**Taiwania 64(2)**: 209-216, 2019
*DOI: 10.6165/tai.2019.64.209*

**NOTE**

# Development of transferable expressed sequence tag-simple sequence repeat (EST-SSR) markers for delimitating two recently diverged gingers endemic to Taiwan

**Yi-Ting TSENG, Min-Xin LUO, Bing-Hong HUANG, Pei-Chun LIAO***

*School of Life Science, National Taiwan Normal University, 88 Ting-Chow Rd., Sec. 4, Taipei, Taiwan, R.O.C.*
**Corresponding author's email: pcliao@ntnu.edu.tw; Tel: +886-2-77346330; Fax: +886-2-29312904*

(Manuscript received 24 March 2019; accepted 2 May 2019; online published 14 May 2019)

ABSTRACT: Species delimitation may be difficult for recently divergent species, particularly those with few distinguishable morphological characteristics. *Zingiber kawagoii* and *Z. shuanglongensis* are two recently divergent species endemic to Taiwan whose extremely similar vegetative characteristics hinder their distinction during non-flowering seasons. Given their recent divergence, the speciation process and introgressions between these two gingers warrant exploration. However, such studies are hampered by the absence of appropriate molecular markers. To solve this dilemma, we developed 20 transferable expressed sequence tag-simple sequence repeat loci (EST-SSR). Seven highly differentiated loci were further identified for rapid species delimitation. A preliminary test using discriminant analysis of principal components (DAPC) validated the effective discrimination of these EST-SSR loci, while Bayesian clustering analysis (BCA) revealed obvious introgression events between species, particularly on positive-outlier loci. These results imply adaptive introgressions between these species. However, more sampling and further experiments are necessary to validate this inference and resolve questions regarding the introgression and speciation mechanisms. The development of genetic markers in this study provides appropriate experimental conditions and a basis for further research.

KEY WORDS: Endemic; $F_{ST}$ outliers; island; Microsatellite DNA; Multilocus marker; Species delimitation; *Zingiber*.

## INTRODUCTION

Compared with geographic isolation, adaptive differentiation can produce more obvious genetic barriers in specific genes or linkage regions, which are called barrier loci (Ravinet *et al.*, 2017). During the initial stage of speciation, incomplete gamete incompatibility may cause hybridization or introgression between incipient species. Positively divergent selection may create new genomic barriers and provide an opportunity to maintain interspecific differentiation in sympatry (Bierne *et al.*, 2011). After a period of time, the genomic differences will gradually expand and eventually complete the speciation.

This phenomenon of heterogeneous divergent rates across the genome has been reported between two recently divergent island gingers endemic to Taiwan, *Zingiber kawagoii* Hayata and *Z. shuanglongensis* C. L. Yeh & S. W. Chung (Huang *et al.*, 2018). Transcriptomic evidence showed that 0.21% of expression genes (429/255711 transcripts) are differentially expressed between these species and are affected by positive divergent selection (Huang *et al.*, 2018). These low proportions of adaptive genes play key roles in maintaining species divergence (gamete incompatibility or hybridization disadvantage) (Comeault, 2018, Huang *et al.*, 2018). According to 53,686 exonic SNPs in the transcriptome, it is estimated that *Z. kawagoii* and *Z. shuanglongensis* differentiated less than 0.2 million years ago (Mya) (Huang *et al.*, 2018), considerably shorter than the formation of Taiwan (5~6 Mya (Sibuet and Hsu, 1997) or 2 Mya (Osozawa *et al.*, 2012)). Based on this evidence of positive selection, short divergence time, and overlapping distribution ranges, an alternative hypothesis of in situ divergence between species in response to similar environmental pressures (Huang *et al.*, 2018) was proposed against Yeh *et al.*'s (2012) hypothesis of independent originations of these two endemic gingers in southwestern China.

