gene fragments as gene-related loci. Finally, we considered these gene-related sites that show high F_{ST} value in multiple comparisons as key genes regulating reproductive photoperiodism phenotypes.

2.6 Relationship of genotype and reproductive photoperiodism phenotype

After identifying the key genes that regulate the reproductive photoperiod, we reviewed the genotypes of each individual at the key gene sites using our mapping results. Further, we named these genotypes based on their type of variation relative to the reference genome, thereby collating the gene pool of *N. nepalensis* in Taiwan at the key gene sites. We then compared the genotype of each individual at the key gene sites

doi:10.6342/NTU202302705

with their reproductive photoperiod phenotype, confirming the role of each gene in regulating the reproductive photoperiod. In addition, we used Python 3.11.3 to generate a random forest classifier to examine whether the key genes we identified could determine an individual's reproductive photoperiod. Specifically, we provided the genotype of each individual at the key genes, as well as information on their reproductive photoperiod phenotype and population for the creation of the random forest classifier. To enhance the reliability of the classifier, we also designed PCR primers for the key variant sites using primer3 v0.4.0 (Untergasser et al., 2012, Table 1), amplifying these segments, and then identifying the genotype of some individuals at the key genes using Sanger sequencing, thereby increasing the sample size for generating the classifier. During the development of the random forest classifier, we divided the individual data used to create the model into a 70% training set and a 30% test set, generating a classifier composed of 100 decision trees with max depth = 4. We assessed its accuracy to quantify the predictive and explanatory power of the key genes we identified for the reproductive photoperiod phenotype.

doi:10.6342/NTU202302705

2.7 Defining beneficial alleles

Having understood the relationship between the reproductive photoperiodism phenotype and the genotypes at each key variant site, we aim to further clarify the source of the alleles that contributed to the adaptation of *N. nepalansis* in Taiwan. To do so, we first need to select the beneficial alleles from the identified variants. Specifically, if an allele allows a population to express a reproductive photoperiod phenotype adapted to the local environment (for example, enabling the Mt. Hehuan population to display a year-round breeding phenotype), we would define such an allele as advantageous for that population.

2.8 Determining the source of genetic variation

To understand the source of the genetic variation that enables local adaptations among the various *N. nepalensis* populations in Taiwan, we consider the results of the phylogenetic tree reconstruction and the distribution of beneficial alleles among populations. Specifically, if a beneficial allele is present in both the local and ancestral populations, we consider the adaptation as arising from standing genetic variation. If the

doi:10.6342/NTU202302705

beneficial allele is found exclusively in the local population, we interpret this as an adaptation resulting from new mutation.

doi:10.6342/NTU202302705

3. Results

3.1 Population genetic structure of *N.nepalensis*

The results of population genetic structure analysis performed on the whole genome (Figure 3) show that individuals from Taiwan are all classified into the same cluster. Apart from the Vietnamese and Fujian, Sichuan population being classified into the same cluster, the other clusters are generally consistent with geographical distribution. This suggests that at the whole-genome level, there are certain differences in genetic composition among populations, and the differences among Taiwanese populations (Mt. Yangming, Wulai, Lalashan, Mt. Hehuan, and Dahanshan Forest Road) are relatively small.

We then carried out population genetic structure analysis on the aforementioned groups using the inversion region (Figure 4). The results show that the division of most clusters is also generally consistent with geographical distribution. However, the Taiwanese populations are no longer classified into the same cluster but align closely with the genotype of the inversion. Specifically, most homozygotes present clear classifications, while heterozygotes demonstrate complex classification situations.

doi:10.6342/NTU202302705

Combining the above two results, we believe that there is a certain degree of gene flow among Taiwanese populations, but greater genetic differences are shown within the inversion region.

3.2 Taiwan population is composed of multiple immigrations

Understanding the evolutionary history of *N. nepalensis* populations in Taiwan can help us clarify the origins of genetic variation. Here, using the whole-genome sequencing data from individuals in Malaysia, Vietnam, Sichuan, Fujian, Amami Oshima, Okinawa, Mt. Yangming, Wulai, Mt. Lala, Mt. Hehuan, and Dahan Forest Road, we reconstructed the phylogenetic tree of *N. nepalensis* (Figure 5). The results indicated a closer phylogenetic relationship among China (Sichuan and Fujian), Japan (Amami Oshima and Okinawa), and Taiwan (Mt. Yangming, Wulai, Mt. Lala, Mt. Hehuan, and Dahan Forest Road). Considering the flying capability of *N. nepalensis* and the fact that Taiwan is a young island, we inferred that the *N. nepalensis* populations in Taiwan likely originated from Eurasia, which was connected to Taiwan during the glacial period.

Since chromosomal inversions can reduce recombination, phylogenetic trees based on inversion region sequences can effectively help us understand the evolutionary

doi:10.6342/NTU202302705

history among populations with gene flow. Therefore, we subsequently reconstructed the haplotype phylogenetic tree of the Fujian and Taiwanese populations based on the inversion region (Figure 6). The results showed that both copies of all chromosomal inversion homozygotes in Taiwan were divided into the same clade, while other Taiwanese haplotypes were divided into another clade. We believe this indicates that the *N. nepalensis* population in Fujian might be the common ancestor of the Taiwanese populations, and the Taiwanese populations might have been formed by more than one immigration event. we believe the clade of chromosomal inversion homozygotes belongs to same immigration event (we call it A immigration here), while the clade of other Taiwanese individuals belongs to another immigration event (we call it B immigration here). This not only shows that chromosomal inversions were introduced into Taiwan more than once, but also shows that there are two different genetic compositions of chromosomal inversion fragments within the Taiwanese populations. In terms of the inversion region, Mt. Yangming belongs to the A immigration, while Mt. Hehuan belongs to the B immigration, and Wulai has some individuals from the A immigration and some from the B immigration.

3.3 Genetic variation in the regulation of the reproductive photoperiodism

doi:10.6342/NTU202302705

After understanding the evolutionary history of *N. nepalensis* in Taiwan, we then tried to identify genetic differences related to adaptive traits. Here, we calculated the F_{ST} values between different *N. nepalensis* populations and compared them with genomic features to identify genetic variations related to the reproductive cycle. We first calculated F_{ST} values using single nucleotide polymorphisms (SNPs). The results showed that in the comparison between the seasonal breeders and year-round breeders in Wulai, there were no sites with $F_{ST} > 0.8$ (Figure 7). This suggests that the differences in reproductive photoperiodism phenotypes may not be caused by SNPs, but by structural variations or other differences.

We further calculated F_{ST} using structural variation, and the results showed that a candidate site was found on chromosome 4 in the comparison between seasonal breeders and year-round breeders in Wulai (Figure 8). In the comparison between seasonal breeders in Mt. Yangming and year-round breeders in Mt. Hehuan, no candidate site was found on chromosome 4, and 153 candidate sites were concentrated within an approximately 13 MB region on chromosome 3, which is the inversion region (Figure 9). In the comparison between seasonal breeders in Wulai and year-round breeders in Mt. Hehuan, there were six candidate sites, five of which were in the inversion region, and another was on chromosome 4 (Figure 10). There were 8 candidate sites between seasonal breeders in Mt. Yangming and year-round breeders in

doi:10.6342/NTU202302705

Wulai (Figure 11), and 2 candidate sites between seasonal breeders in Mt. Yangming and seasonal breeders in Wulai, located outside the inversion region on chromosome 3 and on chromosome 4, respectively (Figure 12).

After identifying the candidate sites, we further intersected these sites with genomic features to determine if these sites were related to known genes. The results of the intersection showed that 25 candidate sites were related to gene sequences in the structural annotation (Table 2), 24 of which were located in the inversion region, and the other was on chromosome 4. Among the 24 candidate sites located in the inversion region, 23 sites only had high F_{ST} values in the comparison between seasonal breeders in Mt. Yangming and year-round breeders in Mt. Hehuan. Another candidate site (corresponding to the peroxisomal membrane protein PMP34) also showed high F_{ST} values in the comparison between seasonal breeders in Wulai and year-round breeders in Mt. Hehuan. This indicates that this gene may play a significant role in determining the reproductive photoperiodism of *N.nepalensis*. The candidate site on chromosome 4 (corresponding to the ceh-30 gene) also had high F_{ST} values in the comparisons between seasonal breeders in Wulai and Mt. Yangming, and between seasonal breeders and year-round breeders in Wulai.

doi:10.6342/NTU202302705

By combining the results of high F_{ST} value sites and intersections in each group, we believe that the chromosome 3 inversion, *PMP34*, and *ceh-30* are key variations that regulate the reproductive photoperiod of *N. nepalensis* in Taiwan.

3.4 Relationship between genotype and reproductive photoperiodism phenotype

To clarify how specifically the chromosome 3 inversion, *PMP34*, and *ceh-30* affect the reproductive photoperiodism phenotype of *N.nepalensis* in Taiwan (i.e., clarifying the relationship between the genotypes and phenotypes of the three key variations), we identified the genotypes of these key variations in all individuals with identified reproductive photoperiodism phenotypes using the results of read mapping. We named them based on their variation types relative to the reference genome to clarify the gene pool of key sites in Taiwan's *N.nepalensis* (Table 3). Next, we summarized the relationships between the genotypes, phenotypes, and populations of the key variations in these individuals and the genotypes identified by PCR (Table 4). The results show that the reproductive photoperiod of the Mt. Yangming and Mt. Hehuan populations is determined only by the genotypes of chromosome 3 inversion and *PMP34*. In contrast, the Wulai population is determined by the combined genotypes of chromosome 3

doi:10.6342/NTU202302705

inversion, *PMP34*, and *ceh-30*. This suggests that there is more than one genetic mechanism controlling the reproductive photoperiod in *N. nepalensis* in Taiwan. The results of using random forest model to classify the reproductive photoperiodism phenotype of individuals in (Table 4) further confirm this. Classification models that include population information as one of the data features perform better than those without population information (Table 5).

The results of the aforementioned phylogenetic tree show that the gene pool of the Wulai population includes the genetic composition of both the A and B immigration. We believe this is likely the reason why the regulatory mechanism of reproductive photoperiodism of Wulai population differs from those of the Mt. Yangming and Mt. Hehuan populations.

Therefore, based on the previous haplotype phylogenetic tree (Fig. 7), we divided the individuals in Taiwan into the A immigration type (Table 6) and the B immigration type (Table 7), and generated random forest models respectively. Since the sample size is reduced after data segmentation, we not only generated models with original data but also duplicated the data of Wulai to triple the number of samples, thus generating models with larger sample sizes. This helps stimulate the explanatory power and predictive power of each key gene on the reproductive photoperiodism after the actual increase in experimental samples.

doi:10.6342/NTU202302705

The results show that in the A immigration type model, the accuracy can be improved to 1.0 after increasing the sample size (original model = 0.91, duplicated data model = 1.0, Table 8), and *PMP34* and *ceh-30* both have certain importance in this model (Fig. 14). However, in the B immigration type model, whether the sample size is increased or not, the model's accuracy is not up to 1.0 (original model = 0.83, duplicated data model = 8.3, Table 9), and only *ceh-30* shows higher importance (Fig. 15). Based on the above results, we believe that the A immigration type's reproductive photoperiodism is mainly affected by the genotypes of *PMP34* and *ceh-30*, while the B immigration type may also be regulated by other sites with smaller effects apart from *ceh-30*.

