

Article

## Introgression Between Cultivars and Wild Populations of *Momordica charantia* L. (Cucurbitaceae) in Taiwan

Pei-Chun Liao<sup>1</sup>, Chi-Chu Tsai<sup>2</sup>, Chang-Hung Chou<sup>3</sup> and Yu-Chung Chiang<sup>4,\*</sup>

<sup>1</sup> Department of Biological Science and Technology, National Pingtung University of Science and Technology, Pingtung 912, Taiwan; E-Mail: pcliao@mail.npust.edu.tw

<sup>2</sup> Kaohsiung District Agricultural Research and Extension Station, Council of Agriculture, Pingtung 908, Taiwan; E-Mail: tsaicc9017@yahoo.com.tw

<sup>3</sup> Research Center for Biodiversity, China Medical University, Taichung 404, Taiwan; E-Mail: choumasa@mail.cmu.edu.tw

<sup>4</sup> Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung 804, Taiwan

\* Authors to whom correspondence should be addressed; E-Mail: yuchung@mail.nsysu.edu.tw; Tel.: +886-7-5253625; Fax: +886-7-5253609.

Received: 20 March 2012; in revised form: 16 May 2012 / Accepted: 18 May 2012 /

Published: 24 May 2012

---

**Abstract:** The landrace strains of *Momordica charantia* are widely cultivated vegetables throughout the tropics and subtropics, but not in Taiwan, a continental island in Southeast Asia, until a few hundred years ago. In contrast, the related wild populations with smaller fruit sizes are native to Taiwan. Because of the introduction of cultivars for agricultural purposes, these two accessions currently exhibit a sympatric or parapatric distribution in Taiwan. In this study, the cultivars and wild samples from Taiwan, India, and Korea were collected for testing of their hybridization and evolutionary patterns. The cpDNA marker showed a clear distinction between accessions of cultivars and wild populations of Taiwan and a long divergence time. In contrast, an analysis of eight selectively neutral nuclear microsatellite loci did not reveal a difference between the genetic structures of these two accessions. A relatively short divergence time and frequent but asymmetric gene flows were estimated based on the isolation-with-migration model. Historical and current introgression from cultivars to wild populations of Taiwan was also inferred using MIGRATE-n and BayesAss analyses. Our results showed that these two accessions shared abundant common ancestral polymorphisms, and the timing of the divergence and colonization of the Taiwanese wild populations is consistent with the geohistory of the Taiwan Strait land bridge of the Last Glacial Maximum (LGM). Long-term and recurrent

introgression between accessions indicated the asymmetric capacity to receive foreign genes from other accessions. The modern introduction of cultivars of *M. charantia* during the colonization of Taiwan by the Han Chinese ethnic group enhanced the rate of gene replacement in the native populations and resulted in the loss of native genes.

**Keywords:** divergence time; effective population size; IM model; introgression; *Momordica charantia*

---

## 1. Introduction

The tropical and subtropical bitter melon, *Momordica charantia*, has been widely cultivated and domesticated for a long time [1]. This species, which is commonly used as a vegetable and an ingredient in traditional medicines, is cultivated in tropical and temperate China [2]. The phylogenetically related species *M. balsamina* (balsam apple) was used as a vegetable at least 500 years ago [3]. Ethnobotanical investigations reveal that cultivars were derived from the local wild types and that the encouragement of local variant populations, through plant utilization and habitat modifications, accelerated greater diversity among cultivated strains than among the local wild types [4]. The close association between cultivated strains and localities fixed the differences between strains and may have decreased the within-strain genetic diversity, which is the most significant difference between domesticated and wild species [5,6]. Individuals in the wild population found in Taiwan have small fruits morphologically different from the cultivars. Currently, the wild individuals are observed in markets in Taiwan but are primarily collected from wild populations rather than cultivated. Increasingly, more studies indicate that the wild bitter melon contains antioxidants [7–9], which help to suppress the inflammatory responses [10–13] and lower blood-glucose levels in diabetes [14,15]. These studies of bitter melon focus on the medical properties but rarely explore the origin, speciation, and hybridization/introgression of the local strains or varieties. Recent genetic evidence has shown that the Cucurbitaceae originated in Africa, and the genus *Momordica* was derived from South Africa, tropical Africa and tropical Asia [16,17].

Hybridization between crops and wild populations is common. Domesticated crops are usually artificially selected for adaptive traits, such as pathogen resistance [18] and higher fertility [19]. These domestication characteristics might be closely associated [20–22] with the role of “supergenes” [23]. Therefore, the hybrids of the domesticated crops and the wild populations would have greater fitness. If introgression occurs, the genes of the domesticated crop could quickly replace the genes of the wild populations by a small acceleration of immigration (e.g., human-mediated spread) [24] and result in genetic assimilation [23], which is the phenomenon of replacing a pure conspecific of one of the hybridizing taxa. Because the artificial hybridization between *M. charantia* cultivars and wild populations is successful when carried out by the Hualien District Agricultural Research and Extension Station (<http://www.hdais.gov.tw/bred>) for the purpose of improving the cultivars, natural hybridization is likely to have occurred in nature.

In addition, the introduction of the cultivars into Taiwan Island by the Han Chinese ethnic group began hundreds of years ago at which time the wild population was already indigenous. Therefore, we

wondered when the native wild population colonized Taiwan Island. Taiwan Island is a continental island located off the coast of Southeast Asia. It was lifted by orogenesis by tectonic compression of the Philippine Sea Plate and the Eurasian Plate *circa* (*ca.*) 6~5 million years ago (Mya) [25] or *ca.* 3 Mya [26]. The surrounding sea level change caused by the Pleistocene climate oscillation caused successive connection and disconnection of Taiwan Island to the Asian continent [27]. This process promoted the colonization and then isolation of several species from the Asian continent in Taiwan [28]. The native wild population is most likely the descendent of the ancient colonizers. In this study, we estimated the divergence time based on plastid DNA and nuclear markers. The divergence time estimated by maternally inherited plastid DNA could exclude the effect of pollen flow and reveal the time of colonization, while the divergence time estimated from the nuclear multilocus markers reflects the degree of isolation and the divergence between populations.

To evaluate the possibility of sympatric hybridization and to test the hypotheses of the relation of colonization of wild populations in Taiwan with the geological history, the chloroplast DNA (cpDNA) and the nuclear microsatellite loci are used as genetic markers to examine the genetic similarity of the cultivars and the wild populations and to trace their evolutionary histories by coalescent approaches. Because multiple strains of *M. charantia*, which might have distinct genetic compositions due to artificial selection, have been domesticated, different cultivated strains (instead of wild populations) are used to compare with the wild samples collected from Taiwan and other countries (India and Korea). In this study, we used evolutionary analyses to examine the hybridization and to explore the origin of the native population of *M. charantia*. This study demonstrates a crisis of genetic assimilation by introgression from a widespread crop.

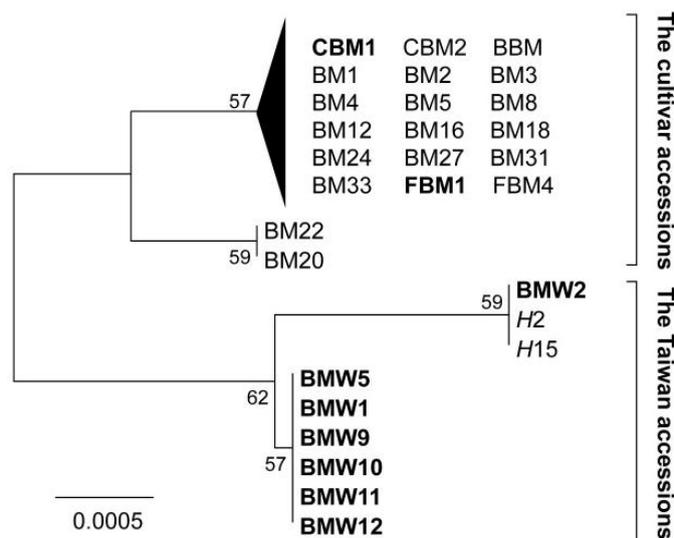
## 2. Results

### 2.1. CpDNA Polymorphisms and Neighbour Joining Tree

The cpDNA cp *atpB-rbcL* sequences from 164 individuals were aligned directly and five chlorotypes were found. The neighbor joining tree inferred from the cp *atpB-rbcL* sequences indicated that the wild population of Taiwan is distinct from other samples of *M. charantia*, including the cultivars and the wild samples of India and Korea, but grouped with the hybrid lines (Figure 1). The two major clades composed of samples of wild populations and hybrids from Taiwan and the samples of cultivars and wild samples of India and Korea are named “the Taiwan accession” and “the cultivar accession”, respectively. Because the cpDNA is maternally inherited, the interference of recent gene flow by pollen flow is eliminated. The BM22 (green cultivar) and BM20 (white cultivar) have identical genotype of cp *atpB-rbcL*, which means identical or similar maternal parents of these two cultivars. However, the parental lines of cultivars were not recognized and we cannot make sure whether these two cultivars are maternally identical by descents. The phylogenetic relationship displayed the long-term evolutionary differentiation between the Taiwan accession and the cultivar accession, which included wild samples from India and Korea. The grouping of hybrids indicated that the wild samples of Taiwan play the role of mother species (pollen receiver). The net average distance ( $D_A$ ) between the two accessions is  $2.077 \times 10^{-3} \pm 1.414 \times 10^{-3}$  (1000 bootstrap replicates under the maximum composite likelihood [MCL] substitution model). The divergence time ( $T$ ) between these

two accessions is 0.519 Mya ( $\pm 0.354$  Mya) using the formula  $T = D_A/2\mu$ , where  $\mu$  is the mutation rate, while a longer time to the most recent common ancestor (TMRCA, 2.002 Mya, 95% HPD: 1.146 Mya~2.796 Mya) between the two accessions was estimated by using the BEAST. The large difference between the divergence time and the coalescent time between the accessions might be due to the high variance of the single-locus estimation or to the long time allowed for gene flow at the beginning of speciation [29].