These two ginger species are morphologically distinguishable in reproductive characteristics by the size and color of bracts and labels, among other features (Lin, 2017). However, when not flowering, it is difficult to distinguish these two species from the leaves only. Lin (2017) indicated that these two species can be distinguished by the color of the transverse section of the rhizome (yellow in *Z. kawagoii* and purple in *Z. shuanglongensis*). However, we have found that the characteristics of the purple rhizome transverse section are not unique to *Z. shuanglongensis* but also appear in some populations of *Z. kawagoii*, with large variation even within populations (personal observation). These observations highlight the indistinguishability of vegetative characteristics between these species.

In the absence of a complete genome draft, multilocus

**Table 1** Information of the sampling populations, sample sizes, and results of assignment test by Bayesian clustering analysis.

| Population | Pop code | Lat (°N) | Lon (°E) | Alt (m) | *Z. kawagoii* | | | *Z. shuanglongensis* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *n* | Mismatch | | *n* | Mismatch | |
| | | | | | | neutral | adaptive | | neutral | adaptive |
| Da-Han Trail | DH | 22.41166 | 120.7338 | 642 | 6 | 0.129 | 0.138 | 10 | 0.007 | 0.046 |
| Dali-Datong Tribe, Taroko | N | 24.19224 | 121.6367 | 929 | 3 | 0.005 | 0.005 | - | - | - |
| Jin-Guang-Shan Trail | AG | 23.02981 | 120.5008 | 546 | 8 | 0.003 | 0.004 | - | - | - |
| Li-Jia Trail | AC | 22.80517 | 121.0318 | 989 | - | - | - | 9 | 0.007 | 0.042 |
| Mt. Du-Lan | DL | 22.89401 | 121.1861 | 1107 | - | - | - | 1 | 0.013 | 0.061 |
| Mt. Wei-Liao | AI | 22.87945 | 120.6455 | 784 | 8 | 0.008 | 0.005 | - | - | - |
| Ren-Lun Trail | RL | 23.7223 | 120.9026 | 1415 | 9 | 0.006 | 0.004 | 9 | 0.004 | 0.006 |
| Rui-Sui, Hualian | AB | 23.51116 | 121.3305 | 483 | 7 | 0.002 | 0.008 | - | - | - |
| Shuang-Liu Waterfall | S | 22.21405 | 120.7961 | 255 | 3 | 0.010 | 0.012 | - | - | - |
| Teng-Jhih | Y | 23.06425 | 120.7167 | 1308 | - | - | - | 9 | 0.005 | 0.014 |
| Entrance of Jin-Shui-Ying Trail | JSY | 22.40758 | 120.7564 | 1460 | 3 | 0.003 | 0.006 | - | - | - |
| Yi-Ma Trail | YM | 22.61459 | 120.9537 | 608 | - | - | - | 9 | 0.006 | 0.004 |
| Total Population | | | | | 47 | 0.021 | 0.023 | 47 | 0.006 | 0.024 |

*n*, sample size; Mismatch, the proportion of genetic components assigned to another species estimated from the *Q*-matrix of STRUCTURE by neutral and adaptive loci.

genome scans are commonly used to screen loci with extremely high population genetic differentiation index values (i.e. $F_{ST}$ outliers), which are considered potential adaptive divergence loci (Barrett and Hoekstra, 2011). Because these outlier loci may be related or linked to traits of adaptive speciation, they are suitable as genetic markers for species delimitation, particularly for recently divergent species. The characteristic of codominance leads the expressed sequence tag-simple sequence repeats (EST-SSRs) to bean appropriate and commonly used multilocus marker for the speciation and introgression studies. For example, it was used for determining the direction of introgression and identifying the species boundary of two parapatric pine species *Pinus massoniana* and *P. hwangshanensis* in eastern China (Zhang *et al.*, 2014); similar study has also been applied to the introgression issue between two walnut species (*Juglans regia* and *J. sigillata*) in southwestern China (Yuan *et al.*, 2018). Moreover, EST-SSRs on functionally annotated genes can even be used for studies of adaptive introgression. A beautiful example is that a positive outlier *CONSTANT*-like gene (*COL*) is considered to be adaptively divergent in drought tolerance between two oaks (*Quercus ellipsoidalis* and *Q. rubra*) in Kansas, USA, while the introgression of *COL* in the contact zone improves the adaptability to changing environmental conditions in both species (Khodwekar and Gailing, 2017). These studies indicate the importance of developing transferable EST-SSRs in the study of identifying species boundary, adaptation, and introgression in phylogenetically close species.