3.5 Distribution of beneficial alleles among populations

Based on the relationships between genotypes, phenotypes, and populations in Table 4, we believe that the allele "5" on *PMP34* in the Mt. Yangming population is a beneficial allele that helps the population to display the seasonal breeding phenotype. On the other hand, for the Hehuan Mountain population, the alleles "D" and "N" on *PMP34* are beneficial alleles that enable it to display the seasonal breeding phenotype.

doi:10.6342/NTU202302705

We further sorted out the gene pool of the Fujian population on the key variations: chromosome 3 inversion, *PMP34*, and *ceh–30* (Table 2). The results show that the beneficial allele "5" of the Mt. Yangming population also exists in the Fujian population, while the beneficial alleles "D" and "N" of the Mt. Hehuan population do not exist in the Fujian population. Therefore, we believe that the Mt. Yangming population acquired the necessary alleles for adaptation from standing genetic variation, while the adaptation of the Mt. Hehuan population came from new mutations.

doi:10.6342/NTU202302705

4. Discussion

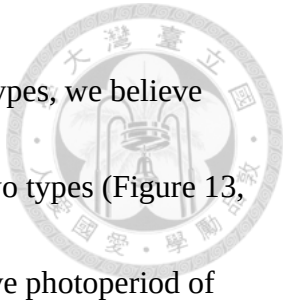
4.1 Both standing genetic variations and new mutations contribute to adaptation.

The distribution of beneficial alleles at the *PMP34* locus among different populations (Table 2) shows that the genetic variations aiding the adaptation of the Taiwanese *N. nepalensis* encompass both those unique to Taiwan and those shared with the Fujian population. This implies that both standing genetic variation and new mutations have jointly participated in local adaptation, revealing that even though standing genetic variation is considered to dominate in most adaptations (Barrett & Schluter, 2008; Hermisson & Pennings, 2005; Lai et al., 2019)the importance of new mutations in adaptation should not be overlooked.

4.2 Mutiple immigration events lead to multiple genetic mechanisms

In this study, we discovered that different populations regulate the same functional trait through different genetic mechanisms. Specifically, the reproductive photoperiod of the Wulai population is influenced by the genotype of ceh–30, while the Mt. Yangming and Mt. Hehuan populations are not (Table 4, Table 5). According to the

doi:10.6342/NTU202302705



results of models generated separately for the A and B immigration types, we believe that the relative importance of key genes may differ between these two types (Figure 13, Figure 14). That is, *PMP34* and *ceh-30* can determine the reproductive photoperiod of the A immigration type, while the influence of *PMP34* on the reproductive photoperiodism of the B immigration type is smaller, and it may be influenced by other genes in addition to *ceh-30*. Although we need to increase the actual experimental samples to verify the above points, we believe that these results suggest that the phenomenon of multiple mechanisms is likely caused by the two-time immigration. This also explains why in the original analysis, the Wulai population has a unique regulatory mechanism, as the Wulai population has both the A and B immigration genetic composition.

The results mentioned above highlight the influence of population formation history in shaping the current genetic structure and phenotypic diversity. As the process of population establishment is composed of a series of historical events, stochastic events (such as genetic drift, founder effect, etc.), and selection events, lacking understanding of this process may lead to misinterpretations of the reasons behind the formation of spatial heterogeneity in the current population's genotype or phenotype (Keller & Taylor, 2008). Take the invasion of *Butomus umbellatus* in North America, which exists in diploid and triploid forms as an example, the frequency of diploids in

doi:10.6342/NTU202302705

North America was observed to be significantly higher than the frequency in its native European population. The initial explanation was that sexual reproduction was favored in the new environment, thereby increasing the frequency of diploids in the introduced population. However, further genetic analysis found that the genetic diversity within the diploid population was extremely low. This indicates that the expansion of the diploid population in North America mainly relied on asexual reproduction rather than sexual reproduction, and the increase in diploid frequency in the introduced population could be caused by the founder effect, not a result of selection (Brown & Eckert, 2005).

4.3 Glacial cycle causes multiple immigration events

Through the phylogenetic tree reconstructed from the inversion region (Figure 6), we found that the *N. nepalensis* population in Taiwan is constituted of multiple immigrations from Fujian. Since Taiwan is a continental island and the phylogenetic tree shows that Taiwan's *N. nepalensis* originated from Fujian, we believe that glacial cycle is the cause of this phenomenon. Many studies have shown that species can disperse to nearby lands through land bridges formed during glacial period, such as the widespread distribution of *Nigella arvensis* in the Aegean Islands of Greece and the European continent (Comes et al., 2008)、the Bering Land Bridge (BLB) allowing the

doi:10.6342/NTU202302705

spread of caballine horses in the European continent and North America (Vershinina et al., 2021), and the dispersion history of tree frogs between Taiwan and Eurasia (Yu et al., 2020). Although we lack dating evidence to support the divergence time on the phylogenetic tree corresponding to the disappearance of the land bridge, considering that the flying ability of *N. nepalensis* is unlikely to allow it to cross the ocean for dispersion and that many actual cases show that glacial cycles allow organisms to migrate from continents to continental islands, viewing the periodic geographical isolation phenomenon caused by glacial cycles as the cause of multiple immigration events forming Taiwan's *N. nepalensis* seems to be the most reasonable explanation.

4.4 Cryptic diversity within Taiwan *N. nepalensis*

This study reveals that for the reproductive photoperiodism of *N. nepalensis* in Taiwan, there is more than one genetic regulatory mechanism, and this phenomenon is likely due to the complex history of population formation. Such diversity cannot be observed solely from the levels of morphology or behavior, which is somewhat similar to the concept of cryptic species: taxonomic groups that are morphologically indistinguishable but divergent in evolutionary history (Struck et al., 2018). Early studies mainly identified the existence of cryptic species through behavioral differences,

doi:10.6342/NTU202302705

such as distinguishing the cryptic species of fireflies through flashing patterns (Lloyd, 1969) and distinguishing two bat species through differences in echolocation and social calls (Russo & Jones, 2000), etc. With the development of DNA sequencing and genetic analysis technology, more and more studies are using genetic information to identify cryptic species, and their presence is more common than originally thought. For example, DNA barcoding revealed that what was previously considered to be the same butterfly species, *Astraptes fulgerator*, is actually a complex composed of at least ten cryptic species (Hebert et al., 2004). Research on cryptic species holds significant importance in conservation biology. One reason is that when a species is divided into different cryptic species, these cryptic species are not as common as originally thought, which may turn them into a combination of even more endangered species (Bickford et al., 2007), for example, the cryptic species *Pongo tapanuliensis*, estimated to number less than 800, was discovered within the threatened orangutan species *Pongo abelii* (Nater et al., 2017). On the other hand, different species may require different conservation strategies (Schönrogge et al., 2002), and different ecotypes of the same species (Engelhard et al., 2011), different populations (Visser et al., 2003) or even different populations within a metapopulation (Buchinger et al., 2022) may also be affected differently by environmental changes. Therefore, we believe that a thorough understanding of the functional traits and genetic diversity within a species is

doi:10.6342/NTU202302705

indispensable for vulnerability assessment and conservation of organisms. This also highlights the importance of clarifying the evolutionary history of adaptive genetic variation, as in this study, because this may help us discover the diversity hidden within species or even populations.

4.5 Historical climatic events and chromosomal inversion jointly shape the sympatric intraspecies diversity

Since gene flow within small spatial scales is often less constrained (Jensen et al., 2005; Van Strien et al., 2014; Wright, 1943), most past studies on genetic differentiation and phenotypic heterogeneity in space have focused on variations at larger spatial scales (Eckert et al., 2015; Richardson et al., 2014). However, for our main research populations (Mt. Yangming, Wulai, Mt. Hehuan), their geographical range spans less than 150 kilometers (from Mt. Yangming to Mt. Hehuan), and even within less than 30 kilometers, different genetic regulatory mechanisms for the reproductive photoperiod are manifested (Mt. Yangming & Wulai). The results of separately examining the genetic regulation mechanisms of the A and B immigration types reveal that the phenotypic diversity of the reproductive photoperiodism is very likely to have originated independently, and immigrated in at different times or places, resulting in

doi:10.6342/NTU202302705

two different genetic regulatory mechanisms in Taiwan populations. Meanwhile, the population genetic structure also reveals the contribution of chromosomal inversion to maintaining genetic differences (Figure 3, Figure 4). Since chromosomal inversion captures loci related to adaptation (*PMP34*), local adaptation of Taiwan *N.nepalensis* can still be maintained even under gene flow.

The intermittent geographic isolation caused by the glacial cycle and the presence of chromosomal inversion allows both standing genetic variation and new mutation to play important roles in local adaptation and increase the sympatric intraspecific genetic diversity and phenotypic diversity. In summary, our research results provide a possibility for the origin of adaptive genetic variation and the maintenance mechanism of local adaptation under gene flow.

doi:10.6342/NTU202302705

5. Conclusion

This study, through identifying genetic variations that assist in local adaptation in Taiwan *N. nepalensis* populations, found multiple genetic regulatory mechanisms in the reproductive photoperiod of Taiwan's *N. nepalensis*, and both standing genetic variation and new mutations helped Taiwan's burying beetles achieve local adaptation. The phylogenetic tree further revealed that Taiwan's *N. nepalensis* was composed of multiple migration events, and the population genetic structure showed that chromosomal inversion maintained genetic differences among populations. We believe that the aforementioned two reasons led to the maintenance of multiple genetic mechanisms assisting adaptation within the sympatric species. This study illustrates how complex population formation history and chromosomal inversion jointly shape intra-species genetic diversity and maintain local adaptation under gene flow, providing new insights into the mechanisms of biological adaptation.

doi:10.6342/NTU202302705

6. References

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 1–6. https://doi.org/10.1186/1471-2105-12-246/FIGURES/3

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/J.TREE.2007.09.008

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. https://doi.org/10.1016/J.TREE.2006.11.004

Bitter, M. C., Kapsenberg, L., Gattuso, J. P., & Pfister, C. A. (2019). Standing genetic variation fuels rapid adaptation to ocean acidification. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-13767-1

Bo-Fei, C. (2019). *Social evolution and genomic investigation of breeding adaptation in burying beetles* [Unpublished doctoral dissertation]. National Taiwan Normal University.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., Rudolf, V. H. W., Schreiber, S. J., Urban, M. C., & Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, *26*(4), 183–192. https://doi.org/10.1016/J.TREE.2011.01.009

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, *22*(3), 140–147. https://doi.org/10.1016/J.TREE.2006.11.002

Brown, J. S., & Eckert, C. G. (2005). Evolutionary increase in sexual and clonal reproductive capacity during biological invasion in an aquatic plant Butomus umbellatus (Butomaceae). *American Journal of Botany*, *92*(3), 495–502. https://doi.org/10.3732/AJB.92.3.495

Buchinger, T. J., Hondorp, D. W., & Krueger, C. C. (2022). Local diversity in phenological responses of migratory lake sturgeon to warm winters. *Oikos*, *2022*(6), e08977. https://doi.org/10.1111/OIK.08977

doi:10.6342/NTU202302705

Chaturvedi, A., Zhou, J., Raeymaekers, J. A. M., Czypionka, T., Orsini, L., Jackson, C. E., Spanier, K. I., Shaw, J. R., Colbourne, J. K., & De Meester, L. (2021). Extensive standing genetic variation from a small number of founders enables rapid adaptation in Daphnia. *Nature Communications*, *12*(1), 4306. https://doi.org/10.1038/s41467-021-24581-z

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, *32*(8), 1220–1222. https://doi.org/10.1093/BIOINFORMATICS/BTV710

Comes, H. P., Tribsch, A., & Bittkau, C. (2008). Plant speciation in continental island floras as exemplified by Nigella in the Aegean Archipelago. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3083–3096. https://doi.org/10.1098/RSTB.2008.0063

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

de Kort, C. A. D. (1990). Thirty-five years of diapause research with the Colorado potato beetle. *Entomologia Experimentalis et Applicata*, *56*(1), 1–13. https://doi.org/10.1111/J.1570-7458.1990.TB01376.X

Eckert, A. J., Maloney, P. E., Vogler, D. R., Jensen, C. E., Mix, A. D., & Neale, D. B. (2015). Local adaptation at fine spatial scales: an example from sugar pine (Pinus lambertiana, Pinaceae). *Tree Genetics and Genomes*, *11*(3), 1–17. https://doi.org/10.1007/S11295-015-0863-0/FIGURES/3

Engelhard, G. H., Ellis, J. R., Payne, M. R., Ter Hofstede, R., & Pinnegar, J. K. (2011). Ecotypes as a concept for exploring responses to climate change in fish assemblages. *ICES Journal of Marine Science*, *68*(3), 580–591. https://doi.org/10.1093/ICESJMS/FSQ183