**Figure 1.** The neighbor-joining tree of the cultivars and wild samples of Taiwan, India, and Korea of *Momordica charantia*, inferred by the chloroplast *atpB-rbcL* spacer. Samples denoted in regular, bold, and italic are the cultivars, wild samples, and hybrids, respectively. The CBM1 and the FBM1 are wild samples from India and Korea, respectively. Each wild population and landrace cultivar variety included five to eight individuals with the same haplotype in the figure. Numbers between the branches are the supporting values for monophyly by 1000 bootstrap replications.



**Table 1.** Genetic diversity and neutrality tests of the chloroplast *atpB-rbcL* spacer in the cultivar accession and the Taiwan accession of *Momordica charantia* (gaps are excluded).

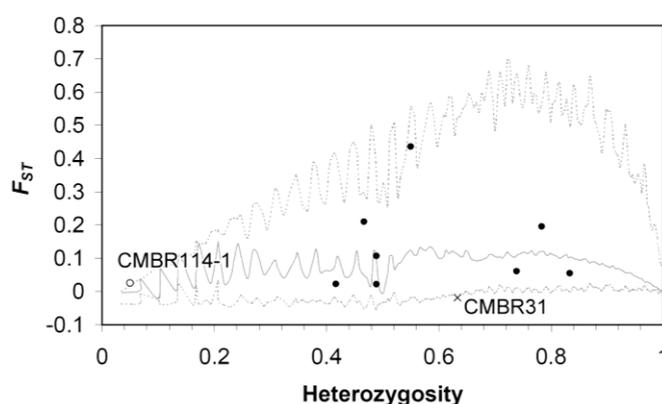
Taxa	<i>N</i> <sup>a</sup>	<i>H</i> <sup>b</sup>	<i>S</i> <sup>c</sup>	$\theta$ <sup>d</sup>	$\pi$ <sup>e</sup>	Tajima's <i>D</i> <sup>f</sup>
The cultivar accession	20	3	2	0.0005	0.0002	-1.2414
The Taiwan accession	7	2	1	0.0004	0.0003	-1.0062
Hybrid	2	1	0	0	0	-1.0488
Total	29	5	4	0.001	0.0011	0.1977

<sup>a</sup> Number of lines; <sup>b</sup> Number of haplotypes; <sup>c</sup> Segregating sites; <sup>d</sup> Genetic diversity index estimated from segregating sites; <sup>e</sup> Nucleotide diversity estimated by site-by-site pairwise comparison; <sup>f</sup> None of the values are significant and the null (neutral) hypothesis cannot be rejected.

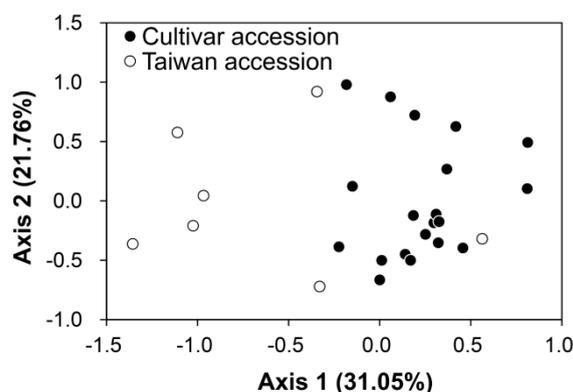
In total, five haplotypes of 164 cp *atpB-rbcL* sequences were obtained from 29 lines of cultivars, hybrids, and wild populations of *M. charantia*. The aligned cpDNA *atpB-rbcL* data matrix was 954 bp in length. The sequences from this study can be found under the GenBank accession numbers: HE585487-HE585491. Four segregating sites were examined, and low genetic diversity was estimated.

The nucleotide diversity ( $\pi$ ) and genetic diversity ( $\theta_w$ ) of total samples (including hybrids) is 0.00108 and 0.00100, respectively. The index of Tajima’s  $D$  of the total sample is 0.1977 and is not significant so the null hypothesis of neutral evolution cannot be rejected. Genetic diversity indices of two accessions are listed in Table 1.

**Figure 2.** Distribution of population genetic differentiation index  $F_{ST}$  as a function of heterozygosity by simulation, while considering the neutral markers only, and forcing the mean  $F_{ST}$  under a stepwise mutation model. The dashed lines and the solid line represent the 0.95 quantiles and the median, respectively. Each dot represents a microsatellite locus; loci above the 0.95 quantile line (upper dotted line) were classified as under divergent selection, while those below the lower dotted line were classified as loci under balancing selection.



**Figure 3.** Population structure estimated by the principle coordinate analysis (PCoA). Scatter plot of the first axis (31.05% explanation) and the second axis (21.76% explanation) of the PCoA based on variations of 8 microsatellite loci (excluding the  $F_{ST}$  outlier loci CMBR31 and CMBR114-1). The black and white dots indicated the individuals from “the cultivar accession” and “the Taiwan accession”, respectively.



**Table 2.** Primer sequences, annealing temperature ( $T_m$ ), and polymorphic type of eleven microsatellite loci and the expected heterozygosity ( $H_{exp}$ ), genetic diversity index ( $\theta$ ), number of alleles and effective alleles ( $N_a$  and  $N_e$ ) of eight neutral evolving and polymorphic loci. The forward and reverse primers are denoted in “F” and “R” in front of the primer sequences, respectively.

Locus	Primer set	$T_m$	Type *	cultivars				wild population (Taiwan)			
				$H_{exp}$	$\theta$	$N_a$	$N_e$	$H_{exp}$	$\theta$	$N_a$	$N_e$
CMBR21	F: 5'-AGATTCTGGTTGTTGGGCAG-3' R: 5'-CAGCGATGATCAACAGAAACA-3'	59 °C	P	0.528	1.506	3	2.062	0.264	2.075	2	1.324
CMBR22	F: 5'-CCAAAACGACCAAATGTTCC-3' R: 5'-ATACAGACACGCCTTCCACC-3'	56 °C	P	0.554	1.523	5	2.174	0.615	1.613	3	2.333
CMBR30	F: 5'-CACTGCATACACACATCCA-3' R: 5'-AAAAGAAGGAGGAGGAGGG-3'	58 °C	P	0.097	5.186	2	1.105	0.440	1.530	2	1.690
CMBR47	F: 5'-ATCCCAACCCATCACTCTCA-3' R: 5'-TGGGGACAGGTGAGAATATTAGA-3'	59 °C	P	0.738	2.089	6	3.571	0.791	2.527	5	3.769
CMBR57	F: 5'-GCTCTGAAGAGTGGAATGAGAGA-3' R: 5'-CCATTTGGGAAGTAGGCATC-3'	59 °C	P	0.708	1.917	6	3.226	0.659	1.726	4	2.579
CMBR66	F: 5'-TCAAGCAAAAACCATAATCAGAA-3' R: 5'-TCCCTTTTCATCATTTCTCTTCA-3'	52 °C	P	0.272	2.026	3	1.361	0.659	1.726	3	2.579
CMBR82	F: 5'-ACGACTCTTGGAATCGGTC-3' R: 5'-TTTAGAAAAGAATCACGAAGAGAGC-3'	54 °C	P	0.344	1.717	3	1.504	0.527	1.506	2	1.960
CMBR152	F: 5'-CCCACATTGGTCTCAACAAG-3' R: 5'-AAAAAATTTGGCATTAGCTATAAAAA-3'	54 °C	P	0.590	1.567	4	2.353	0.440	1.530	2	1.690
CMBR31	F: 5'-AAACAAACCAAACCAAACCG-3' R: 5'-AAAAAGAAGCGGGAGTAATGA-3'	58 °C	BS	0.610	1.602	4	2.469	0.703	1.896	3	2.882
CMBR114-1	F: 5'-TGCTTTGCCTTAACCGTCTT-3' R: 5'-TGAGTGCCCAAGATGTTGTC-3'	52 °C	PS	0.097	5.186	2	1.105	0	0	1	1.000
CMBR114-2	F: 5'-TGCTTTGCCTTAACCGTCTT-3' R: 5'-TGAGTGCCCAAGATGTTGTC-3'	52 °C	M	-	-	-	-	-	-	-	-
CMBR145	F: 5'-TGTGACAATGTGCAACCAG-3' R: 5'-AAAAATGGTGTAAACGACATGG-3'	56 °C	M	-	-	-	-	-	-	-	-

\* P: polymorphic; M: Monomorphic; PS: positive selection; BS: balancing selection.