In this study, we intend to develop transferable EST-SSRs for *Z. kawagoii* and *Z. shuanglongensis*. Their highly transferable, polymorphic, and codominant characteristics make EST-SSRs appropriate for future studies on speciation mechanisms. We also hope to screen highly differentiated loci of the two species from these EST-SSRs, not only for genetic identification but also for future applications in hybridization and introgression research.

## MATERIALS AND METHODS

### *Sampling*

In total, 94 individuals were collected in this study, including 47 samples of *Z. kawagoii* from eight populations and 47 samples of *Z. shuanglongensis* from six populations (Table 1). Fresh leaves were collected and immediately dried in silica gel to prevent DNA degradation before further genetic experiments. Voucher specimens were collected from each population and stored in the laboratory of Population Genetics and Molecular Ecology at National Taiwan Normal University, Taipei, Taiwan.

### *Development of transferable EST-SSR primers*

Primers for EST-SSR loci were developed from contigs of transcriptomic data of both species (NCBI Bioproject accession number PRJNA437070) (Huang *et al.*, 2018). The assembly and expression profiles are available at Mendeley Data (https://data.mendeley.com; doi:10.17632/mxfgjzyxm9.1). Both transcriptomic contig sequence datasets were set as local libraries in BLAST-n searches reciprocally to search for SSR-containing contigs in SciRoKo version 3.4 (Kofler *et al.*, 2007). To facilitate the acquisition of polymorphic and transferable loci, we used the following criteria for choosing candidate loci in silico: (1) common sequences between both species (e-value <0.001), (2) different repeat numbers between species, (3) single BLAST hit sequences, (4) repeat numbers >6, (5) motif length >3, (6) GC content 40~60%, (7) exclusion of loci with A/T combinations only, and (8) predicted length of the PCR product >100 bp and <500 bp (best between 200 bp and 300 bp). Primers were designed in the SSR flanking regions, and the region for primer design was a minimum of 50 bp from the repeat motifs.

The designed primers were used to amplify candidate SSR loci from a few samples to examine the amplifiable and transferable SSR loci. PCR products were checked by 1% agarose gel electrophoresis. The amplifiable loci were chosen for genotyping.

***Genotyping***

We performed genotyping by PCR with the designed primers labeled with the fluorescent dyes FAM, VIC, PET, and NED and analyzed the lengths of the PCR products by electrophoresis in a fluorescence detection system. We adopted an economic method for the cost-effective usage of fluorescent dye. By adding an M13 tag (5'-TGTAAAACGACGGCCAGT-3') to the forward primer 5' tail, a universal M13 primer with 5' labeled fluorescent dye could be incorporated into the PCR products (Schuelke, 2000). The final volume of each PCR mixture was 20 μL and contained 10 pmol of each primer, 0.2 nmol of dNTP and 1 U of Taq DNA polymerase (Bernardo, Taiwan). The amplification procedure was as follows: initial denaturation for 3 min at 94°C; 35 cycles of denaturation of 94°C for 30 s, annealing at Tm for 30 s, and extension at 72°C for 30 s; and a final extension at 72°C for 5 min. The PCR products with different fluorescent dyes were pooled for subsequent analysis by capillary electrophoresis on an Applied Biosystems 3730 DNA Analyzer (Applied Biosystem, USA).

The fragment size was analyzed by Peak Scanner version 1.0 (Applied Biosystems, USA) at the National Center for Genome Medicine, Academia Sinica, Taiwan, and determined by the size standard ABI GS500 LIZ (Applied Biosystem, USA). For peak picking and noise reduction, a minimum peak height of 100 was adopted for allelic calling. Those peaks with a size falling within expectations were manually checked and adjusted.