Fuhrmann, N., Prakash, C., & Kaiser, T. S. (2023). Polygenic adaptation from standing genetic variation allows rapid ecotype formation. *ELife*, *12*. https://doi.org/10.7554/ELIFE.82824

Gomulkiewicz, R., Holt, R. D., & Barfield, M. (1999). The Effects of Density Dependence and Immigration on Local Adaptation and Niche Evolution in a Black-Hole Sink Environment. *Theoretical Population Biology*, *55*(3), 283–296. https://doi.org/10.1006/TPBI.1998.1405

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical

doi:10.6342/NTU202302705

skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14812–14817. https://doi.org/10.1073/PNAS.0406166101/SUPPL_FILE/06166SUPPAPPENDIX2.PDF

Hermisson, J., & Pennings, P. S. (2005). Soft SweepsMolecular Population Genetics of Adaptation From Standing Genetic Variation. *Genetics*, *169*(4), 2335–2352. https://doi.org/10.1534/GENETICS.104.036947

Holt, R. D., & Gomulkiewich, R. (1997). How Does Immigration Influence Local Adaptation? A Reexamination of a Familiar Paradigm. *Https://Doi.Org/10.1086/286005*, *149*(3), 563–572. https://doi.org/10.1086/286005

Hwang, W., & Shiao, S. F. (2011). Dormancy and the influence of photoperiod and temperature on sexual maturity in Nicrophorus nepalensis (Coleoptera: Silphidae). *Insect Science*, *18*(2), 225–233. https://doi.org/10.1111/J.1744-7917.2010.01356.X

Jensen, J. L., Bohonak, A. J., & Kelley, S. T. (2005). Isolation by distance, web service. *BMC Genetics*, *6*, 13. https://doi.org/10.1186/1471-2156-6-13

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Team, B. I. G. S. P. & W. G. A. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jump, A. S., Marchant, R., & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science*, *14*(1), 51–58. https://doi.org/10.1016/J.TPLANTS.2008.10.002

Kawecki, T. J. (2000). Adaptation to marginal habitats: contrasting influence of the dispersal rate on the fate of alleles with small and large effects. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1450), 1315–1320. https://doi.org/10.1098/RSPB.2000.1144

Keller, S. R., & Taylor, D. R. (2008). History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, *11*(8), 852–866. https://doi.org/10.1111/J.1461-0248.2008.01188.X

Lai, Y. T., Yeung, C. K. L., Omland, K. E., Pang, E. L., Hao, Y., Liao, B. Y., Cao, H. F., Zhang, B. W., Yeh, C. F., Hung, C. M., Hung, H. Y., Yang, M. Y., Liang, W., Hsu, Y. C., Yao, C. Te, Dong, L., Lin, K., & Li, S. H. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 2152–2157.

doi:10.6342/NTU202302705

https://doi.org/10.1073/PNAS.1813597116/SUPPL_FILE/PNAS.1813597116.SAPP.PDF

Lande, R., & Shannon, S. (1996). THE ROLE OF GENETIC VARIATION IN ADAPTATION AND POPULATION PERSISTENCE IN A CHANGING ENVIRONMENT. *Evolution*, *50*(1), 434–437. https://doi.org/10.1111/j.1558-5646.1996.tb04504.x

Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for structural variant discovery. *Genome Biology*, *15*(6), 1–19. https://doi.org/10.1186/GB-2014-15-6-R84/FIGURES/8

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Lloyd, J. E. (1969). Flashes of Photuris Fireflies: Their Value and Use in Recognizing Species. *The Florida Entomologist*, *52*(1), 29. https://doi.org/10.2307/3493705

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544. https://doi.org/10.1016/J.TREE.2019.02.008

Martin, M., Patterson, M., Garg, S., Fischer, S. O., Pisanti, N., Klau, G. W., Schöenhuth, A., & Marschall, T. (2016). WhatsHap: fast and accurate read-based phasing. *BioRxiv*, 085050. https://doi.org/10.1101/085050

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/GR.107524.110

Nater, A., Mattle-Greminger, M. P., Nurcahyo, A., Nowak, M. G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T. B., Roos, C., Lameira, A. R., Wich, S. A., Askew, J., Davila-Ross, M., Fredriksson, G., de Valles, G., Casals, F., Prado-Martinez, J., Goossens, B., … Krützen, M. (2017). Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current Biology*, *27*(22), 3487-3498.e10. https://doi.org/10.1016/J.CUB.2017.09.047

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/MOLBEV/MSU300

doi:10.6342/NTU202302705

Paaby, A. B., & Rockman, M. V. (2014). Cryptic genetic variation: Evolution's hidden substrate. In *Nature Reviews Genetics* (Vol. 15, Issue 4, pp. 247–258). Nature Publishing Group. https://doi.org/10.1038/nrg3688

Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, *13*(6), 288–294. https://doi.org/10.1016/J.TPLANTS.2008.03.004

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, *28*(18), i333–i339. https://doi.org/10.1093/BIOINFORMATICS/BTS378

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, *29*(3), 165–176. https://doi.org/10.1016/J.TREE.2014.01.002

Russo, D., & Jones, G. (2000). The two cryptic species of Pipistrellus pipistrellus (Chiroptera: Vespertilionidae) occur in Italy: Evidence from echolocation and social calls. *Mammalia*, *64*(2), 187–197. https://doi.org/10.1515/MAMM.2000.64.2.187/MACHINEREADABLECITATION/RIS

Salminen, T. S., & Hoikkala, A. (2013). Effect of temperature on the duration of sensitive period and on the number of photoperiodic cycles required for the induction of reproductive diapause in Drosophila montana. *Journal of Insect Physiology*, *59*(4), 450–457. https://doi.org/10.1016/J.JINSPHYS.2013.02.005

Sanford, E., Holzman, S. B., Haney, R. A., Rand, D. M., & Bertness, M. D. (2006). LARVAL TOLERANCE, GENE FLOW, AND THE NORTHERN GEOGRAPHIC RANGE LIMIT OF FIDDLER CRABS. *Ecology*, *87*(11), 2882–2894. https://doi.org/10.1890/0012-9658

Schönrogge, K., Barr, B., Wardlaw, J. C., Napper, E., Gardner, M. G., Breen, J., Elmes, G. W., & Thomas, J. A. (2002). When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnean Society*, *75*(3), 291–300. https://doi.org/10.1046/J.1095-8312.2002.00019.X

doi:10.6342/NTU202302705

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140. https://doi.org/10.1111/J.1365-294X.2011.05350.X

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K. H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution*, *33*(3), 153–163. https://doi.org/10.1016/J.TREE.2017.11.007

Tsai, H. Y., Rubenstein, D. R., Fan, Y. M., Yuan, T. N., Chen, B. F., Tang, Y., Chen, I. C., & Shen, S. F. (2020). Locally-adapted reproductive photoperiodism determines population vulnerability to climate change in burying beetles. *Nature Communications 2020 11:1*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15208-w

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115. https://doi.org/10.1093/NAR/GKS596

Van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2014). Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity 2015 114:1*, *114*(1), 27–37. https://doi.org/10.1038/hdy.2014.62

Vershinina, A. O., Heintzman, P. D., Froese, D. G., Zazula, G., Cassatt-Johnstone, M., Dalén, L., Der Sarkissian, C., Dunn, S. G., Ermini, L., Gamba, C., Groves, P., Kapp, J. D., Mann, D. H., Seguin-Orlando, A., Southon, J., Stiller, M., Wooller, M. J., Baryshnikov, G., Gimranov, D., … Shapiro, B. (2021). Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. *Molecular Ecology*, *30*(23), 6144–6161. https://doi.org/10.1111/MEC.15977

Visser, M. E., Adriaensen, F., Van Balen, J. H., Blondel, J., Dhondt, A. A., Van Dongen, S., Du Feu, C., Ivankina, E. V., Kerimov, A. B., De Laet, J., Matthysen, E., McCleery, R., Orell, M., & Thomson, D. L. (2003). Variable responses to large-scale climate change in European Parus populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1513), 367–372. https://doi.org/10.1098/RSPB.2002.2244

Wright, S. (1943). ISOLATION BY DISTANCE. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/GENETICS/28.2.114

Yu, G. H., Du, L. N., Wang, J. S., Rao, D. Q., Wu, Z. J., & Yang, J. X. (2020). From mainland to islands: colonization history in the tree frog Kurixalus (Anura: Rhacophoridae). *Current Zoology*, *66*(6), 667–675. https://doi.org/10.1093/CZ/ZOAA023

doi:10.6342/NTU202302705

7. Figures

Figure 1: Map of *N. nepalensis* populations. The map shows all the populations we have collected beetles. Inversion of chromosome 3 has been identified among Fujian and Taiwan populations.

doi:10.6342/NTU202302705

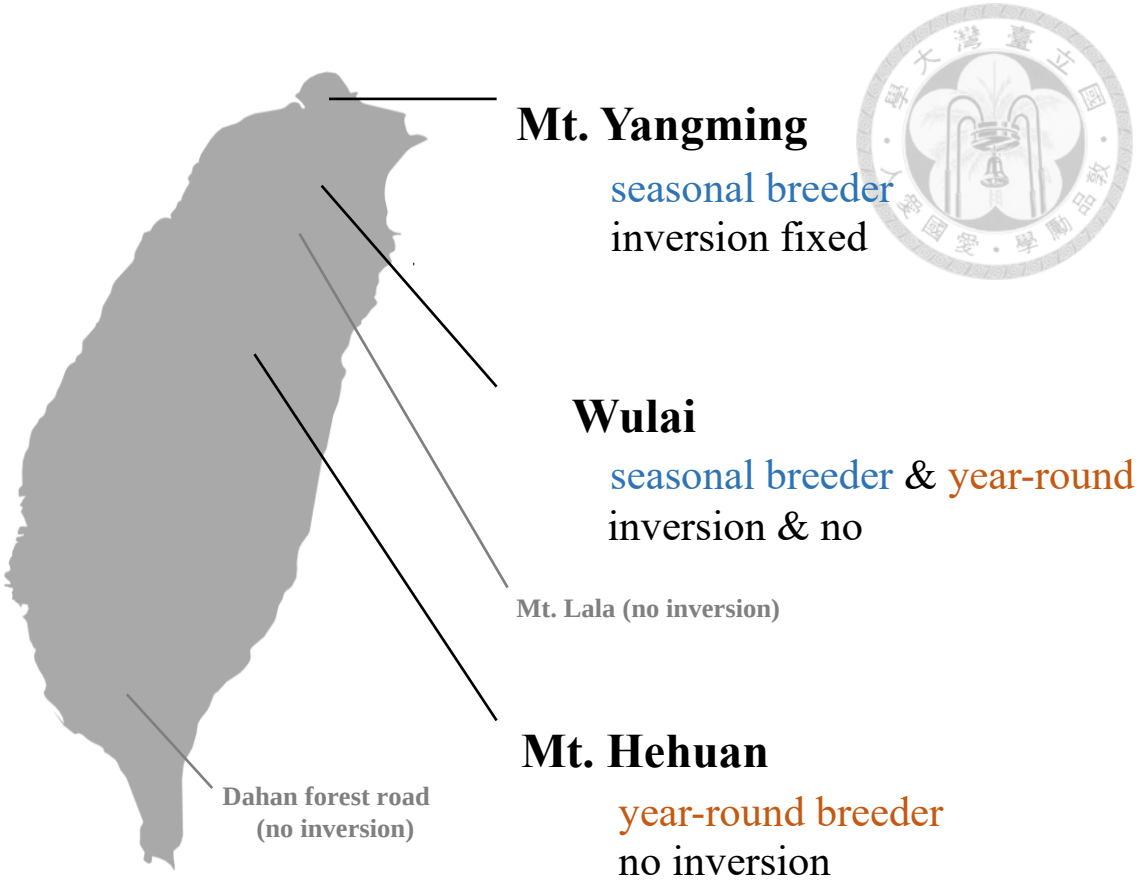


Figure 2: Chromosome inversion and reproductive photoperiodism of Taiwan *N. nepalensis* populations.