## 2.2. Nuclear Microsatellite Polymorphisms

Eleven of 72 primer sets developed from Ritschel *et al.* [30] were successfully amplified in all samples of this study and 12 loci were obtained from these 11 loci, which were all confirmed by direct sequencing. Except for the two loci (CMBR114-1 and CMBR114-2) that were amplified from the primer set CMBR114, one locus was obtained from each of the primer sets. However, two loci (CMBR114-2 and CMBR145) are monomorphic. The neutrality of evolution of each locus was tested by detecting the  $F_{ST}$ -outliers by criteria of 99% confidence intervals (CI). CMBR114-1 and CMBR31, which are thought to have evolved under positive selection (extremely high  $F_{ST}$ ) and balancing selection (extremely low  $F_{ST}$ ), respectively (Figure 2), were detected as the  $F_{ST}$ -outlier loci. The locus CMBR114-1 is monomorphic in wild population of Taiwan samples but polymorphic in cultivars. The fixation of CMBR114-1 in Taiwan accession is probably reasoning for the extremely high  $F_{ST}$  among accessions. We estimated the genetic diversity by expected heterozygosity ( $H_{exp}$ ) and genetic diversity index ( $\theta$ ) of the remaining eight microsatellite loci for each accession and obtained a wide range of diversity for each locus: for the cultivar accession  $H_{exp}$  ranges from 0.097 to 0.738, and  $\theta$  ranges from 1.506 to 5.186; and for the Taiwan accession  $H_{exp}$  ranges from 0.264 to 0.791, and  $\theta$  ranges from 1.506 to 2.527 (Table 2). Note that the estimates for the cultivar accession are based on the genetic diversity of cultivated strains, but the estimates for the Taiwan accession represent the genetic diversity of the wild population.

## 2.3. Genetic Mixing Between the Cultivars and the Wild Population of Taiwan

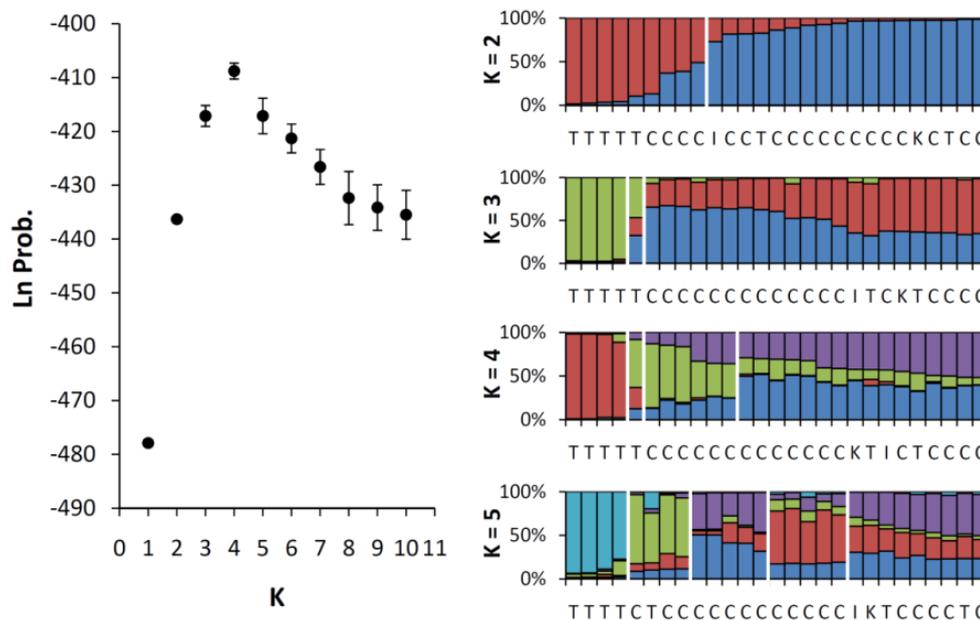
The AMOVA results showed that most of the genetic variation results from within-accession variation (99.44%), which indicates that the genetic differentiation between the two accessions is not significant (fixation index  $R_{ST} = 0.0056$ ,  $P = 0.6149$ ). From the review of genetic differentiation by Storz *et al.* (2005), genetic differentiation will be significant if the genetic variation of one of two populations (here, the accessions of *M. charantia*) is significantly lower than the other. In our study, the high percentage of genetic variation within accessions combined with insignificant genetic differentiation between accessions seems to imply that the degree of genetic variation of the Taiwan accession, the wild bitter melon, is not very different from the genetic variation among multiple cultivar lines of *M. charantia*. However, the accession-specific  $R_{ST}$  values of the two accessions are 0.00139 and 0.01758 in the cultivar accession and the Taiwan accession, respectively. The difference in  $R_{ST}$  between the two accessions is larger than 10-fold, indicating a larger genetic differentiation of the wild population, which seems unreasonable because different sources (lines) of the cultivars have been suggested to have higher differentiation than the local populations. Therefore, clustering patterns of these samples were rechecked with PCoA. We also used the PCoA to reassign the clustering of collected samples without any assumption of prior classification. The result of the PCoA revealed mixture patterns of individuals between the two accessions in two axes with 52.81% explanation (31.05% and 21.76% for the first and the second axis, respectively) (Figure 3). This result implies that there is small genetic differentiation with certain admixture between the two accessions, which is consistent with the result of the AMOVA (Table 3). To understand whether there are cryptic groupings in certain cultivated lines of *M. charantia* and wild populations in Taiwan, we performed another

assignment test based on the genetic composition with the assistance of STRUCTURE v. 2.2 [31]. In the STRUCTURE analyses, both admixture and no admixture models were tested in 10 replicates for  $K = 1\sim 10$ , where  $K$  is the grouping number, and the likelihood values of each  $K$  were evaluated. The best clustering was  $K = 4$  (the highest likelihood,  $\ln L = -362.7$ , estimated  $\ln \text{Prob.} = -411.6$ , Figure 4), but the grouping was not consistent with the grouping of the cpDNA tree. This result indicates that the genetic composition of nuclear DNA is not assigned as well as the grouping of maternally inherited cpDNA and implies contributions of pollen flow to genetic admixture.

**Table 3.** Analysis of molecular variance (AMOVA) between the cultivar accession and the Taiwan accession. The variance explained by differences among accessions and its significance were calculated using probabilities derived from 1000 permutations.

Source of variation	d.f.	Sum of Squares	Variance Components	Percentage of Variance	$R_{ST}$	$P$
Between accessions	1	104.955	0.528	0.56	0.0056	0.615
Within accessions	52	4887.786	93.996	99.44		
Total	53	4992.741	94.524	100		

**Figure 4.** Genetic subdivision inferred using the Bayesian-clustering method implemented in the STRUCTURE program [31] for eight neutral microsatellite loci. The left plot is the log of the estimated posterior probability of  $K$ , which indicates a best fit of  $K = 4$ ; the right panels are individual genotypes grouped by  $K = 2\sim 5$ ; the marks C, T, I, and K of each individual plot indicate the cultivar samples, the wild samples of Taiwan, India, and Korea, respectively.

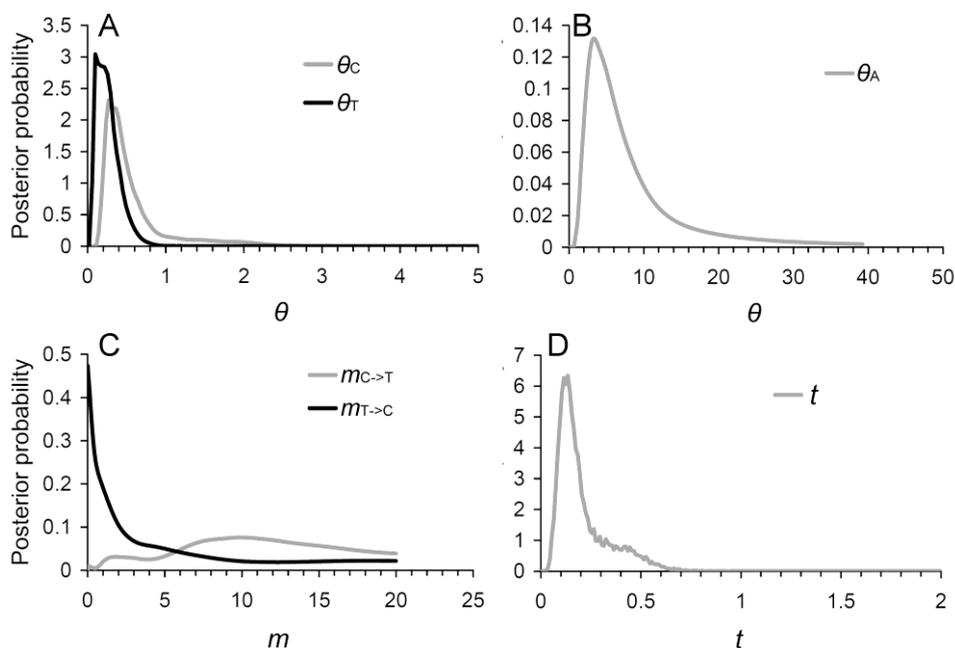


#### 2.4. Evidence of Asymmetric Introgression Inferred by the Isolation-with-Migration Model

The isolation-with-migration (IM) model [32] that allows for the population size change and gene flow after divergence was used for examining the degrees and directions of gene flow between the

cultivar accession and the Taiwan accession as well as their divergence time. For fitting the assumption of the IM model, two putatively positively selected genes were removed in this analysis. According to the eight neutral loci, the unscaled divergence time  $t = 0.135$  (95% CI: 0.065–0.545) has the highest posterior probability (Figure 5D) estimated by the program IMa [33]. If we used a rough mutation rate of  $7.47 \times 10^{-6}$  per year for microsatellites, the divergence time is *ca.* 18.07 kilo-years ago (kya) (95% CI: 8.70 kya~72.96 kya), which is much smaller than the 0.519 Mya estimated from cpDNA *atpB-rbcL* spacer. The difference in divergence time inferences might be due to the intrinsic factor of rapid evolutionary rate of microsatellites compared with cpDNA, resulting in faster coalescence of the wide ranges of evolutionary rate of microsatellites, and extrinsic factors such as the pollen flow. In addition, we also noticed that the posterior probability is zero when  $t = 0$ , indicating that these two accessions are truly diverged. This result does not support the inference of the admixture pattern of genetic structure analyses (PCoA, AMOVA, and STRUCTURE).