***Data analyses***

The genotyping data were used to calculate the genetic diversity at the species level in terms of the observed and expected heterozygosity (*Ho* and *He*), inbreeding coefficient (*F*), diversity index $\theta_H$ (= $4N\mu$ for diploids, where $N$ is the effective population size and $\mu$ is the mutation rate) (Ohta and Kimura, 1973), and $M$ ratio of the Garza-Williamson (G-W) statistic ($M = k/r$, where $k$ is the number of alleles at a given locus and $r$ is the is the allelic range + 1) (Garza and Williamson, 2001). All of the genetic diversity indices were estimated using Arlequin v. 3.5.1.3 (Excoffier and Lischer, 2010).

Bayesian clustering analysis (BCA) and discriminant analysis of principal components (DAPC) were performed to examine whether these transferable SSR loci could delimitate species effectively. BCA is a population model-based approach under the assumption of Hardy-Weinberg and linkage equilibria (Falush *et al.*, 2003), while DAPC uses partial synthetic variables to minimize variation within groups without a priori grouping (Manel *et al.*, 2005, Jombart *et al.*, 2010). We performed DAPC with the R (R Core Team, 2015) package adegenet (Jombart, 2008) and conducted BCA using STRUCTURE 2.3.4 (Hubisz *et al.*, 2009). For DAPC, we performed $10^6$ runs of the $k$-means algorithm to obtain the optimal number of PCs (i.e. $a$-score optimization), and all given genetic variables were set as unscaled (scale = FALSE). Two axes of linear discriminants in DA were retained. For STRUCTURE, we estimated the posterior probability of the grouping number ($K$ = 1~20) by 10 independent runs using $10^6$ steps of Markov chain Monte Carlo (MCMC) replicates after 10% burn-in for each run to evaluate consistency. The best grouping number was evaluated by $\Delta K$ (Evanno *et al.*, 2005) in STRUCTURE HARVESTER ver. 0.6.94 (Earl and Vonholdt, 2012). We then determined the neutral and outlier loci with the fdist approach by LOSITAN (Antao *et al.*, 2008). The positive outliers, as defined by an observed $F_{ST}$ to He ratio greater than the upper bound of the 95% CI of $10^6$ simulations, were used for re-conducting the BCA and DAPC.

## RESULTS AND DISCUSSION

***Primer selection and genetic diversity***

Among the 51 designed primer pairs, 24 primer pairs were amplifiable in both species and polymorphic between species. We chose 20 loci for further genetic diversity analyses, of which 20 and 17 loci were polymorphic in *Z. kawagoii* and *Z. shuanglongensis*, respectively (Table 2). There were 5.4 ± 2.1 and 6.3 ± 2.7 alleles per polymorphic locus on average in *Z. kawagoii* and *Z. shuanglongensis*, respectively. Overall, the genetic variation of *Z. kawagoii* was similar to that of *Z. shuanglongensis*, but the observed heterozygosity was slightly higher in *Z. kawagoii* (*Ho* = 0.167 ± 0.088) than in *Z. shuanglongensis* (*Ho* = 0.147 ± 0.110), while the expected heterozygosity was slightly higher in *Z. shuanglongensis* (*He* = 0.595 ± 0.187) than in *Z. kawagoii* (*He* = 0.539 ± 0.218). These values resulted in a higher estimate of the inbreeding coefficient for *Z. shuanglongensis* (*F* = 0.728 ± 0.162) than *Z. kawagoii* (*F* = 0.656 ± 0.205) (Table 2). The interspecific divergence estimated by $F_{ST}$ and Nei's unbiased genetic distance was 0.206 and 0.834, respectively.

Compared with the very low genetic variation (intraspecific $\pi$ = 0.00002 and 0 in *Z. kawagoii* and *Z. shuanglongensis*, respectively) and extremely high interspecific divergence (FST = 0.996~1.000) of chloroplast DNA (Huang *et al.*, 2018), EST-SSR obtained relatively higher intraspecific genetic variation and lower interspecific genetic differentiation. This result implies a lower lineage sorting rate and greater