The location of Taiwan populations, and the chromosome 3 inversion and reproductive photoperiodism of Mt.Yangming, Wulai, and Mt.Hehuan populations. For Mt.Yangming, chromosome 3 inversion is fixed in the population, and all the beetles are seasonal breeder, which diapause in summer. For Mt.Hehuan, no inversion exist and all the beetles are year-round breeder. For Wulai, both haplotypes and both phenotypes exist in the population.

doi:10.6342/NTU202302705

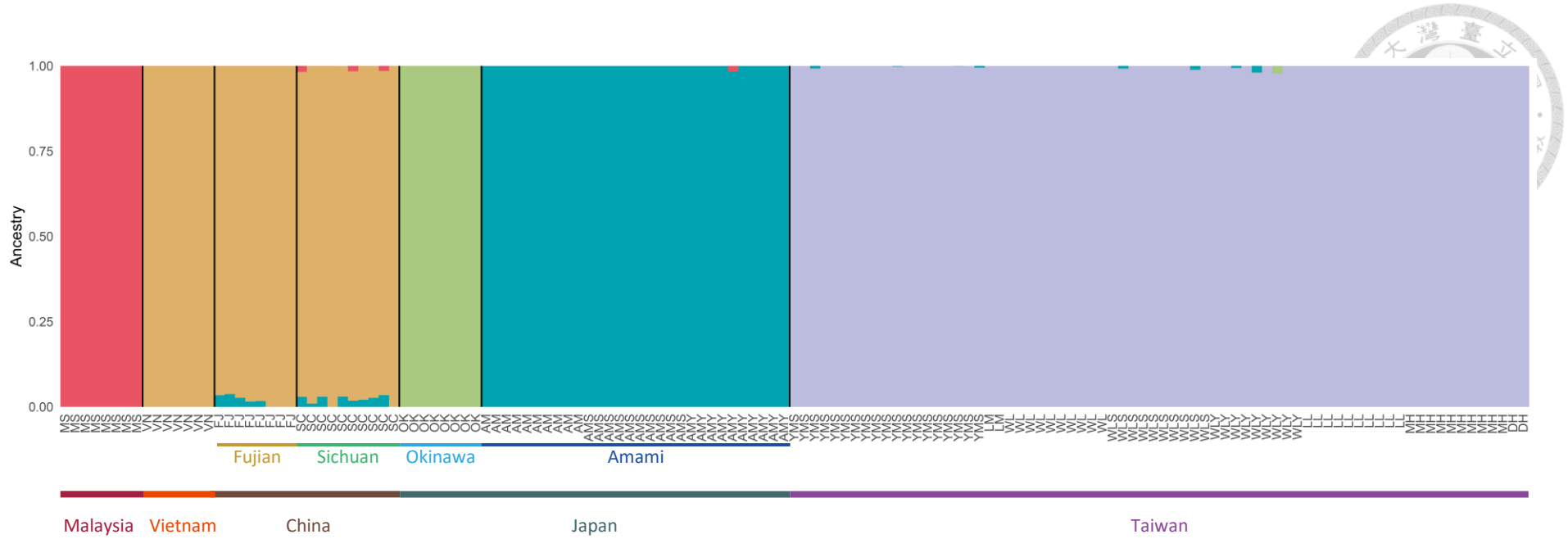


Figure 3: Population structure of *N.nepalensis*.

Structure analysis with Admixture (K = 5). Each individual is depicted as a vertical bar divided into differently colored genetic clusters. The length of each color segment is proportional to the probability of being assigned to that particular cluster. Most clusters largely corresponded with geographical divisions. The results show that all individuals from Taiwan are classified within the same cluster. Apart from the populations of Vietnam, Fujian, and Sichuan which are grouped together into a single cluster, the partitioning of the remaining clusters largely aligns with geographical distribution. This suggests that, at the whole-genome scale, there are certain differences in genetic composition between populations from different regions, and the differences among various groups within Taiwan are relatively small.

doi:10.6342/NTU202302705

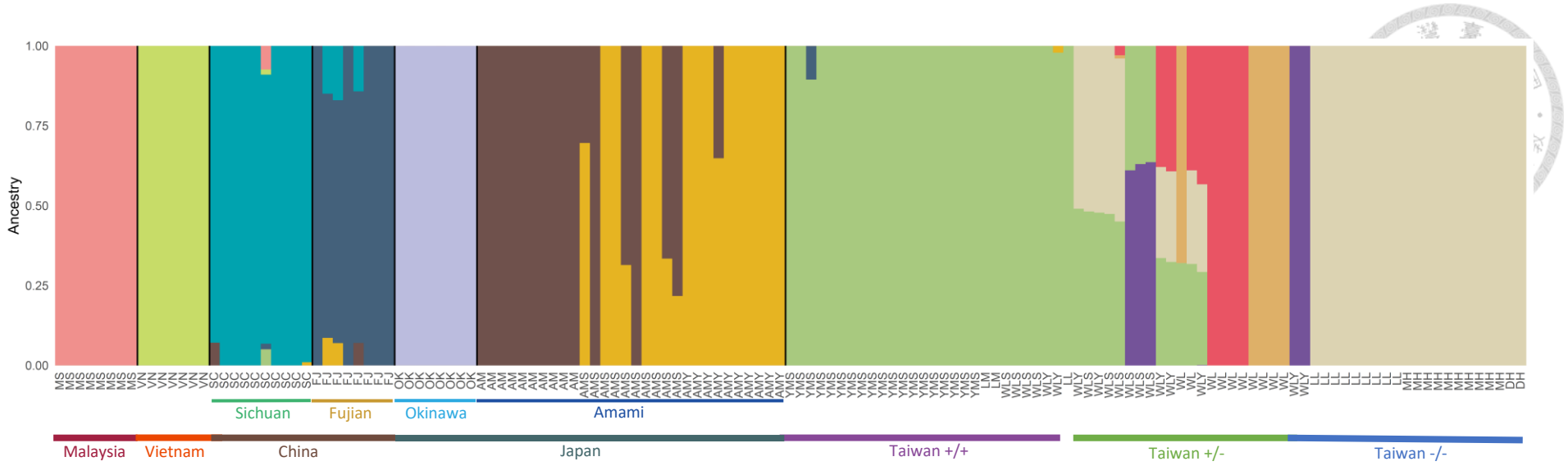


Figure 4: Population structure of the inversion region of *N.nepalensis*.

Structure analysis with Admixture (K = 12). Each individual is depicted as a vertical bar divided into differently colored genetic clusters. The length of each color segment is proportional to the probability of being assigned to that particular cluster. Most clusters largely corresponded with geographical divisions. However, the Taiwan population was classified into different clusters, which generally align with the genotype of the inversion. Most homozygotes exhibited clear classifications, while heterozygotes presented complex classification situations. ”+”represent the inversion haplotype, and “-“ represent the no inversion haplotype.

doi:10.6342/NTU202302705

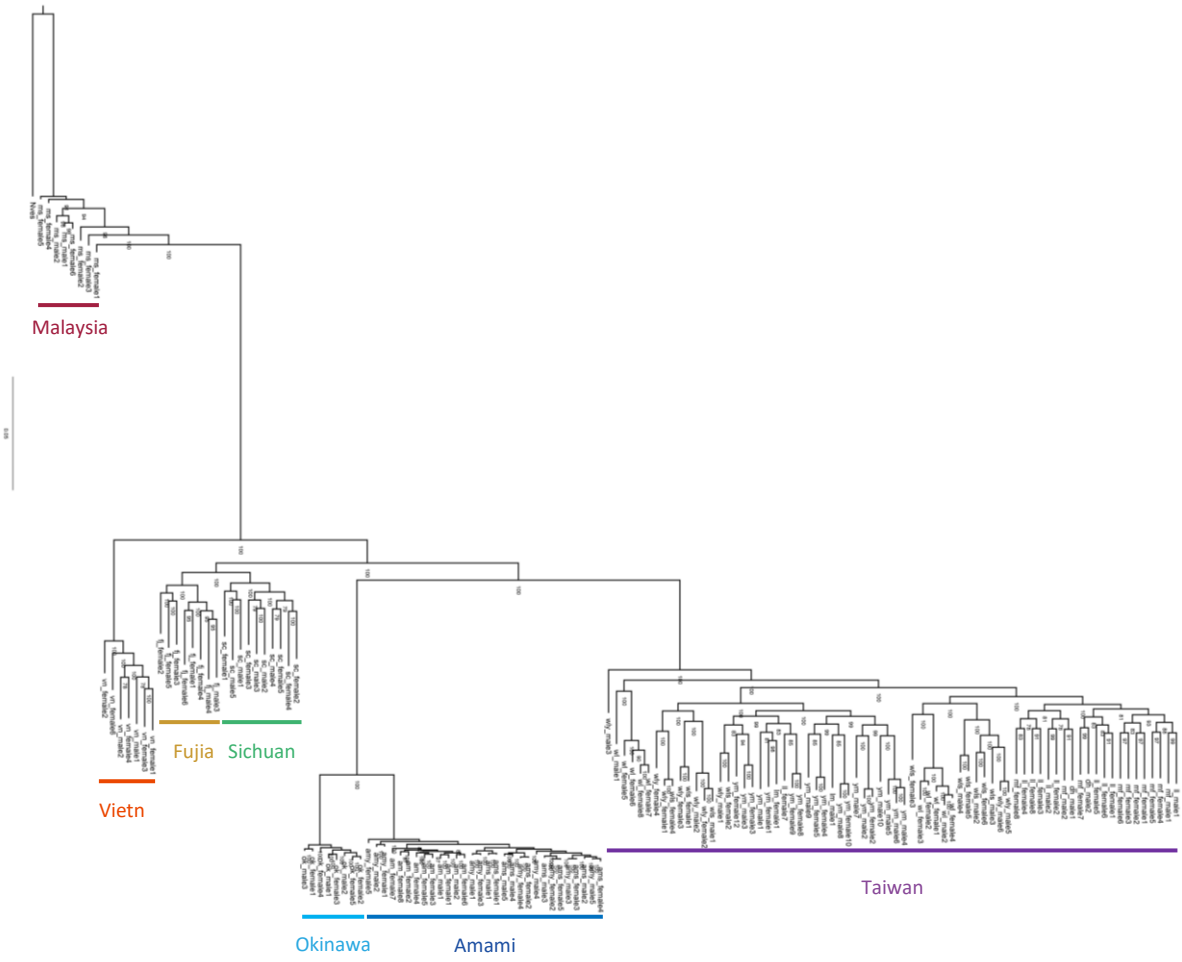


Figure 5: Phylogenetic tree of *N.nepalensis*.