**Figure 5.** Marginal posterior probability density distribution of (A) the current effective population size of the cultivar accession and the Taiwan accession; (B) the ancestral population size; (C) migration rates of the cultivar accession and the Taiwan accession; and (D) divergence time of these two accessions, each scaled by the geometric mean of the mutation rates of all loci analyzed.



In the analysis of gene flow, the direction from the cultivar accession to the Taiwan accession has a maximum posterior probability of migration rate  $m = 9.95$  while the opposite direction of migration has a maximum posterior probability of  $m = 0.01$  (Table 4 and Figure 5C). However, this result cannot be used to reject the null model  $m_{T \rightarrow C} = 0$  (Log-likelihood ratio 2LLR = 0.002,  $P = 0.967$ ), indicating a unidirectional gene flow (or introgression) from the cultivar accession to the Taiwan accession (Table 4). The estimated effective population size of the ancestor ( $\theta_A = 3.3572$ ) is considerably larger than both current populations of the cultivar accession ( $\theta_C = 0.3314$ ) and the Taiwan accession ( $\theta_T = 0.1767$ ) (Figure 5A,B). The fourfold larger population size of the cultivar accession compared

with the native wild population of Taiwan is most likely due to the multiple sources of cultivated lines. However, the equal effective population sizes of the cultivar and the Taiwan accessions ( $\theta_C = \theta_T$ ) cannot be rejected (2LLR = 0.0056,  $P = 0.470$ ), which suggests that these two accessions preserved equal weighted genetic diversity (Table 4). Both the estimation of the equal population size and unidirectional gene flow implied that the wild population of Taiwan receives a large amount of polymorphism from the cultivar accession, and the abundant foreign polymorphism received by the Taiwan population explains its high accession-specific  $R_{ST}$ . Except for the model  $m_{T \rightarrow C} = 0$  and  $\theta_C = \theta_T$ , the other nested models were rejected at the level of  $P < 0.05$  by the log-likelihood ratio tests (Table 4).

**Table 4.** Log-likelihood nested model tests by the full model ( $\theta_C \neq \theta_T \neq \theta_A$ ,  $m_1 \neq m_2 \neq 0$ , where  $m_1 = m_{C \rightarrow T}$  and  $m_2 = m_{T \rightarrow C}$ ). Statistical nonsignificance (not rejecting the null model) is indicated in bold.

Model	T	$\theta_C$	$\theta_T$	$\theta_A$	$m_{C \rightarrow T}$	$m_{T \rightarrow C}$	df	2LLR †	P
Full model									
mode	0.135	0.3314	0.1767	3.3572	9.95	0.01			
95% CI Low	0.065	0.1736	0.0589	1.708	1.61	0.05			
95% CI High	0.545	1.9409	0.6086	29.5471	19.37	18.85			
Null models									
$m_1 = m_2$	0.1167	0.5634	0.1469	2.1057	3.2863	3.2863	1	4.3693	0.0366
$m_2 = 0$	<b>0.0625</b>	<b>0.2105</b>	<b>0.2230</b>	<b>2.6982</b>	<b>19.9976</b>	<b>0.0001</b>	<b>1</b>	<b>0.0056</b>	<b>0.4702</b>
$m_1 = 0$	0.0995	0.4138	0.2023	2.5454	0.0001	9.4322	1	14.8471	$5.829 \times 10^{-5}$
$m_1 = m_2 = 0$	0.1785	0.7870	0.4202	1.8314	0.0001	0.0001	2	718.0318	$6.03 \times 10^{-157}$
$\theta_C = \theta_T$	<b>0.0625</b>	<b>0.2163</b>	<b>0.2163</b>	<b>2.7202</b>	<b>19.9992</b>	<b>0.0002</b>	<b>1</b>	<b>0.3680</b>	<b>0.5441</b>
$\theta_C = \theta_T = \theta_A$	0.2213	0.6101	0.6101	0.6101	7.8461	2.8492	2	16.3993	$2.747 \times 10^{-4}$
$\theta_C = \theta_T, m_1 = m_2$	0.0902	0.1906	0.1906	2.1566	19.5826	19.5826	2	9.8940	0.0071
$\theta_C = \theta_T, m_1 = m_2 = 0$	0.1785	0.7026	0.7026	1.8250	0.0001	0.0001	3	746.7677	$7.58 \times 10^{-162}$
$\theta_C = \theta_T = \theta_A, m_1 = m_2$	0.2213	0.6107	0.6107	0.6107	5.4662	5.4662	3	51.8731	$3.19 \times 10^{-11}$
$\theta_C = \theta_T = \theta_A, m_1 = m_2 = 0$	0.1785	0.8188	0.8188	0.8188	0.0001	0.0001	4	803.2170	$7.72 \times 10^{-173}$
$\theta_C = \theta_A$	0.2537	1.0702	0.4156	1.0702	7.3973	0.0478	1	5.9515	0.0147
$\theta_C = \theta_A, m_1 = m_2$	0.2980	1.5204	0.3881	1.5204	1.4217	1.4217	2	25.8596	$2.42 \times 10^{-6}$
$\theta_C = \theta_A, m_1 = m_2 = 0$	0.1785	0.9223	0.4204	0.9223	0.0001	0.0001	3	759.3913	$1.39 \times 10^{-164}$
$\theta_T = \theta_A$	0.2213	0.6338	0.5800	0.5800	7.8300	2.8436	1	15.4141	$8.63 \times 10^{-5}$
$\theta_T = \theta_A, m_1 = m_2$	0.2213	0.6347	0.5806	0.5806	5.4683	5.4683	2	50.8823	$8.93 \times 10^{-12}$
$\theta_T = \theta_A, m_1 = m_2 = 0$	0.1785	0.7864	0.8914	0.8914	0.0001	0.0001	3	801.3394	$1.11 \times 10^{-173}$

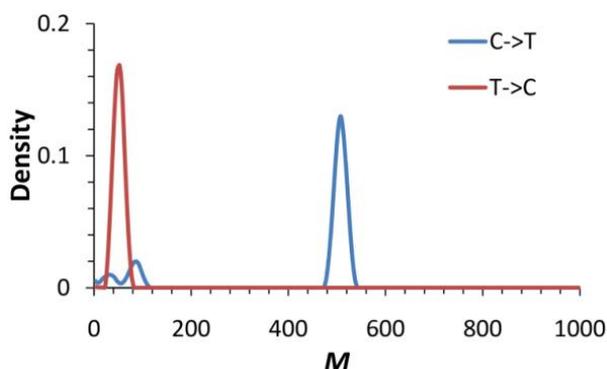
† The log-likelihood ratio result from IMA analysis [33] approximates a  $\chi^2$  distribution. For models where migration estimates are set to zero, the expected distribution is a mixture, and the 2LLR approximates a  $1/2\chi^2$  distribution [33]. T: unscaled divergent time;  $\theta_C$ ,  $\theta_T$ , and  $\theta_A$ : unscaled population size of cultivar accession, Taiwan accession, and ancestors, respectively;  $m_{C \rightarrow T}$  and  $m_{T \rightarrow C}$ : migration rate from cultivar accession to Taiwan accession and the opposite direction, respectively; df: degrees of freedom; 2LLR denotes the log-likelihood ratio test expressed by two-times difference of logarithmic likelihoods between alternative and null models; P is the one-tailed probability value for the  $\chi^2$  test used for LRT.

### 2.5. Historical and Contemporary Gene Flow

Patterns of historical gene flow estimated by MIGRATE-n v. 3.0 [34] are consistent with the inference by IMA [33] that the degrees of historical gene flow from the cultivar accession to the Taiwan accession ( $M_{C \rightarrow T} = 507.5$ , 97.5% CI: 475.0~535.0, where the suffixes C and T indicate the

cultivar accession and the Taiwan accession, respectively) are larger than the opposite direction ( $M_{T \rightarrow C} = 52.5$ , 97.5% CI: 25.0~75.0) (Table 5 and Figure 6). The migration rate  $M$  estimated by using the MIGRATE-n [34] was scaled by  $\mu$  and such high and significant different values of  $M$  indicated the historical asymmetric gene flow (*i.e.*, introgression) between two accessions. This result is consistent with the inference of unidirectional gene flow by IMa analysis. Such asymmetric gene flow was also detected in the recent migration events by the assignment test of BayesAss [35] that (1) the inter-accession gene flow was detected at a frequency (rate) of migrations of 0.166 (95% CI: 0.00787~0.325) between accessions under  $10^7$  simulations, and (2) the rate of gene flow from the cultivar accession to the Taiwan accession is 0.0707 (95% CI: 0.0026~0.1963), and the opposite direction is 0.0151 (95% CI: 0.0005~0.0522) (Table 5). The rate of migration within accessions (intra-accessions gene flow) was 0.9849 (95% CI: 0.9478~0.9995) and 0.9293 (95% CI: 0.8037~0.9974) in the cultivar accession and the Taiwan accession, respectively. The smaller degree of gene flow between accessions than the intra-accession gene flow also implied more or less isolation (barrier) of reproduction between accessions in spite of the secondary contacts due to agricultural reasons.

**Figure 6.** Posterior distribution of migration rate ( $M$ ) over all loci between the cultivar accession and the Taiwan accession estimated by using the MIGRATE-n [34] analysis.



**Table 5.** Summary of immigration rate between accessions of *Momordica charantia* estimated by IMa [33], MIGRATE-n [34], and BayesAss [35].