**Table 2** Summary statistics of genetic diversity indices in 20 SSR loci for each and both ginger species.

| Locus | *Zingiber kawagoii* ($n$ = 47) | | | | | | *Zingiber shuanglongensis* ($n$ = 47) | | | | | | Total ($n$ = 94) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. alleles | *Ho* | *He* | $\theta_H$ | *M* | *F* | No. alleles | *Ho* | *He* | $\theta_H$ | *M* | *F* | No. alleles | *Ho* | *He* | $\theta_H$ | *M* | *F* |
| 30870 | 6 | 0.234 | 0.766 | 8.664 | 0.316 | 0.691 | 1 | 0 | 0 | N.A. | N.A. | N.A. | 6 | 0.117 | 0.656 | 3.721 | 0.316 | 0.822 |
| 29072 | 6 | 0.149 | 0.547 | 1.934 | 0.024 | 0.708 | 7 | 0.298 | 0.837 | 18.291 | 0.027 | 0.535 | 9 | 0.223 | 0.773 | 9.199 | 0.034 | 0.711 |
| 29724 | 3 | 0.255 | 0.511 | 1.594 | 0.158 | 0.495 | 6 | 0.106 | 0.677 | 4.278 | 0.018 | 0.834 | 7 | 0.181 | 0.721 | 5.906 | 0.021 | 0.749 |
| 28165 | 6 | 0.191 | 0.613 | 2.841 | 0.273 | 0.684 | 1 | 0 | 0 | N.A. | N.A. | N.A. | 6 | 0.096 | 0.639 | 3.339 | 0.273 | 0.850 |
| 28188 | 7 | 0.213 | 0.678 | 4.319 | 0.333 | 0.683 | 8 | 0.085 | 0.757 | 7.986 | 0.029 | 0.873 | 9 | 0.149 | 0.803 | 12.420 | 0.032 | 0.815 |
| 31488 | 8 | 0.170 | 0.803 | 12.416 | 0.320 | 0.786 | 7 | 0.234 | 0.625 | 3.047 | 0.241 | 0.621 | 10 | 0.202 | 0.835 | 17.794 | 0.345 | 0.758 |
| 29442 | 4 | 0.234 | 0.693 | 4.820 | 0.308 | 0.659 | 2 | 0.085 | 0.467 | 1.258 | 0.500 | 0.816 | 5 | 0.160 | 0.682 | 4.442 | 0.313 | 0.766 |
| 29958 | 7 | 0.128 | 0.644 | 3.435 | 0.206 | 0.800 | 8 | 0.064 | 0.809 | 13.168 | 0.024 | 0.906 | 11 | 0.096 | 0.784 | 10.171 | 0.033 | 0.878 |
| 28631 | 9 | 0.298 | 0.819 | 14.845 | 0.030 | 0.621 | 6 | 0.043 | 0.704 | 5.196 | 0.020 | 0.903 | 10 | 0.170 | 0.821 | 15.129 | 0.033 | 0.793 |
| 29134 | 9 | 0.128 | 0.681 | 4.416 | 0.024 | 0.802 | 6 | 0.191 | 0.652 | 3.635 | 0.316 | 0.703 | 11 | 0.160 | 0.671 | 4.120 | 0.029 | 0.762 |
| 29971 | 5 | 0.149 | 0.235 | 0.354 | 0.313 | 0.359 | 4 | 0.021 | 0.380 | 0.799 | 0.019 | 0.938 | 6 | 0.085 | 0.311 | 0.553 | 0.027 | 0.726 |
| 26647 | 4 | 0.128 | 0.514 | 1.618 | 0.571 | 0.749 | 6 | 0.128 | 0.618 | 2.921 | 0.316 | 0.791 | 7 | 0.128 | 0.570 | 2.202 | 0.368 | 0.776 |
| 27807 | 2 | 0 | 0.082 | 0.094 | 0.500 | 1.000 | 4 | 0.170 | 0.392 | 0.851 | 0.400 | 0.561 | 4 | 0.085 | 0.597 | 2.577 | 0.400 | 0.857 |
| 31154 | 7 | 0.085 | 0.521 | 1.683 | 0.018 | 0.824 | 12 | 0.106 | 0.677 | 4.298 | 0.030 | 0.729 | 14 | 0.096 | 0.775 | 9.392 | 0.034 | 0.876 |
| 13019 | 6 | 0.234 | 0.582 | 2.356 | 0.375 | 0.593 | 6 | 0.255 | 0.561 | 2.099 | 0.316 | 0.540 | 7 | 0.245 | 0.620 | 2.954 | 0.368 | 0.605 |
| 29842 | 6 | 0.362 | 0.684 | 4.502 | 0.273 | 0.465 | 4 | 0.447 | 0.655 | 3.706 | 0.400 | 0.311 | 6 | 0.404 | 0.761 | 8.225 | 0.273 | 0.469 |
| 31415 | 3 | 0.128 | 0.468 | 1.266 | 0.097 | 0.724 | 8 | 0.170 | 0.713 | 5.575 | 0.027 | 0.739 | 9 | 0.149 | 0.738 | 6.776 | 0.031 | 0.798 |
| 31989 | 2 | 0.021 | 0.021 | 0.022 | 0.500 | -0.011 | 6 | 0.085 | 0.528 | 1.745 | 0.024 | 0.827 | 7 | 0.053 | 0.317 | 0.570 | 0.028 | 0.832 |
| 10841 | 4 | 0.149 | 0.536 | 1.827 | 0.308 | 0.719 | 12 | 0.149 | 0.615 | 2.877 | 0.300 | 0.755 | 12 | 0.149 | 0.770 | 8.956 | 0.300 | 0.807 |
| 15990 | 4 | 0.085 | 0.383 | 0.815 | 0.250 | 0.776 | 1 | 0 | 0 | N.A. | N.A. | N.A. | 5 | 0.043 | 0.603 | 2.676 | 0.020 | 0.929 |
| Mean | 5.4 | 0.167 | 0.539 | 1.512 | 0.260 | 0.656 | 6.3 | 0.147 | 0.595 | 1.510 | 0.168 | 0.728 | 8.1 | 0.149 | 0.672 | 1.769 | 0.164 | 0.779 |
| s.d. | 2.1 | 0.088 | 0.218 | 3.884 | 0.164 | 0.205 | 2.7 | 0.110 | 0.187 | 4.453 | 0.175 | 0.162 | 2.6 | 0.080 | 0.146 | 4.662 | 0.155 | 0.099 |