The phylogenetic tree using *N. vespilloides* as the outgroup for Malaysia, Vietnam, Fujian, Sichuan, Okinawa, Amami, Mt.Yangming, Wulai, Mt.Lala, Mt. Hehuan, and Dahan Forest Road. This result is similar to the structure analysis, indicating that at the whole-genome scale, the genetic differences and divergence among various Taiwanese groups are not significant.

doi:10.6342/NTU202302705

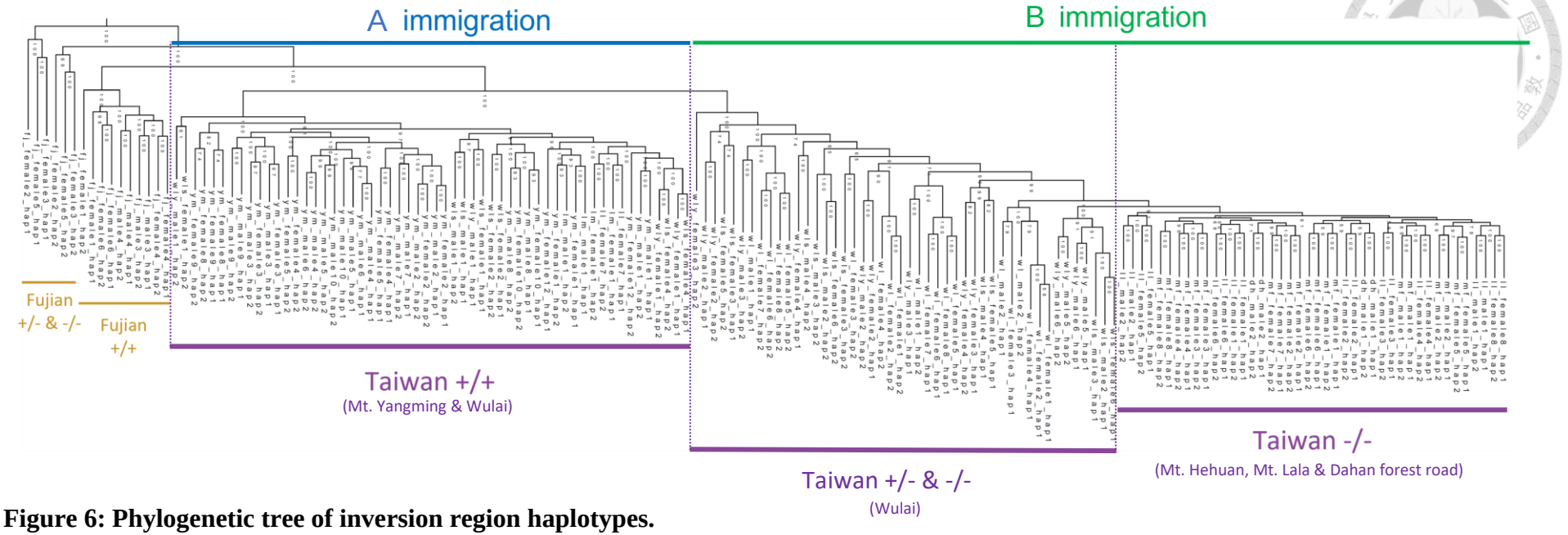


Figure 6: Phylogenetic tree of inversion region haplotypes.

The green and blue horizontal lines above indicate the haplotypes with genotypes of A and B immigration we inferred. All haplotype pairs belonging to the same individual are within the same area divided by the purple dashed line, and yellow and purple horizontal lines, and the texts underneath indicated the inversion genotypes of the individuals, the “+/+” represent inversion homozygotes, “-/-” represent non-inversion homozygotes, and “+/-” represent heterozygotes. This indicates that both the inversion haplotype and the non-inversion haplotype flowed into Taiwan during the A immigration, while only the inversion haplotype was involved in the B immigration, and there is a certain degree of genetic variation between the inverted haplotypes from "+/-" and "+/+" individuals.

doi:10.6342/NTU202302705

Seasonal breeder (Wulai) vs. Year-round breeder (Wulai)

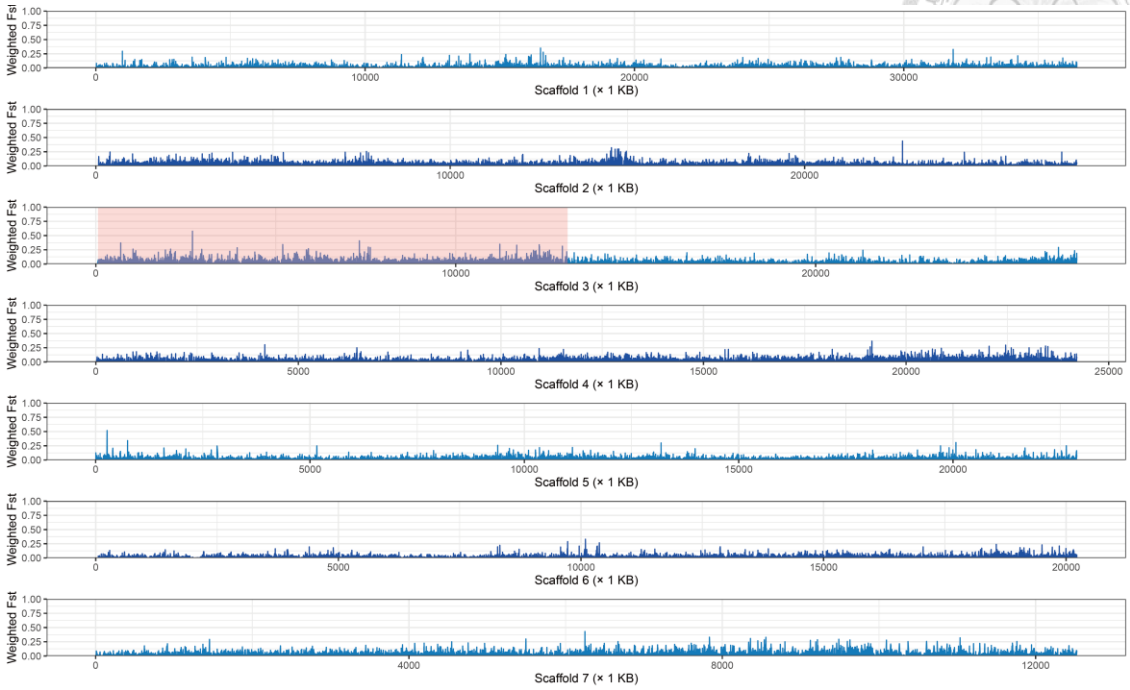


Figure 7: F_{ST} of SNPs in seasonal breeder (Wulai) vs. year-round breeder (Wulai). F_{ST} of SNPs in seasonal breeder (Wulai, n = 10) vs. year-round breeder (Wulai, n = 9). The pink area in Scaffold 3 represents the inversion region.

doi:10.6342/NTU202302705

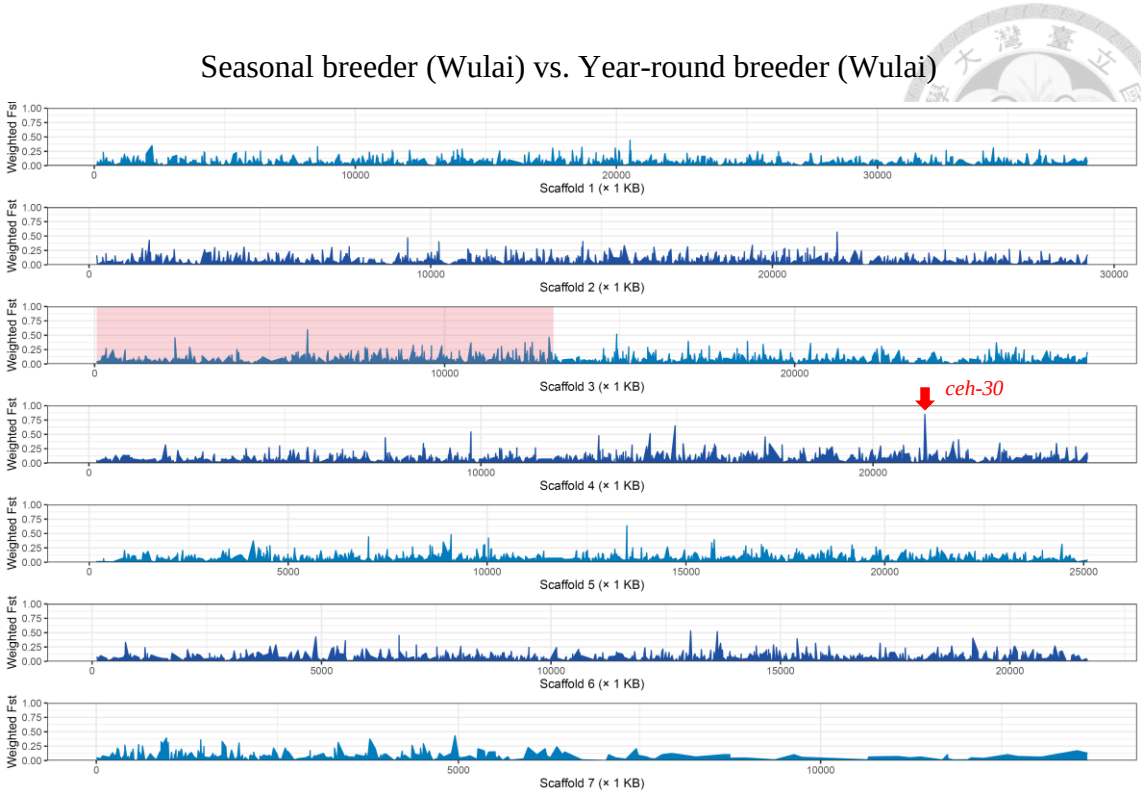


Figure 8: F_{ST} of structural variation in seasonal breeder (Wulai) vs. year-round breeder (Wulai).

F_{ST} of structural variation in seasonal breeder (Wulai, n = 10) vs. year-round breeder (Wulai, n = 9). The pink area in scaffold 3 represents the inversion region, red arrows point out the sites with $F_{ST} > 0.8$, and the text next to the arrow indicates the name of the key gene that the site overlaps with.

doi:10.6342/NTU202302705

Seasonal breeder (Mt.Yangming) vs. Year-round breeder (Mt.Hehuan)

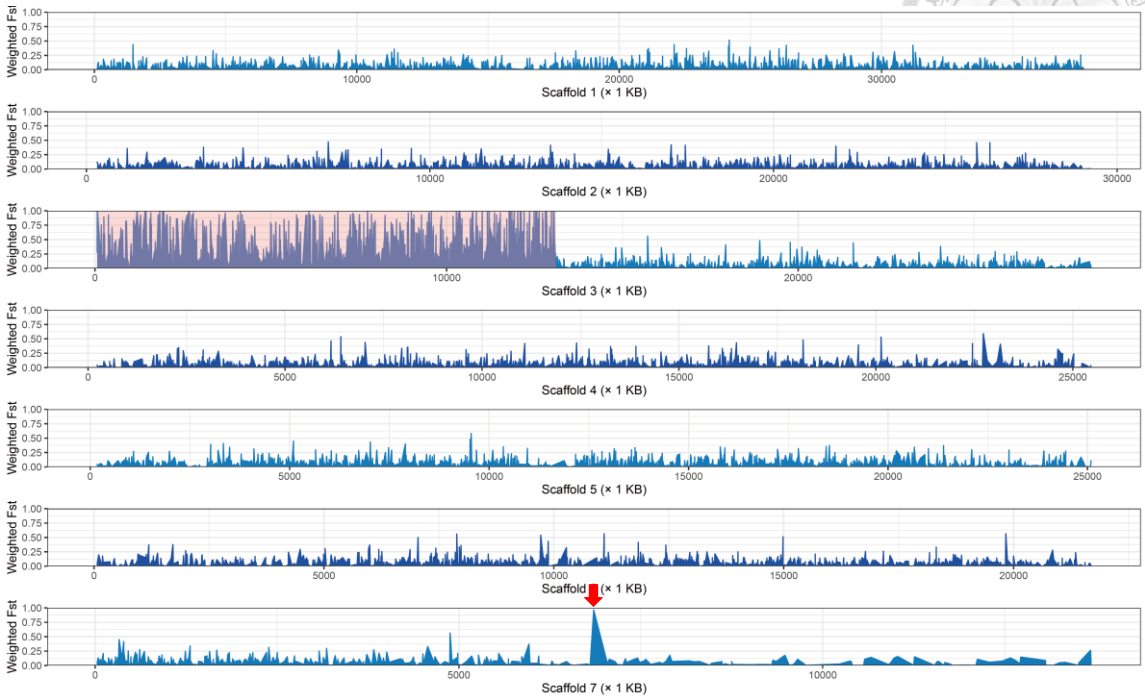


Figure 9: F_{ST} of structural variation in seasonal breeder (Mt.Yangming) vs. year-round breeder (Mt.Hehuan).