Source	IMa <sup>a</sup>		MIGRATE-n <sup>a</sup>		BayesAss <sup>a</sup>	
	Cultivar accession	Taiwan accession	Cultivar accession	Taiwan accession	Cultivar accession	Taiwan accession
Cultivar accession	-	2.655	-	507.5	0.985	0.071
Taiwan accession	0.195	-	52.5	-	0.015	0.929

<sup>a</sup> Immigration rates were revealed in  $m$  of IMa [33],  $M (= m/\mu)$  of MIGRATE-n [34], and the proportion of migrants in the data sets of two accessions in a migration rate of 0.166 (95% CI: 0.008~0.325) of BayesAss [35], respectively.

### 3. Discussion

The first aim of this study was to test whether the cultivars of bitter gourds are genetically differentiated from the wild populations of Taiwan or whether these two accessions are a genetic mixture. Although the cultivars and the wild populations of Taiwan are taxonomically defined as the

same taxon, which is also supported by AMOVA (Table 3), PCoA (Figure 3), and Bayesian clustering analysis (STRUCTURE) (Figure 4) by nuclear SSR, the phylogenetically distinguished accession of the wild population in Taiwan from cultivars indicated that the wild population of Taiwan has been differentiated from the inland populations (including two lines of wild populations from India and Korea) for a long time. The inland populations of Asia are thought to be sources of the domesticated lines of bitter gourds. Because of the maternally inherited characters, the cpDNA tree reflects the faithful phylogenetic relationships without the interference of recent pollen flow. The cpDNA phylogeny revealed different evolutionarily significant units between the cultivar accession and the wild population of Taiwan (the Taiwan accession), and the incongruent grouping between the cpDNA and nuclear SSR implied the recent introgression between the two accessions due to pollen flow. The domestication history of *M. charantia* spans thousands of years [1], and the morphological transitions in the fruit size, shape, color, taste, and many other traits are not reflected by the selected loci used in this study. Although only a small number of the SSR loci were used, which cannot completely represent the whole genome variation, the undifferentiated patterns detected in eight neutral loci implied that these morphological transitions are not the outcome of species (population) divergence but a consequence of selection (either natural or artificial) for specific traits. Accordingly, the impact of the translocation of diversified cultivars on wild populations might be encrypted in phenotypes and therefore worth further consideration.

There are two explanations for the admixture pattern of the genetic composition of the two accessions: (1) incomplete divergence by sharing abundant common ancestral polymorphisms and (2) gene flow after divergence. Although the posterior probability of the splitting time ( $t$ ) estimated by using the IMA analysis is zero at minimum  $t$  (Figure 5D), the high estimates of historical  $M$  by MIGRATE-n analysis suggest that the divergence between the two accessions is not complete. In other words, the divergence of nuclear genomes was begun at *ca.* 18 kya, but the speciation process is still progressing. This age roughly matches the time of LGM, during which the climate was colder and drier than at present, and the distribution of the ancestors of these two accessions might be restricted to refugia. At that time, the shallow shelf of the Taiwan Strait was exposed and formed a land bridge between mainland Asia and Taiwan Island. The LGM at approximately 18 kya has been suggested to be a major factor in influencing species evolution (e.g., speciation) in Taiwan. After the LGM (*i.e.*, postglaciation), the geographic barrier of the Taiwan Strait isolated the island populations from mainland China. The perfect match of the divergence time of the two accessions and the age of the LGM suggests a relation between the evolutionary history of *M. charantia* and the geohistory of Southeast Asia. In addition, the shorter divergence time estimated by microsatellites than cpDNA implied the continuous gene transport by pollen flow between accessions after the divergence of maternally inherited materials since the middle Pleistocene, *i.e.*, a long-term process of speciation (divergence).

Furthermore, the large ancestral population size ( $\theta_A = 3.357$ ) of the two accessions (as inferred by the coalescent approach under the IM model) is in contrast to the relatively small population sizes of the derived accessions, reflecting a prominent stochastic effect of lineage sorting among multiple loci [36]. The large ancestral population size also implied that these two accessions share an abundant genetic background with their ancestors before divergence. Considering the recent divergence time (*ca.* 18 kya) inferred by microsatellites and that the cultivar accession is not native to Taiwan Island,

the native population of the Taiwan accession most likely colonized Taiwan during the late Pleistocene glaciations (*i.e.*, the LGM). Certain examples, such as lizards [37], earthworms [38], butterflies [39], and crabs [40], indicated that the LGM provided a chance for colonization from mainland Asia to Taiwan, and the subsequent postglacial sea level rise caused the isolation of colonizers from the continental populations and resulted in genetic differentiation. However, the gene flow seems to have continued after the separation of Taiwan Island and mainland Asia as inferred by MIGRATE-n [34] and BayesAss [35]. The past and current gene flow between accessions in the coalescent process illustrates a gradual process of speciation of the local population of Taiwan instead of a sudden allopatric isolation [41]. This inference is also supported by the wide range of the 95% CI of the divergence time (8.70 kya~72.96 kya) inferred by IMA. The postglacial climate warming provided better growth conditions for both continental and island populations. Therefore, geographic isolation and local adaptation might be two key factors contributing to the divergence (speciation) of the two accessions.

In addition, the relatively smaller effective population size of the wild population of Taiwan (approximately a quarter of the cultivars) is truly a reflection of the locally restricted distribution in Taiwan. Reduced genetic diversity is a common feature of domesticated genomes because of a “genetic bottleneck” caused by the domestication process [42] or a selective sweep for local genomic regions surrounding the locus of human-mediated selection [43]. However, the cultivars revealed higher genetic diversity and lower genetic differentiation ( $R_{ST}$ ) than the wild population of Taiwan, which implied multiple sources of different cultivated strains (*i.e.*, independent domestication events from multiple local ecotypes instead of progressing domestication from one single strain) followed by artificial hybridization for breeding purposes. Multiple independent domestication events from local strains might have resulted in the rapid fixation of domestication-related genes while most unlinked loci of the genome retained the degree of diversity that was present before the initial domestication. This mechanism might account for the high genetic diversity of the nuclear microsatellites of the cultivars of *M. charantia*. Another possibility is that crops restored diversity through recent gene flow from wild populations after domestication [44]. We are not sure whether the large genetic diversity was contributed by the introgression from wild populations to cultivars, either at the initial stages of domestication or recurrently because of the lack of wild inland samples of *M. charantia*. However, the assignment tests by Bayesian clustering analysis (Figure 4), which infers the recent gene flow, revealed genetic admixture between the wild population of Taiwan and cultivars. The gene flow from the wild population of Taiwan to the cultivars is inferred to be rare ( $m_{T \rightarrow C} = 0.01$ ) by the IM model, and the hypothesis of no migration ( $m_{T \rightarrow C} = 0$ ) cannot be rejected (Table 4). On the other hand, the opposite direction of gene flow is higher ( $m_{C \rightarrow T} = 9.95$ ), suggesting that the genetic admixture is a result of strong introgression from cultivars to wild populations of Taiwan. This inference is also supported by the analyses of historical and current gene flow by using the MIGRATE-n [34] and BayesAss [35], respectively, which show a continuing asymmetric introgression from cultivars to the wild population of Taiwan. The high migration rates indicate the lack of reproductive isolation between two accessions and different capacities to receive foreign genes. The secondary contact by human mediated introduction (e.g., cultivation) would intensify such asymmetric gene flow, which could be occurring because of the increase in the recent migration rates (Figure S1), although estimates of the times of gene flow are sometimes inaccurate [45,46].

## 4. Experimental Methods

### 4.1. Sample Preparation

Seeds or adult leaves of both *M. charantia* cultivars and wild populations were collected from the field by our investigators or by exchange with cooperating laboratories. Most of the cultivars were provided by the National Plant Genetic Resources Center of the Taiwan Agricultural Research Institute, Council of Agriculture of Taiwan. The lines of varieties are listed in Table 6. Five to eight individuals of each wild population from Taiwan, India, and Korea and the white and green landraces of *M. charantia* were grown or collected from freshly matured leaves, purified genomic DNA, sequenced the *atpB-rbcL* intergenic spacer of cpDNA, and genotyped SSR, independently, to test the homogenization. In total, 164 individuals were tested on this study and listed in Table 6. Soaked seeds of *Momordica* lines were incubated for four days in 30 °C to allow them to germinate. Subsequently, plants were cultivated in a thermostatic chamber in 25 °C for two weeks and then transplanted into greenhouse conditions for the follow-up study. Freshly matured leaves were collected for molecular experiments.