*Ho,* observed heterozygosity; *He,* expected heterozygosity; $\theta_H$, an estimator of $\theta$ (=$4N\mu$ for diploids, where $N$ is the effective population size and $\mu$ is the mutation rate) obtained from the expected homozygosity in stationary populations under a pure stepwise mutation model (Ohta and Kimura, 1973); *M*, Garza-Williamson statistic for denoting a stationary (*M* close to 1) or bottleneck population (*M* close to 0) (Garza and Williamson, 2001); *F*, inbreeding coefficient.

retention of common ancestral polymorphisms in nuclear markers than in plastic DNA. This result is not surprising because the EST-SSRs were developed from conserved transcribed regions between the genomes, and this characteristic can be applicable in subsequent studies of introgression and evolution (L. Y. Zhang *et al.*, 2005).

With respect to the G-W statistic, the M ratio of *Z. kawagoii* (M = 0.260 ± 0.164) was slightly higher than that of *Z. shuanglongensis* (M = 0.168 ± 0.175), revealing a more obvious bottleneck effect in *Z. shuanglongensis* than in *Z. kawagoii* after their divergence (Table 3). This inference is consistent with the current narrower geographic distribution of *Z. shuanglongensis* compared with *Z. kawagoii* (Lin, 2017). *Zingiber shuanglongensis* is mostly distributed in southern and southeastern Taiwan (except the Ren-Lun Trail populations, Table 1), whereas *Z. kawagoii* is widely distributed at low altitudes in northern, western, and southern Taiwan, with a sparse distribution in the eastern region. In terms of the census population size and current distribution, the widespread *Z. kawagoii* implies a rapid spatial expansion after a bottleneck event. However, larger sample sizes from more populations are needed to verify this inference.