F_{ST} of structural variation in seasonal breeder (Mt.Yangming, n = 19) vs. year-round breeder (Mt.Hehuan, n = 10). The pink area in scaffold 3 represents the inversion region, red arrows point out the sites with $F_{ST} > 0.8$. There are actually 24 genes overlapped with the high F_{ST} site in the inversion region, and no gene overlapped with the site at scaffold 7.

doi:10.6342/NTU202302705

Seasonal breeder (Wulai) vs. Year-round breeder (Mt.Hehuan)

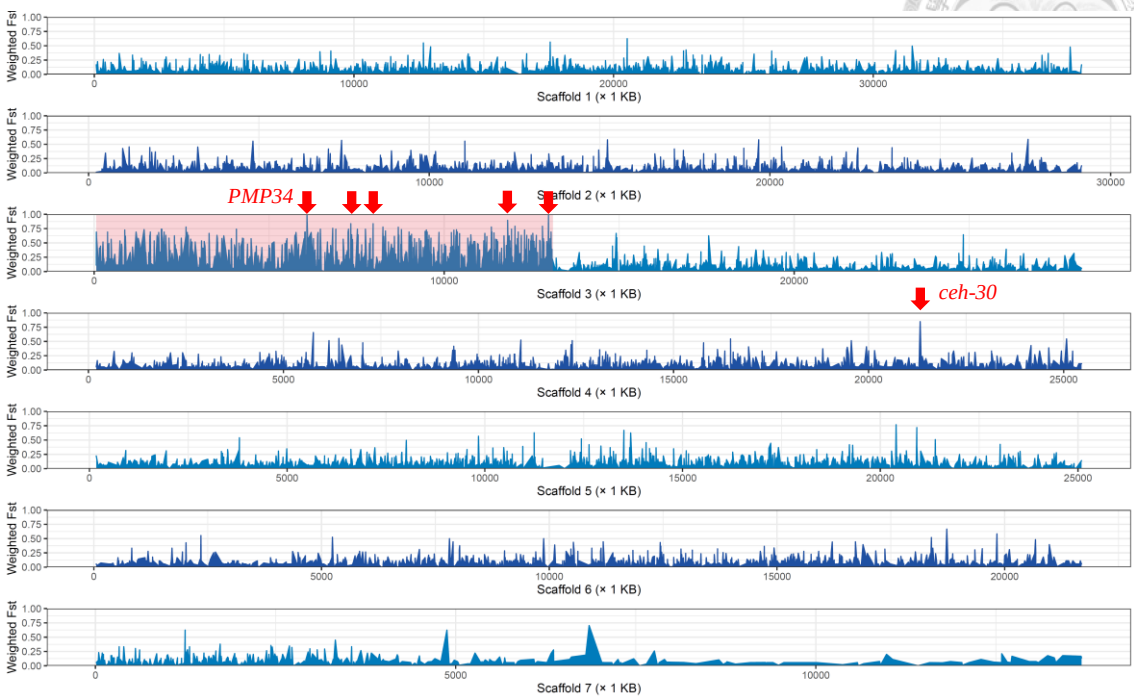


Figure 10: F_{ST} of structural variation in seasonal breeder (Wulai) vs. year-round breeder (Mt.Hehuan).

F_{ST} of structural variation in seasonal breeder (Wulai, n = 10) vs. year-round breeder (Mt.Hehuan, n = 10). The pink area in scaffold 3 represents the inversion region, red arrows point out the sites with F_{ST} > 0.8,and the text next to the arrow indicates the name of the key gene that the site overlaps with.

doi:10.6342/NTU202302705

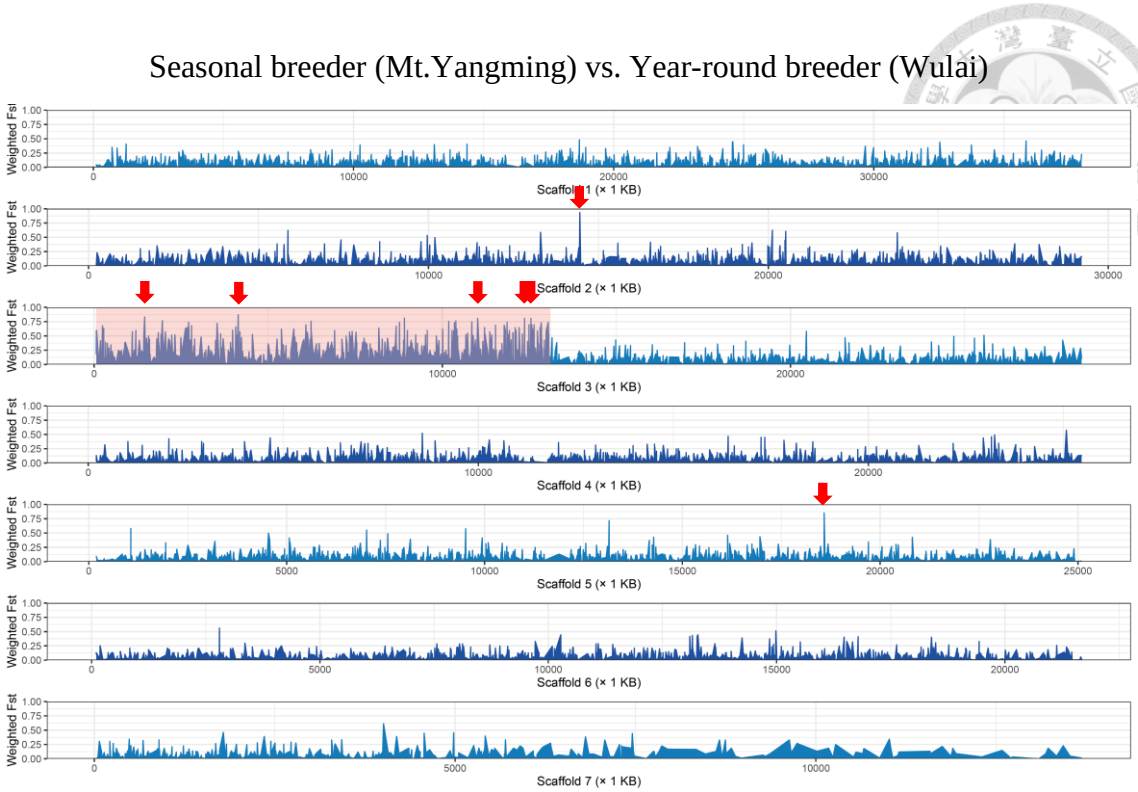


Figure 11: F_{ST} of structural variation in seasonal breeder (Mt.Yangming) vs. year-round breeder (Wulai).

F_{ST} of structural variation in seasonal breeder (Mt.Yangming, n = 19) vs. year-round breeder (Wulai, n = 9). The pink area in scaffold 3 represents the inversion region, red arrows point out the sites with $F_{ST} > 0.8$.

doi:10.6342/NTU202302705

Seasonal breeder (Mt.Yangming) vs. Seasonal breeder (Wulai)

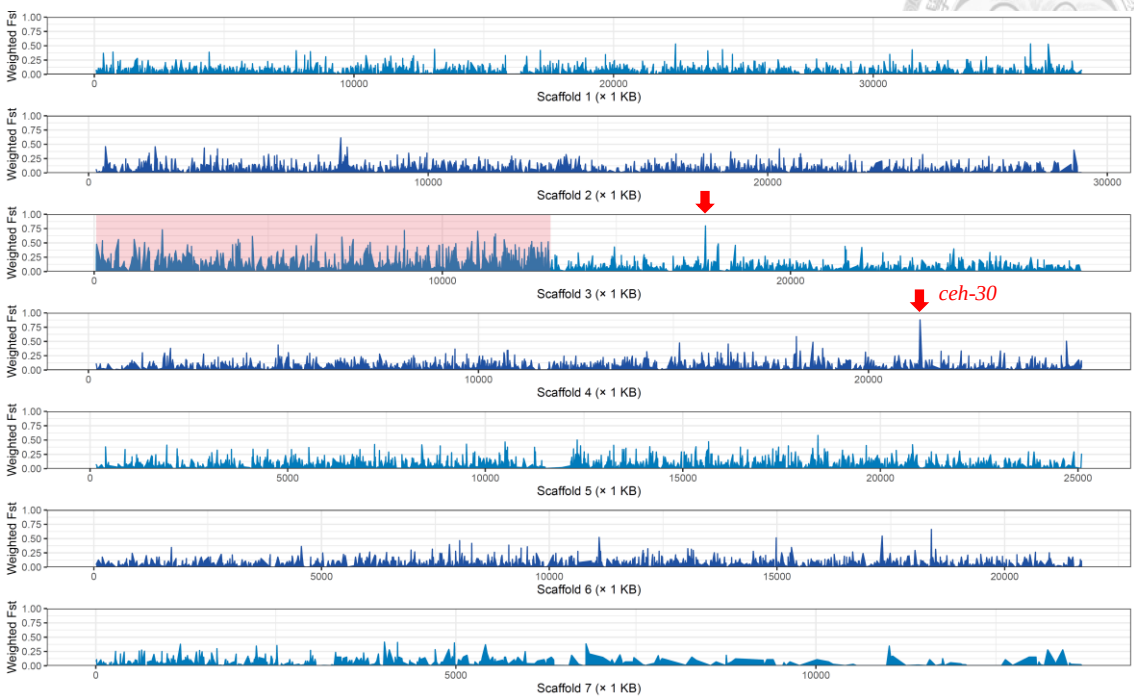


Figure 12: F_{ST} of structural variation in seasonal breeder (Mt.Yangming) vs. seasonal breeder (Wulai).

F_{ST} of structural variation in seasonal breeder (Mt.Yangming, n = 19) vs. seasonal breeder (Wulai, n = 10). The pink area in scaffold 3 represents the inversion region, red arrows point out the sites with $F_{ST} > 0.8$, and the text next to the arrow indicates the name of the key gene that the site overlaps with. The result indicated that gene *ceh-30* might function differently in Mt. Yangming and Wulai populations.

doi:10.6342/NTU202302705

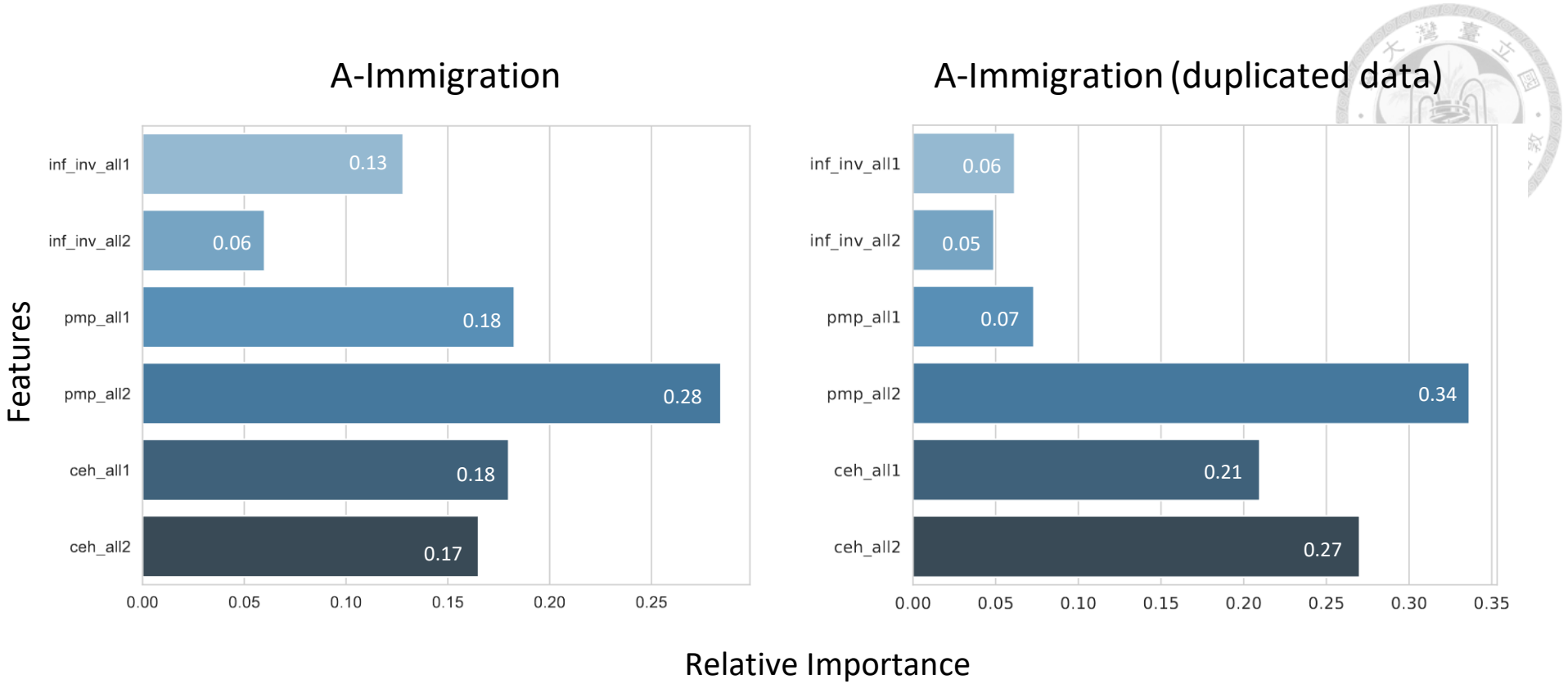


Figure 13: Feature importance of random forest model of A-immigration type.