**Table 6.** Resources and accessions of the *Momordica charantia* in this study.

Species	Strain	Accessions	Sample size	Resource <sup>a, b</sup>	Abbreviation
<i>M. charantia</i> (white cultivar)	Pai Bitter Gourd	TVI006693	6	NPGRC, TARI, COA	BM1
<i>M. charantia</i> (white cultivar)	Pin Tung He Tzu Ku Kua No. 8	TVI007217	6	NPGRC, TARI, COA	BM2
<i>M. charantia</i> (white cultivar)	Pai Pi Bitter Gourd	TVI007364	6	NPGRC, TARI, COA	BM3
<i>M. charantia</i> (white cultivar)	Chin Lien Bitter Gourd	TVI007723	6	NPGRC, TARI, COA	BM4
<i>M. charantia</i> (white cultivar)	Pin Tung He Tzu Bitter Gourd	TVI007895	6	NPGRC, TARI, COA	BM5
<i>M. charantia</i> (white cultivar)	Ming Hua	TVI006970	6	NPGRC, TARI, COA	BM8
<i>M. charantia</i> (white cultivar)	Lin Nei Tzu Liu Chung	TVI009016	5	NPGRC, TARI, COA	BM12
<i>M. charantia</i> (white cultivar)	Chiang Men Ta Ting Bitter Gourd	TVI010175	6	NPGRC, TARI, COA	BM16
<i>M. charantia</i> (white cultivar)	Ping Tung Li Kang Bitter Gourd	TVI009615	5	NPGRC, TARI, COA	BM18
<i>M. charantia</i> (white cultivar)	Small Bitter Melon	TVI009995	5	NPGRC, TARI, COA	BM20
<i>M. charantia</i> (white cultivar)	Chen Chu	TVI006900	6	NPGRC, TARI, COA	BM24
<i>M. charantia</i> (white cultivar)	Kuang Han Te Ta Chang Pai Bitter Gourd	TVI009892	5	NPGRC, TARI, COA	BM27
<i>M. charantia</i> (white cultivar)	Pai Pi Tsu Mi Bitter Gourd	TVI009534	5	NPGRC, TARI, COA	BM31
<i>M. charantia</i> (white cultivar)	Ping Tung Bitter Gourd	TVI007365	5	NPGRC, TARI, COA	BM33
<i>M. charantia</i> (white cultivar)	Chinese Gui-Nin No. 2	SYSU-BM-1	5	Exchange from South China Botanical Garden, China.	CBM2
<i>M. charantia</i> (green cultivar)	Chin Pi Bitter Gourd	TVI008603	6	NPGRC, TARI, COA	BM22
<i>M. charantia</i> (green cultivar)	Chin Pi Bitter Gourd	SYSU-BM-2	6	Exchange from Kunming Institute of Botany, China.	BBM
<i>M. charantia</i> (green cultivar)	Kao Mien Bitter Gourd	TVI009317	5	NPGRC, TARI, COA	FBM4
<i>M. charantia</i> (wild population from Korea)	Han Ch'eng K'u Kua	CN93MJF135	5	NPGRC, TARI, COA	FBM1
<i>M. charantia</i> (wild population from India)	<i>Momordica charantia</i> variety	TVI009464	5	NPGRC, TARI, COA	CBM1
<i>M. charantia</i> (wild population from Taiwan)	Ping Tung Chiu Chow Bitter Gourd	SYSU-BW-1	8	N 22°39'14", E 120°38'10", 205 m altitude	BMW1
<i>M. charantia</i> (wild population from Taiwan)	Nan Tou Bitter Gourd	SYSU-BW-2	8	N 23°55'06", E 120°53'04", 700 m altitude	BMW2

Table 6. Cont.

Species	Strain	Accessions	Sample size	Resource <sup>a, b</sup>	Abbreviation
<i>M. charantia</i> (wild population from Taiwan)	Yeh Sheng Bitter Gourd	TVI009560	5	NPGRC, TARI, COA	BMW5
<i>M. charantia</i> (wild population from Taiwan)	Hua Lien Yeh Sheng Bitter Melon No. 1	TVI009930	5	NPGRC, TARI, COA	BMW9
<i>M. charantia</i> (wild population from Taiwan)	Hua Lien Yeh Sheng Bitter Melon No. 5	TVI009934	5	NPGRC, TARI, COA	BMW10
<i>M. charantia</i> (wild population from Taiwan)	Hua Lien Yeh Sheng Bitter Melon No. 10	TVI009959	5	NPGRC, TARI, COA	BMW11
<i>M. charantia</i> (wild population from Taiwan)	Ping Tung Bitter Gourd	SYSU-BW-3	8	N 22°24'57", E 120°39'57", 686 m altitude	BMW12
Hybrid between wild population and cultivars ( <i>H</i> )	Hua Lien Bitter Gourd No. 2	HB002	5	HDARES, COA	H2
Hybrid between wild population and cultivars ( <i>H</i> )	Hua Lien Bitter Gourd No. 15	HB015	5	HDARES, COA	H15

<sup>a</sup> NPGRC, TARI, COA: The National Plant Genetic Resources Center, Taiwan Agricultural Research Institute, Council of Agriculture of Taiwan, respectively. <sup>b</sup> HDARES, COA: Hualien District Agricultural Research and Extension Station, Council of Agriculture of Taiwan.

#### 4.2. Molecular Techniques

Total DNA was extracted from leaves according to the protocol of the Plant Genomic DNA Extraction Miniprep System (Viogene, Taipei, Taiwan). Following the commercial protocol, 100 mg fresh leaves from each individual were used for purifying genomic DNA. Totally, extracted genomic DNA from 164 individuals was dissolved in 200 µL of TE buffer (pH 8.0) and stored at -20 °C for further experiments. Seventy-two primer sets developed from *Cucumis melo* L. [30] were tested for our samples by PCR, isolation, and sequencing and 12 microsatellite loci with perfect tandem repeat sequences were found. Eleven microsatellite primer sets, which totally extract 12 microsatellite loci, were successfully amplified in all cultivated and wild samples of *M. charantia* (Table 2). The thermocycling profiles were as follows: initial denaturing at 94 °C for 5 min, followed by 30 cycles of 40 s for denaturation at 94 °C, 1 min annealing at the optimized annealing temperature (Table 2), 1 min extension at 72 °C, and a subsequent final extension for 10 min at 72 °C. We chose certain amplicons for direct sequencing to confirm the polymorphic patterns of tandem repeats. The other amplicons were scored using an ABI 3730 for genotyping, and GeneMapper 3.7 software (Applied Biosystems, USA) was used for fragment analysis. The PCR amplification of sequences of the cpDNA *atpB-rbcL* intergenic spacer was performed using the Gene Amp<sup>®</sup> PCR System 9700 (Applied Biosystems, CA, USA) and the primers *atpB*-107 and *rbcL*-188 were derived from Chiang *et al.* [47]. The PCR mixture (50 µL) contained 500 mM KCl, 15 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 100 mM Tris-HCl (pH 8.3), 2 mM dNTPs, 2 µM primers (forward and reverse), 20 ng template DNA, 1 µg RNase, and 0.5 U Taq polymerase (Clontech Laboratories, Inc., CA, USA). The program for amplifying *atpB-rbcL* was as follows: initial denaturation at 94 °C for 5 min followed by 35 cycles of 40 sec at 94 °C, 1 min annealing at 55 °C, 1 min at 72 °C and a subsequent 10 min final extension at 72 °C. The purified PCR products were sequenced in both directions using a BigDye<sup>®</sup> Terminator v 3.1 Cycle Sequencing Kit (Applied Biosystems, CA, USA), the ABI 377XL automated sequencer

(Applied Biosystems, CA, USA) and the ABI PRISM<sup>®</sup> 3700 DNA Sequencer (Applied Biosystems, CA, USA) at the Genomics BioSci & Tech. Co., Ltd., Taipei, Taiwan.

### 4.3. Data Analyses

#### 4.3.1. Genetic Diversity and Neutrality Tests

The amplified cpDNA sequences from 164 individuals were aligned directly by ClustalX [48] and edited for further analyses using BioEdit ver. 5.0.6 [49]. Chlorotype sequences were generated using DnaSP ver. 4.0 [50]. The average number of mutations between pairs ( $\pi$ ) and the index of genetic diversity estimated by segregating sites ( $\theta_w$ ) were calculated using DnaSP ver. 4.0 [50]. The Tajima's  $D$  neutrality tests [51], estimated by calculating  $\pi$  and  $\theta_w$ , were performed using DnaSP ver. 4.0 [50].

#### 4.3.2. Phylogenetic Analyses of Chloroplast DNA

Neighbor-joining phylogeny and Bayesian phylogenetic relationships of chlorotypes were constructed using MEGA v. 5.05 [52]. The settings for the MCL model [53,54] for substitutions and pairwise deletion for indels, and the amount of support for monophyly evaluated by 1000 replicates in bootstrap resampling were used for the NJ tree reconstruction. For estimating the coalescent time of the cultivars and the wild samples of *M. charantia*, the phylogenetic Bayesian Markov chain Monte Carlo simulations were performed using BEAST v. 1.6.1 [55]. The Markov chains were run for 10 million generations and were sampled every 1000 generations, with the first 1000 samples discarded as burn-in. The HKY model was used based on the evaluation by using the corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC) with the assistance of MEGA v. 5.05 [52]. The Yule's birth-rate model with gamma distribution was set for prior trees with a randomly generated starting tree. Three replicates were run and combined for estimating the divergence time and the TMRCA with a setting of the substitution rate as  $2 \times 10^{-9}$  per site per year [56].

#### 4.3.3. Neutrality Test by Detecting Outlier Loci for Microsatellite DNA

Before extending our analysis to genetic structure and speciation model tests, it was necessary to perform a neutrality test on the polymorphic microsatellite loci (*i.e.*, the loci with any repeat-number variation) to ensure the exclusion of the influence of natural selection. We used the method of Beaumont and Nichols [57] to identify the  $F_{ST}$  outlier for each locus individually. The idea of outlier loci is based on the hypothesis of extremely high  $F_{ST}$  between populations for positively selected loci (the positive outliers) than for the neutral loci and a reduced  $F_{ST}$  between populations for loci under balancing selection (the negative outliers). The  $F_{ST}$  outlier approach is performed by LOSITAN [58] using two strategies: (1) we ran LOSITAN once to calculate the mean  $F_{ST}$  by empirical data followed by forcing the simulation according to the mean  $F_{ST}$ ; (2) we ran LOSITAN as in strategy 1 but simulated by removing the loci outside the 99% CI as recommended by Antao *et al.* [58]. The stepwise mutation model (SMM) was used for 100,000 simulations. Each strategy was run three times to obtain a converged inference for ensuring the accuracy of estimation.