***Positive outliers***

By the fdist approach, we found seven loci with relatively high FST values (i.e. positive outliers) compared with the others (Fig. 1). These seven loci were

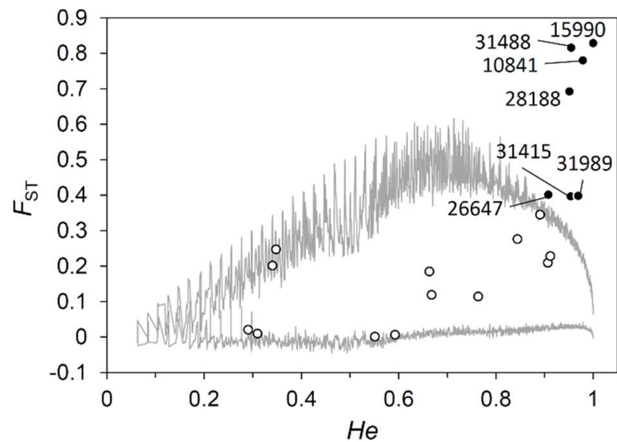


**Fig. 1** Determination of neutral and outlier loci based on Beaumont and Nichols's (1996) method. Loci located within the 95% confidence intervals were taken as neutral loci (open dots), and positive outliers (solid dots) were taken as candidate loci for species delimitation.

considered candidate loci for species delimitation. Positive outliers are usually considered positively selected genes (i.e. adaptive loci) that drive or are associated with species divergence (Mark A. Beaumont and Balding, 2004). Five of the seven adaptive loci were annotated to UniprotKB genes encoding Steroid 5-alpha-reductase DET2 (DET2_GOSHI, locus 31488), Zinc finger CCCH domain-containing protein 2 (C3H2_ORYSJ, locus 31415), RING-H2 finger protein ATL32 (ATL32_ARATH, locus 10841), Uridine kinase-like protein 2 (UKL2_ARATH, locus 31989),