The relative importance of different features in each model. The "duplicated data" model is generated by duplicating the Wulai data twice compared to the original model.

doi:10.6342/NTU202302705

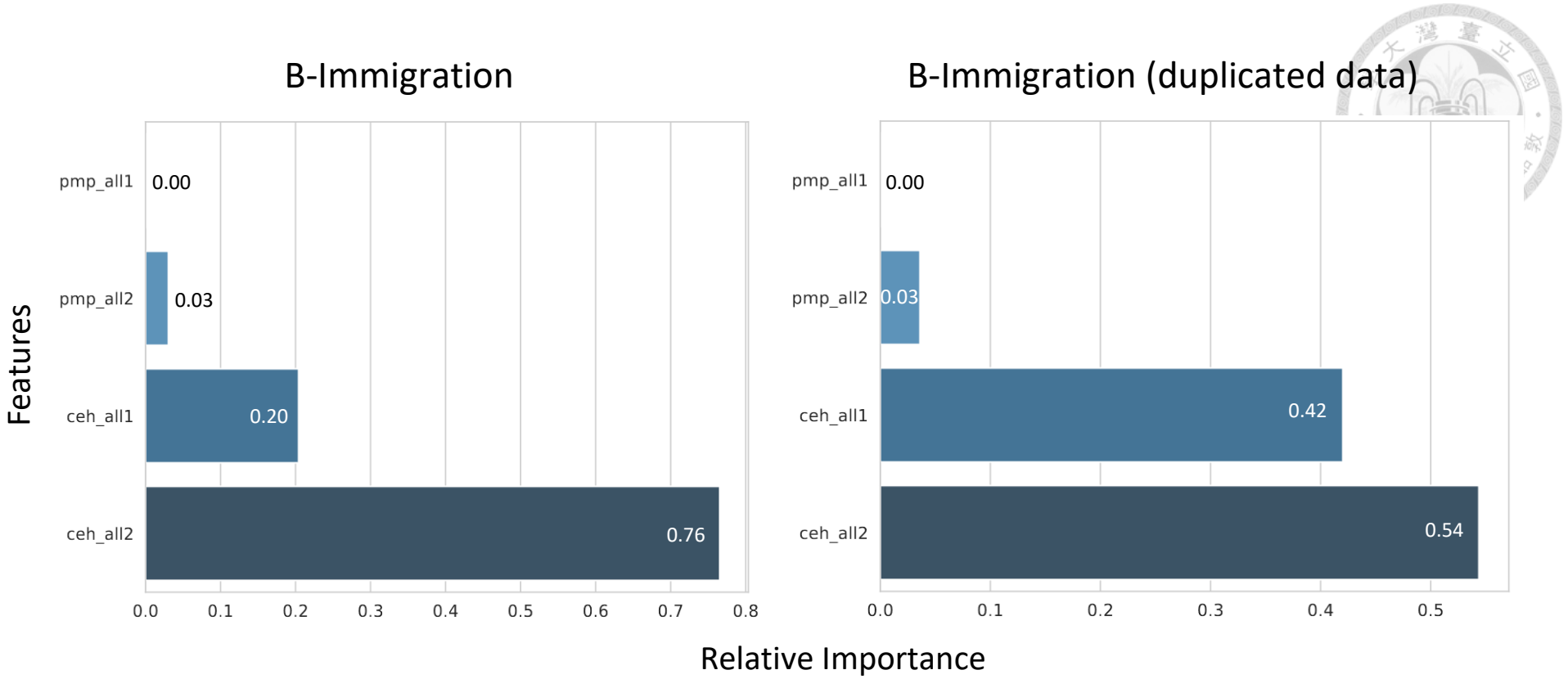


Figure 14: Feature importance of random forest model of B-immigration type.

The relative importance of different features in each model. The "duplicated data" model is generated by duplicating the Wulai data twice compared to the original model. We didn't use the inversion type as a feature since all individuals we used have the same type of inversion.

doi:10.6342/NTU202302705

8. Tables

Table 1: PCR primer for amplifying key loci sequence

Loci	Primer	Sequence (5’ to 3’)
PMP34	forward	CAATAGCATCCGCACCTTTT
	reverse	GCTTCAGTTTCATATGACCACCA
ceh-30	forward	TGGCCAATGAACAGAAAAGA
	reverse	AAATCTATAAAACGTGCTTTCAACTAA

doi:10.6342/NTU202302705

Table 2: Genes overlapped with high F_{ST} loci.

Gene Name
homeobox protein ceh-30
high affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8A isoform X2
titin isoform X1
hemicentin-2-like
40S ribosomal protein S20
protein Wnt-5b-like isoform X1
sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 isoform X1
riboflavin kinase
UTP--glucose-1-phosphate uridylyltransferase isoform X2
peroxisomal membrane protein PMP34
neurexin-1 isoform X2
cytochrome c oxidase subunit 5B, mitochondrial-like
putative fatty acyl-CoA reductase CG5065 isoform X1
protein SDE2 homolog
potassium voltage-gated channel subfamily H member 6-like isoform X4
E3 ubiquitin-protein ligase AMFR-like
semaphorin-1A isoform X1

doi:10.6342/NTU202302705

Table 3: Gene pool of Fujian and Taiwan population at key loci

Population	inversion	PMP34	ceh-30
Fujian	+*	3	1b
			1f
		5*	2*
	-*	6	DS*
		8	DL*
			W*
			S2*
			T2*
Mt. Yangming	+*	5*	2*
		D	DS*
			W*
			S1*
Wulai	+*	5*	2*
			DS*
		D	DL*
	-*	N	W*
			S2*
			T2*
			new
Mt. Hehuang	-*	D	DS*
			DL*
		N	W*
			S1*

The gene pool of Fujian and Taiwan populations at key loci, "*" represent alleles that appear in both Fujian and Taiwan population. The designations of alleles are named according to their variations compared to the reference genome. "+": inversion, "-": wild-type, "D": duplication, "N": wild-type, "1b": 1 SNPs (closer to 3'end than 1f), "1f": 1 SNPs (closer to 5'end than 1b), "2": 2 SNPs (1f & 1b), DS and DL: shorter and longer deletion, W: wildtype, "S1" & "S2": different types of substitution, "T2": translocation and 2SNPs, "new": A type of deletion different from DS and DL.

doi:10.6342/NTU202302705

Table 4: Reproductive photoperiodism phenotype and genotype at key loci of each individual

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*		Genotyping Method
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	
ym_male8	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_female10	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_female9	YM	seasonal	+(Y)	+(Y)	5	5	W	W	sequenced by Illumina
ym_female8	YM	seasonal	+(Y)	+(Y)	5	5	W	W	sequenced by Illumina
ym_female1	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_male1	YM	seasonal	+(Y)	+(Y)	5	5	S1	DS	sequenced by Illumina
ym_female12	YM	seasonal	+(Y)	+(Y)	5	5	W	2	sequenced by Illumina
ym_male9	YM	seasonal	+(Y)	+(Y)	5	5	W	W	sequenced by Illumina
ym_male3	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_female3	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_female5	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_female4	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_male7	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_male2	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_female2	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_male10	YM	seasonal	+(Y)	+(Y)	5	5	W	S1	sequenced by Illumina
ym_male5	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_male6	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_male4	YM	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
ym_female6	YM	seasonal	+(Y)	+(Y)	5	D	S1	S1	sequenced by Illumina

doi:10.6342/NTU202302705

Table 4: Reproductive photoperiodism phenotype and genotype at key loci of each individual (continued)

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*		Genotyping Method
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	
ym_female7	YM	seasonal	+(Y)	+(Y)	5	D	S1	S1	sequenced by Illumina
ym_female11	YM	seasonal	+(Y)	+(Y)	5	5	2	2	sequenced by Illumina
wls_female2	WL	seasonal	+(Y)	+(Y)	5	5	S1	DS	sequenced by Illumina
wls_male1	WL	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
wls_female1	WL	seasonal	+(Y)	+(Y)	5	5	S1	S1	sequenced by Illumina
wls_female4	WL	seasonal	+(Y)	+(Y)	5	5	S1	DS	sequenced by Illumina
wly_male1	WL	year-round	+(Y)	+(Y)	5	5	W	W	sequenced by Illumina
wly_female1	WL	year-round	+(Y)	+(Y)	5	5	W	W	sequenced by Illumina
wls_female5	WL	seasonal	+(W)	-(W)	5	D	S1	DS	sequenced by Illumina
wls_female3	WL	seasonal	+(W)	-(W)	5	N	S1	DS	sequenced by Illumina
wls_male4	WL	seasonal	+(W)	-(W)	5	N	S1	DS	sequenced by Illumina
wls_male3	WL	seasonal	+(W)	-(W)	5	N	S1	S1	sequenced by Illumina
wls_male2	WL	seasonal	+(W)	-(W)	5	N	S1	DS	sequenced by Illumina
wls_female6	WL	seasonal	+(W)	-(W)	5	N	S1	DS	sequenced by Illumina
wly_male3	WL	year-round	+(W)	-(W)	5	D	W	S1	sequenced by Illumina
wly_female4	WL	year-round	+(W)	-(W)	5	N	W	W	sequenced by Illumina
wly_male2	WL	year-round	+(W)	-(W)	5	D	S1	S1	sequenced by Illumina
wly_female2	WL	year-round	+(W)	-(W)	5	D	S1	S1	sequenced by Illumina
wly_female3	WL	year-round	+(W)	-(W)	5	D	W	DS	sequenced by Illumina
wly_male6	WL	year-round	-(W)	-(W)	N	D	W	S1	sequenced by Illumina

doi:10.6342/NTU202302705

Table 4: Reproductive photoperiodism phenotype and genotype at key loci of each individual (continued)

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*		Genotyping Method
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	
wly_male5	WL	year-round	-(W)	-(W)	N	D	W	DS	sequenced by Illumina
wly_male4	WL	year-round	+(W)	-(W)	D	N	W	S1	sequenced by Illumina
wls_female9	WL	seasonal	+	+	5	5	W	DS	sequenced by Illumina
wls_male5	WL	seasonal	+	+	5	5	2	S1	sequenced by Illumina
wly_female5	WL	year-round	+	+	5	5	W	W	sequenced by Illumina
wly_female6	WL	year-round	+	+	5	5	W	S1	sequenced by Illumina
wly_male7	WL	year-round	+	+	5	5	W	DS	sequenced by Illumina
wls_male6	WL	seasonal	+	+	5	5	W	DS	sequenced by Illumina
wls_male7	WL	seasonal	+	+	5	5	S1	new	sequenced by Illumina
wly_female7	WL	year-round	+	+	5	5	W	new	sequenced by Illumina
wls_female7	WL	seasonal	+	+	5	D	S1	new	sequenced by Illumina
wls_male8	WL	seasonal	+	+	5	5	S1	DS	sequenced by Illumina
wls_male9	WL	seasonal	+	+	5	5	S1	DS	sequenced by Illumina
wls_female8	WL	seasonal	+	+	5	5	W	W	sequenced by Illumina
TNY_00300	MH	year-round	-(M)	-(M)	N	D	S1	S1	PCR & sequenced by Sanger
TNY_00301	MH	year-round	-(M)	-(M)	N	D	DS	DS	PCR & sequenced by Sanger
TNY_04844	MH	year-round	-(M)	-(M)	N	D	S1	S1	PCR & sequenced by Sanger
TNY_04864	MH	year-round	-(M)	-(M)	D	D	S1	S1	PCR & sequenced by Sanger
TNY_05157	MH	year-round	-(M)	-(M)	D	D	W	S1	PCR & sequenced by Sanger
TNY_05183	MH	year-round	-(M)	-(M)	D	D	W	S1	PCR & sequenced by Sanger

doi:10.6342/NTU202302705

Table 4: Reproductive photoperiodism phenotype and genotype at key loci of each individual (continued)