#### 4.3.4. Genetic Diversity and Genetic Structure

The expected heterozygosity and index of genetic diversity ( $\theta$ ) of the “neutral” microsatellite loci were estimated using the stepwise mutation model (SMM) by Arlequin v. 3.0.1 [59]. The genetic structure between the two accessions was evaluated by an analysis of molecular variance (AMOVA) [60], a PCoA, and the assignment test by Markov chain Monte Carlo (MCMC) simulation performed by using the STRUCTURE program [31]. The candidate adaptive loci were excluded in these analyses. The AMOVA was performed with Arlequin v. 3.0.1 [59] based on SMM [61]. Based on the phylogenetic relationships reconstructed by cpDNA, two accessions (the cultivar accession and the Taiwan accession) were set. Between-accession and within-accession variations were estimated as a percentage of the total genetic variation. Statistical significance for evaluating fixation indices ( $R_{ST}$ ) from random variation (no genetic differentiation) was tested using 1000 permutations. PCoA was performed using GenAlEx 6.3 [62] in a covariance standardized setting. In the PCoA, polymorphic loci were treated as independent traits to estimate the eigenvalues of the switched axes. The three-dimensional PCoA plot was redrawn with the assistance of JMP v. 7.0 [63]. The genetic structure was further examined by using the Bayesian clustering method implemented in STRUCTURE v. 2.2 [31]. Both admixture and no admixture models were used. The posterior probability of the grouping number ( $K = 1\sim 10$ ) was estimated by using the MCMC method with 10 independent runs to evaluate the consistency of the results using 3 million steps with a 500,000-step burn-in for each run. Better grouping numbers were evaluated by log estimates of the posterior probability of  $K$ .

#### 4.3.5. Isolation with Migration (IM) Model Test

After the candidate loci under selection were excluded, the remaining eight neutral loci, which were detected in the mode of SMM, were used for the IM analysis [32]. We performed the IM analysis by using the IMA program [33], in which the nested models were tested by using the likelihood-ratio test (LRT) to test whether the speciation mode of divergence with gene flow is fitted to the cultivar accession and the Taiwan accession of *M. charantia*. Four specific questions were addressed to resolve questions concerning the split time, migrations (gene flow), and population sizes: (1) Are the cultivar accession and the Taiwan accession really divergent? (2) Is there a change in population size after their divergence? (3) Has any migration occurred since the divergence of two accessions? (4) How do the degrees of migration (gene flow) differ between the two accessions? The IMA uses the MCMC sampling strategy to simulate parameters (scaled by mutation rate) of population size ( $\theta_A = 4N_A\mu$ ;  $\theta_1 = 4N_1\mu$ ;  $\theta_2 = 4N_2\mu$ , where  $N$  is the effective population size, and  $\mu$  is the mutation rate), migration rate ( $M_1 = m_1/\mu$ ;  $M_2 = m_2/\mu$ , where  $m$  is the unscaled migration rate), and divergence time ( $t = T/\mu$ , where  $T$  is the unscaled divergence time). The model of molecular evolution employed was the SMM with an inheritance scalar of 1 for the nuclear microsatellite loci. MCMC runs were carried out with 100,000-step burn-in followed by 10 million iterations of the Markov chain. Because the mutation rate for the microsatellite loci might vary and cannot be determined, the default mutation rate ( $\mu = 1$ ) was set. Most effective sample sizes (ESS) for the microsatellite loci are never higher than 50; thus, the lack of trends in the parameter plot for  $t$  was used to determine whether convergence had been reached. Log-likelihood ratio (2LLR) tests were performed to test the significance of various nested models

(null models) implemented in the IMA. We used null models that included various combinations of equal population sizes of the ancestral ( $\theta_A$ ) and the current populations ( $\theta_C$  for the cultivar accession and  $\theta_T$  for the Taiwan accession), equal bidirectional migrations ( $m_{C \rightarrow T} = m_{T \rightarrow C}$ ), and a fixed value of zero for migrations (Table 4). The 2LLR values approximately fit a  $\chi^2$  distribution, and the associated probabilities were calculated with a  $\chi^2$  distribution, as recommended [33].

#### 4.3.6. Inferring the Past and Recent Introgression

The historical and recent gene flows were estimated with the assistance of the MIGRATE-n v. 3.0 [34] and the BayesAss programs [35], respectively. The MIGRATE-n was executed by sampling genealogies under assumptions of constant population size and migration rate over coalescent time to obtain the maximum likelihood of migration rate. The Brownian mutation model and the prior distribution of variable  $\theta$  with variable migration rate were incorporated into the Bayesian approach in the MIGRATE algorithm. Varying mutation rates among loci were adopted by empirical estimation (rates 0.4~1.8 per locus). The  $F_{ST}$  estimates were used as starting values for the initial analysis. For all other analyses, the ending parameters of the previous run were used as the starting values for the next run until results equilibrated at approximately the same values. We used the settings of three long chains with 15 million sampled, 15,000 recorded, and 25% discarded at the beginning of the chain. Adaptive heating was applied with temperature specifications of 1.00, 3.71, 9.14 and 20.00 to assure convergence, and the last three replicates were combined for estimation.

The pattern of recent gene flow between accessions was inferred by using the Bayesian inference with the assistance of BayesAss v. 1.3 [35]. This approach allows a dataset free from Hardy-Weinberg or migration-drift equilibrium. Three independent MCMC runs for  $10^7$  iterations with every 2000 iteration of sampling were performed, and the first 10% iterations were discarded as burn-in. Three delta parameters of allele frequency, migration, and inbreeding were set as 0.15, 0.15, and 0.20, respectively, based on the preliminary runs.

## 5. Conclusions

The unresolved genetic structure and the high migration rates between cultivars and the wild populations of Taiwan of *M. charantia* suggest incomplete divergence and a high capacity for gene exchange. In fact, artificial hybridization between the cultivars (as the pollen donor) and the wild populations of Taiwan has been carried out, named “Hua Lien Bitter Gourd No. 2” and “No. 15”, by the Hualien District Agricultural Research and Extension Station (<http://www.hdais.gov.tw/bred>). The successful artificial hybridization suggested the possibility of natural hybridization, especially for the crop-to-wild introgression by pollen flow. Different estimates of divergence time implied that the long-term duration of the speciation process might be caused by recurrent asymmetric gene flow, an idea that is supported by inferences of both historical and current rates of migrations. In fact, the crop-to-wild introgression is much more substantial than we had previously thought [64]. A serious genomic impact of introgression is the reduction of the reproductive fitness of wild individuals whose variants are replaced by “domesticated alleles” [65]. In our microsatellite study, two loci, which have extremely high (CMBR114-1) and extremely low divergence (CMBR31), were suggested to be evolving under directional selection and balancing selection, respectively. Although these two outliers

are most likely due to the hitchhiking effect, the results still indicated that certain loci were either naturally or artificially selected by the domestication process, resulting in fixation or loss of certain alleles. That is, the domestication-related alleles, good or bad, had been successfully incorporated into wild populations by introgression. Although the genetic assimilation might be relaxed by demographic swamping, the recurrent receiving of pollen from crops would speed progress towards gene replacement [24]. Therefore, the crop-to-wild introgression would cause a serious problem of losing native genes. This study provides an empirical case of such phenomenon of gene replacement by asymmetric introgression.

### Acknowledgments

We thank the National Plant Genetic Resources Center of the Taiwan Agricultural Research Institute, Council of Agriculture of Taiwan for giving seeds of most of the cultivars in this study. This work was supported by grants from the National Science Council in Taiwan awarded to Yu-Chung Chiang and Chang-Hung Chou, and partial grants from NSYSU.

### References

1. Marr, K.L.; Mei, X.Y.; Bhattarai, N.K. Allozyme, morphological and nutritional analysis bearing on the domestication of *Momordica charantia* L. (Cucurbitaceae). *Econ. Bot.* **2004**, *58*, 435–455.
2. Lu, A.; Jeffrey, C. *Momordica* Linnaeus. *Flora China* **2011**, *19*, 28–30.
3. Janick, J.; Paris, H.S. The cucurbit images (1515–1518) of the Villa Farnesina, Rome. *Ann. Bot.* **2006**, *97*, 165–176.
4. Joseph, J.K.; Antony, V.T. Ethnobotanical investigations in the genus *Momordica* L. in the southern Western Ghats of India. *Genet. Resour. Crop Evol.* **2008**, *55*, 713–721.
5. Wang, L.L.; Zhang, H.; Song, L.S.; Guo, X.M. Loss of allele diversity in introduced populations of the hermaphroditic bay scallop *Argopecten irradians*. *Aquaculture* **2007**, *271*, 252–259.
6. Whitt, S.R.; Wilson, L.M.; Tenailon, M.I.; Gaut, B.S.; Buckler, E.S. Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12959–12962.
7. Tripathi, U.N.; Chandra, D. Anti-hyperglycemic and anti-oxidative effect of aqueous extract of *Momordica charantia* pulp and *Trigonella foenum graecum* seed in alloxan-induced diabetic rats. *Indian J. Biochem. Biophys.* **2010**, *47*, 227–233.
8. Wu, S.J.; Ng, L.T. Antioxidant and free radical scavenging activities of wild bitter melon (*Momordica charantia* Linn. var. *abbreviata* Ser.) in Taiwan. *LWT Food Sci Technol.* **2008**, *41*, 323–330.
9. Horax, R.; Hettiarachchy, N.; Chen, P.Y. Extraction, quantification, and antioxidant activities of phenolics from pericarp and seeds of bitter melons (*Momordica charantia*) harvested at three maturity stages (immature, mature, and ripe). *J. Agric. Food Chem.* **2010**, *58*, 4428–4433.
10. Kobori, M.; Nakayama, H.; Fukushima, K.; Ohnishi-Kameyama, M.; Ono, H.; Fukushima, T.; Akimoto, Y.; Masumoto, S.; Yukizaki, C.; Hoshi, Y.; *et al.* Bitter gourd suppresses lipopolysaccharide-induced inflammatory responses. *J. Agric. Food Chem.* **2008**, *56*, 4004–4011.