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*		Genotyping Method
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	
TNY_02649	MH	year-round	-(M)	-(M)	N	D	W	S1	PCR & sequenced by Sanger
TNY_02659	MH	year-round	-(M)	-(M)	N	D	S1	S1	PCR & sequenced by Sanger
TNY_01175	MH	year-round	-(M)	-(M)	N	D	S1	DS	PCR & sequenced by Sanger
TNY_01185	MH	year-round	-(M)	-(M)	D	D	S1	S1	PCR & sequenced by Sanger
TNY_01369	MH	year-round	-(M)	-(M)	D	D	W	S1	PCR & sequenced by Sanger
TNY_01373	MH	year-round	-(M)	-(M)	N	D	W	S1	PCR & sequenced by Sanger
TNY_01400	MH	year-round	-(M)	-(M)	N	D	S1	S1	PCR & sequenced by Sanger
TNY_01417	MH	year-round	-(M)	-(M)	D	D	S1	S1	PCR & sequenced by Sanger
TNY_02920	MH	year-round	-(M)	-(M)	I	D	S1	DS	PCR & sequenced by Sanger
TNY_02955	MH	year-round	-(M)	-(M)	N	N	W	W	PCR & sequenced by Sanger
TNY_01952	MH	year-round	-(M)	-(M)	N	N	W	W	PCR & sequenced by Sanger
TNY_05624	MH	year-round	-(M)	-(M)	D	D	W	S1	PCR & sequenced by Sanger
TNY_05638	MH	year-round	-(M)	-(M)	N	D	W	DS	PCR & sequenced by Sanger
TNY_04704	MH	year-round	-(M)	-(M)	N	D	S1	S1	PCR & sequenced by Sanger
TNY_04717	MH	year-round	-(M)	-(M)	N	D	W	W	PCR & sequenced by Sanger
TNY_04727	MH	year-round	-(M)	-(M)	D	D	W	W	PCR & sequenced by Sanger
TNY_04738	MH	year-round	-(M)	-(M)	N	N	W	S1	PCR & sequenced by Sanger

For the inversion allele columns, the alphabet in the parentheses represents which clade the haplotype belongs to in the inversion region haplotype phylogenetic tree (Fig. 5), “Y” refers to the clade all Mt. Yangming individuals are in, “W” refers to the clade that most of the Wulai individuals are in, and “M” refers to the clade all Mt. Hehuan individuals are in. No parentheses mean this individual was not used in the

doi:10.6342/NTU202302705

haplotype phylogeny reconstruction. For the population column, “YM” refers to Mt.Yangming, “WL” refers to Wulai, and “MH” refers to Mt. Hehuan.

doi:10.6342/NTU202302705

Table 5: Summary of random forest classifier of all 77 individuals in Table 4

	With population	Without population
Accuracy	1.00	0.80
Precision	1.00	0.69
Recall	1.00	1.00
F1-score	1.00	0.82
Matthews Correlation Coefficient	1.00	0.66
Log-loss	0.17	0.38
Mean Absolute Error	0.00	0.20
Mean Squared Error	0.00	0.20

The accuracy of a random forest classifier generated from the phenotype, alleles, and population data of 77 individuals is higher than the random forest classifier generated only from the phenotype and alleles data of 77 individuals.

doi:10.6342/NTU202302705

Table 6: Data for generating random forest classifier of A-immigration type

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*	
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
wls_female5	WL	seasonal	w+	w-	5	D	S1	DS
wls_female3	WL	seasonal	w+	w-	5	N	S1	DS
wls_male4	WL	seasonal	w+	w-	5	N	S1	DS
wls_male3	WL	seasonal	w+	w-	5	N	S1	S1
wls_male2	WL	seasonal	w+	w-	5	N	S1	DS
wls_female6	WL	seasonal	w+	w-	5	N	S1	DS
wly_male3	WL	year-round	w+	w-	5	D	W	S1
wly_female4	WL	year-round	w+	w-	5	N	W	W
wly_male2	WL	year-round	w+	w-	5	D	S1	S1
wly_female2	WL	year-round	w+	w-	5	D	S1	S1
wly_female3	WL	year-round	w+	w-	5	D	W	DS
wly_male6	WL	year-round	w-	w-	N	D	W	S1
wly_male5	WL	year-round	w-	w-	N	D	W	DS
TNY_00300	MH	year-round	*m-	*m-	N	D	S1	S1
TNY_00301	MH	year-round	*m-	*m-	N	D	DS	DS
TNY_04844	MH	year-round	*m-	*m-	N	D	S1	S1
TNY_04864	MH	year-round	*m-	*m-	D	D	S1	S1
TNY_05157	MH	year-round	*m-	*m-	D	D	W	S1
TNY_05183	MH	year-round	*m-	*m-	D	D	W	S1
TNY_02649	MH	year-round	*m-	*m-	N	D	W	S1

doi:10.6342/NTU202302705

Table 6: Data for generating random forest classifier of A-immigration type (continued)

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*	
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
TNY_02659	MH	year-round	*m-	*m-	N	D	S1	S1
TNY_01175	MH	year-round	*m-	*m-	N	D	S1	DS
TNY_01185	MH	year-round	*m-	*m-	D	D	S1	S1
TNY_01369	MH	year-round	*m-	*m-	D	D	W	S1
TNY_01373	MH	year-round	*m-	*m-	N	D	W	S1
TNY_01400	MH	year-round	*m-	*m-	N	D	S1	S1
TNY_01417	MH	year-round	*m-	*m-	D	D	S1	S1
TNY_02920	MH	year-round	*m-	*m-	I	D	S1	DS
TNY_02955	MH	year-round	*m-	*m-	N	N	W	W
TNY_01952	MH	year-round	*m-	*m-	N	N	W	W
TNY_05624	MH	year-round	*m-	*m-	D	D	W	S1
TNY_05638	MH	year-round	*m-	*m-	N	D	W	DS
TNY_04704	MH	year-round	*m-	*m-	N	D	S1	S1
TNY_04717	MH	year-round	*m-	*m-	N	D	W	W
TNY_04727	MH	year-round	*m-	*m-	D	D	W	W
TNY_04738	MH	year-round	*m-	*m-	N	N	W	S1

Data of WL(Wulai) individuals have been additionally copied twice while generating the model, and we didn't use population as a feature. Individuals with "*" in inversion were not used in the haplotype phylogeny reconstruction, and we inferred their inversion type is "m-" since all the Mt. Hehuan individuals are "m-" type in the haplotype phylogenetic tree (Fig. 7).

doi:10.6342/NTU202302705

Table 7: Data for generating random forest classifier of B-immigration type

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*	
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
wls_female2	WL	seasonal	y+	y+	5	5	S1	DS
wls_male1	WL	seasonal	y+	y+	5	5	S1	S1
wls_female1	WL	seasonal	y+	y+	5	5	S1	S1
wls_female4	WL	seasonal	y+	y+	5	5	S1	DS
wly_male1	WL	year-round	y+	y+	5	5	W	W
wly_female1	WL	year-round	y+	y+	5	5	W	W
wls_female9	WL	seasonal	*y+	*y+	5	5	W	DS
wls_male5	WL	seasonal	*y+	*y+	5	5	2	S1
wly_female5	WL	year-round	*y+	*y+	5	5	W	W
wly_female6	WL	year-round	*y+	*y+	5	5	W	S1
wly_male7	WL	year-round	*y+	*y+	5	5	W	DS
wls_male6	WL	seasonal	*y+	*y+	5	5	W	DS
wls_male7	WL	seasonal	*y+	*y+	5	5	S1	new
wly_female7	WL	year-round	*y+	*y+	5	5	W	new
wls_female7	WL	seasonal	*y+	*y+	5	D	S1	new
wls_male8	WL	seasonal	*y+	*y+	5	5	S1	DS
wls_male9	WL	seasonal	*y+	*y+	5	5	S1	DS
wls_female8	WL	seasonal	*y+	*y+	5	5	W	W
ym_male8	YM	seasonal	y+	y+	5	5	W	S1
ym_female10	YM	seasonal	y+	y+	5	5	S1	S1

doi:10.6342/NTU202302705

Table 7: Data for generating random forest classifier of B-immigration type (continued)

ID	Population	Phenotype	inversion		*PMP34*		*ceh-30*	
			allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
ym_female9	YM	seasonal	y+	y+	5	5	W	W
ym_female8	YM	seasonal	y+	y+	5	5	W	W
ym_female1	YM	seasonal	y+	y+	5	5	W	S1
ym_male1	YM	seasonal	y+	y+	5	5	S1	DS
ym_female12	YM	seasonal	y+	y+	5	5	W	2
ym_male9	YM	seasonal	y+	y+	5	5	W	W
ym_male3	YM	seasonal	y+	y+	5	5	S1	S1
ym_female3	YM	seasonal	y+	y+	5	5	W	S1
ym_female5	YM	seasonal	y+	y+	5	5	W	S1
ym_female4	YM	seasonal	y+	y+	5	5	W	S1
ym_male7	YM	seasonal	y+	y+	5	5	S1	S1
ym_male2	YM	seasonal	y+	y+	5	5	S1	S1
ym_female2	YM	seasonal	y+	y+	5	5	S1	S1
ym_male10	YM	seasonal	y+	y+	5	5	W	S1
ym_male5	YM	seasonal	y+	y+	5	5	S1	S1
ym_male6	YM	seasonal	y+	y+	5	5	S1	S1
ym_male4	YM	seasonal	y+	y+	5	5	S1	S1
ym_female6	YM	seasonal	*y+	*y+	5	D	S1	S1
ym_female7	YM	seasonal	*y+	*y+	5	D	S1	S1
ym_female11	YM	seasonal	*y+	*y+	5	5	2	2

doi:10.6342/NTU202302705

Data of WL(Wulai) individuals have been additionally copied twice while generating the model. Individuals with "*" in inversion were not used in the haplotype phylogeny reconstruction, and we inferred their inversion type is "y+" since all the inversion heterozygous are "y+" type in the haplotype phylogenetic tree (Fig. 7). We didn't use population as a feature while generating the model, and we didn't use the inversion type as a feature either since all individuals here have the same type of inversion.

doi:10.6342/NTU202302705

Table 8: Summary of random forest model of A-immigration type

	A-immigration	A-immigration (duplicated data)
Accuracy	0.91	1.00
Precision	1.00	1.00
Recall	0.90	1.00
F1-score	0.95	1.00
Matthews Correlation Coefficient	0.67	1.00
Log-loss	0.14	0.18
Mean Absolute Error	0.09	0.00
Mean Squared Error	0.09	0.00

doi:10.6342/NTU202302705

Table 9: Summary of random forest model of B-immigration type

	B-immigration	B-immigration (duplicated data)
Accuracy	0.83	0.83
Precision	0.00	1.00
Recall	0.00	0.50
F1-score	0.00	0.67
Matthews Correlation Coefficient	0.00	0.63
Log-loss	0.52	0.44
Mean Absolute Error	0.17	0.17
Mean Squared Error	0.17	0.17

doi:10.6342/NTU202302705