11. Lii, C.K.; Chen, H.W.; Yun, W.T.; Liu, K.L. Suppressive effects of wild bitter gourd (*Momordica charantia* Linn. var. *abbreviata* Ser.) fruit extracts on inflammatory responses in RAW 264.7 macrophages. *J. Ethnopharmacol.* **2009**, *122*, 227–233.
12. Fang, Q.M.; Zhang, H.; Cao, Y.; Wang, C. Anti-inflammatory and free radical scavenging activities of ethanol extracts of three seeds used as “Bolenguazi”. *J. Ethnopharmacol.* **2007**, *114*, 61–65.
13. Kobori, M.; Amemiya, J.; Sakai, M.; Shiraki, M.; Sugishita, H.; Sakaue, N.; Hoshi, Y.; Yukizaki, C. Bitter gourd induces apoptosis in HL60 human leukemia cells and suppresses the production of inflammatory cytokine in RAW264.7 macrophage like cells. *J. Jpn. Soc. Food Sci. Technol.* **2006**, *53*, 408–415.
14. Abdollahi, M.; Zuki, A.B.Z.; Goh, Y.M.; Rezaeizadeh, A.; Noordin, M.M. The effects of *Momordica charantia* on the liver in streptozotocin-induced diabetes in neonatal rats. *Afr. J. Biotechnol.* **2010**, *9*, 5004–5012.
15. Yuan, X.Q.; Gu, X.H.; Tang, J. Optimization of the production of *Momordica charantia* L. var. *abbreviata* Ser. protein hydrolysates with hypoglycemic effect using Alcalase. *Food Chem.* **2008**, *111*, 340–344.
16. Schaefer, H.; Heibl, C.; Renner, S.S. Gourds afloat: A dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proc. R. Soc. B* **2009**, *276*, 843–851.
17. Schaefer, H.; Renner, S.S. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Mol. Phylogenet. Evol.* **2010**, *54*, 553–560.
18. Zhai, C.; Lin, F.; Dong, Z.; He, X.; Yuan, B.; Zeng, X.; Wang, L.; Pan, Q. The isolation and characterization of *Pik*, a rice blast resistance gene which emerged after rice domestication. *New Phytol.* **2011**, *189*, 321–334.
19. Ramsay, L.; Comadran, J.; Druka, A.; Marshall, D.F.; Thomas, W.T.; Macaulay, M.; Mackenzie, K.; Simpson, C.; Fuller, J.; Bonar, N.; *et al.* *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. *Nat. Genet.* **2011**, *43*, 169–172.
20. Wright, S.I.; Bi, I.V.; Schroeder, S.G.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S. The effects of artificial selection on the maize genome. *Science* **2005**, *308*, 1310–1314.
21. Frary, A.; Doganlar, S.; Daunay, M.C.; Tanksley, S.D. QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor. Appl. Genet.* **2003**, *107*, 359–370.
22. Doganlar, S.; Frary, A.; Daunay, M.C.; Lester, R.N.; Tanksley, S.D. Conservation of gene function in the solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* **2002**, *161*, 1713–1726.
23. Hancock, J.F. Contributions of domesticated plant studies to our understanding of plant evolution. *Ann. Bot.* **2005**, *96*, 953–963.
24. Haygood, R.; Ives, A.R.; Andow, D.A. Consequences of recurrent gene flow from crops to wild relatives. *Proc. R. Soc. Lond. B Biol. Sci.* **2003**, *270*, 1879–1886.
25. Teng, L.S. Geotectonic evolution of late Cenozoic arc-continent collision in Taiwan. *Tectonophysics* **1990**, *183*, 57–76.

26. Lu, C.Y.; Hsu, K.J. Tectonic evolution of the Taiwan mountain belt. *Petrol. Geol. Taiwan* **1992**, *27*, 21–46.
27. Wei, K.-Y. Environmental changes during the Late Quaternary in Taiwan and adjacent seas: An overview of recent results of the past decade (1990–2000). *West Pac. Earth Sci.* **2002**, *2*, 149–160.
28. Chiang, T.Y.; Schaal, B.A. Phylogeography of plants in Taiwan and the Ryukyu archipelago. *Taxon* **2006**, *55*, 31–41.
29. Edwards, S.V.; Beerli, P. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **2000**, *54*, 1839–1854.
30. Ritschel, P.S.; Lins, T.C.; Tristan, R.L.; Buso, G.S.; Buso, J.A.; Ferreira, M.E. Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). *BMC Plant Biol.* **2004**, *4*, doi:10.1186/1471-2229-4-9.
31. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
32. Hey, J.; Nielsen, R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **2004**, *167*, 747–760.
33. Hey, J.; Nielsen, R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2785–2790.
34. Beerli, P.; Palczewski, M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **2010**, *185*, 313–326.
35. Rannala, B.; Wilson, G.A. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **2003**, *163*, 1177–1191.
36. Stevison, L.S.; Kohn, M.H. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Mol. Ecol.* **2009**, *18*, 2457–2475.
37. Lin, S.M.; Chen, C.A.; Lue, K.Y. Molecular phylogeny and biogeography of the grass lizards genus *Takydromus* (Reptilia : Lacertidae) of East Asia. *Mol. Phylogenet. Evol.* **2002**, *22*, 276–288.
38. Chang, C.H.; Chen, J.H. Three new species of octothecate pheretimoid earthworms from Taiwan, with discussion on the biogeography of related species. *J. Nat. Hist.* **2005**, *39*, 1469–1482.
39. Long, Y.; Wan, H.; Yan, F.M.; Xu, C.R.; Lei, G.C.; Li, S.W.; Wang, R.J. Glacial effects on sequence divergence of mitochondrial COII of *Polyura eudamippus* (Lepidoptera: Nymphalidae) in China. *Biochem. Genet.* **2006**, *44*, 361–377.
40. Yin, W.; Fu, C.Z.; Guo, L.; He, Q.X.; Li, J.; Jin, B.S.; Wu, Q.H.; Li, B. Species delimitation and historical biogeography in the genus *Helice* (Brachyura: Varunidae) in the Northwestern Pacific. *Zool. Sci.* **2009**, *26*, 467–475.
41. Wu, C.I. The genic view of the process of speciation. *J. Evol. Biol.* **2001**, *14*, 851–865.
42. Eyre-Walker, A.; Gaut, R.L.; Hilton, H.; Feldman, D.L.; Gaut, B.S. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4441–4446.
43. Lawton-Rauh, A.; Burgos, N. Cultivated and weedy rice interactions and the domestication process. *Mol. Ecol.* **2010**, *19*, 3243–3245.
44. Glemin, S.; Bataillon, T. A comparative view of the evolution of grasses under domestication. *New Phytol.* **2009**, *183*, 273–290.

45. Strasburg, J.L.; Rieseberg, L.H. Interpreting the estimated timing of migration events between hybridizing species. *Mol. Ecol.* **2011**, *20*, 2353–2366.
46. Sousa, V.C.; Grelaud, A.; Hey, J. On the nonidentifiability of migration time estimates in isolation with migration models. *Mol. Ecol.* **2011**, *20*, 3956–3962.
47. Chiang, T.Y.; Schaal, B.A.; Peng, C.I. Universal primers for amplification and sequencing a noncoding spacer between the *atpB* and *rbcL* genes of chloroplast DNA. *Bot. Bull. Acad. Sin.* **1998**, *39*, 245–250.
48. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL\_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *25*, 4876–4882.
49. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
50. Rozas, J.; Sanchez-DelBarrio, J.C.; Messeguer, X.; Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **2003**, *19*, 2496–2497.
51. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595.
52. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739.
53. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526.
54. Tamura, K.; Nei, M.; Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11030–11035.
55. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, doi:10.1186/1471-2148-7-214.
56. Wolfe, K.H.; Li, W.-H.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9054–9058.
57. Beaumont, M.A.; Nichols, R.A. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B Biol. Sci.* **1996**, *263*, 1619–1626.
58. Antao, T.; Lopes, A.; Lopes, R.J.; Beja-Pereira, A.; Luikart, G. LOSITAN: A workbench to detect molecular adaptation based on a  $F_{ST}$ -outlier method. *BMC Bioinform.* **2008**, *9*, doi:10.1186/1471-2105-9-323.
59. Excoffier, L.; Laval, G.; Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **2005**, *1*, 47–50.
60. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491.
61. Weir, B.C.; Cockerham, C.C. Estimating  $F$ -statistics for the analysis of population structure. *Evolution* **1984**, *38*, 1358–1370.
62. Peakall, R.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295.
63. JMP, version 7; SAS Institute Inc.: Cary, NC, USA, 2007.

64. Ellstrand, N.C. Current knowledge of gene flow in plants: Implications for transgene flow. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 1163–1170.
65. Tang, H.B.; Sezen, U.; Paterson, A.H. Domestication and plant genomes. *Curr. Opin. Plant Biol.* **2010**, *13*, 160–166.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).