**Table 2** Primer sequences and locus information, including annotation results

| Locus | Primer sequence | *Tm* | Motif | *L* | UniprotKB gene | Protein | BLAST-P | | BLAST-X | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | e-value | Bit score | e-value | Bit score |
| 30870 | F: M13-CGAACGTCTGCGTCATCGAC<br>R: CAAGAACAAGTACGGAGGTGAAAGG | 53 | CGC | 200 | MYC4_ARATH | Transcription factor MYC4 | 6.67E-136 | 412 | 2.63E-125 | 387 |
| 29072 | F: M13-GGTCAAGGGTTGGATCGAGC<br>R: GCATTTGATGAAGGAGGTTGCTTCG | 53 | CCT | 226 | - | - | - | - | - | - |
| 29724 | F: M13-GATCTCGGTGCAGAGGTACG<br>R: GAGACACGACGAAGACAGAGACG | 55 | GCT | 292 | CDPKO_ORYSJ | Calcium-dependent protein kinase 24 | 0 | 825 | 0 | 812 |
| 28165 | F: M13-CAGTGTTCTCTTCGTCAGATCC<br>R: GTAAACATCTCGGCACCAGATCC | 55 | TCT | 202 | RAP24_ARATH | Ethylene-responsive transcription factor RAP2-4 | 1.7E-62 | 204 | 3.55E-40 | 150 |
| 28188† | F: M13-GGAAGCCCAAGAAATCATGTC<br>R: CAGCTCAGAAAAGTCTTCCTTTACG | 57 | GAA | 238 | - | - | - | - | - | - |
| 31488† | F: M13-GTCTTCATCGCCGGTATGTTCG<br>R: CAGGGCCGGAGTATAGATCAGTAGC | 53 | CCT | 201 | DET2_GOSHI | Steroid 5-alpha-reductase DET2 | - | - | 9.33E-22 | 96.7 |
| 29442 | F: M13-GAAGCCCTGATGGGTTGATTCG<br>R: GCTGAGAATTCCGTGCTGAGAGC | 55 | TGC | 209 | BZP44_ARATH | bZIP transcription factor 44 | 3.3E-38 | 131 | 2.15E-29 | 115 |
| 29958 | F: M13-GGAATCCATTCCCTTCTCTCTC<br>R: CTGCCTGGAGGAATTCACTTGAC | 50 | ACC | 285 | GUN4C_ARATH | Tetrapyrrole-binding protein, chloroplastic | - | - | 1.18E-62 | 204 |
| 28631 | F: M13-GCTCGTCGCCCTTTATCCTC<br>R: GCTCCGACTCGATATTGAATGATCC | 50 | CTG | 258 | - | - | - | - | - | - |
| 29134 | F: M13-GCTGCTCAATTTTCCACTCC<br>R: CCAATAGTTGACCGACATGTGG | 57 | CTC | 341 | ERF1_SOLLC | Ethylene-responsive transcription factor 1 | - | - | 1.52E-42 | 154 |
| 29971 | F: M13-CTCGGGGAACTTACAGCTCTTG<br>R: CTGCCTAACTACATGAAGCCCACC | 57 | TTC | 186 | - | - | - | - | - | - |
| 26647† | F: M13-CCTCACCTGGAGAAGCTGAAC<br>R: GCACACACTCTGACATGCTTGC | 53 | GAA | 223 | - | - | - | - | - | - |
| 27807 | F: M13-CTGGGATTCATCAACGTCGTCG<br>R: GGACGACAACATGGAGCAACC | 53 | TCG | 244 | - | - | - | - | - | - |
| 31154 | F: M13-CGACTCGATCCTTAAAGTGG<br>R: CCTCTCCTTCCCTTTCTTCTTC | 53 | TCG | 359 | - | - | - | - | - | - |
| 13019 | F: M13-GCAGTTTTTCTGGCCTAAGG<br>R: CTTACCTTCCTCTCACTGTCC | 55 | TCT | 223 | IP5P3_ARATH | Type IV inositol polyphosphate 5-phosphatase 3 | - | - | 3.42E-20 | 92 |
| 29842 | F: M13-CCATCCCATCATAAGAGAAGC<br>R: GTCGGACTACATTATCTGATCTCC | 55 | GAA | 213 | SAT5_ARATH | Serine acetyltransferase 5 | - | - | 1.05E-85 | 279 |
| 31415† | F: M13-CGAGAGGCGTTGGTTGGTTATG<br>R: GTTTGAGTGCTGGCTCCATCC | 53 | CTG | 241 | C3H2_ORYSJ | Zinc finger CCCH domain-containing protein 2 | 5.76E-117 | 348 | 8.31E-95 | 298 |
| 31989† | F: M13-GAGAAATAGCGTGGAATCAAGG<br>R: GAAGGGGGAAGAGGAAGAGTAATC | 55 | CTC | 191 | UKL2_ARATH | Uridine kinase-like protein 2, chloroplastic | 0 | 773 | 0 | 704 |
| 10841† | F: M13-CTAACATGAGAATTCGCTGAGG<br>R: CTTCCACATCGAGTGCATTG | 55 | CGG | 198 | ATL32_ARATH | RING-H2 finger protein ATL32 | - | - | 1.21E-19 | 90.5 |
| 15990† | F: M13-CGATGACTCGGTGTCGTAGAG<br>R: CTCTGCTGCAAAGGAAGCCTTTC | 55 | CCG | 227 | IDD4_ARATH | Protein indeterminate-domain 4, chloroplastic | 5.14E-29 | 120 | 1.37E-20 | 97.8 |

†, positive-outlier loci; M13: TGTAAAACGACGGCCAGT; *Tm*, annealing temperature (°C); *L*, sequence length (bp) of the locus predicted by the contig of transcriptome data

and Protein indeterminate-domain 4 (IDD4_ARATH, locus 15990) (Table 2).

The proportion of detected positive outliers was quite high (7/20 = 35%, compared to the 0.21% positively selected transcripts estimated by Ka/Ks >1 (Huang *et al.*, 2018), where Ka and Ks are the nonsynonymous and synonymous substitution rates, respectively), and most of the annotated functions of these outlier loci seem unrelated to environmental pressures. False positives may increase if a species has undergone range expansion (Lotterhos and Whitlock, 2014), which may be the case for *Z. kawagoii*. The divergence of transcriptomes, the

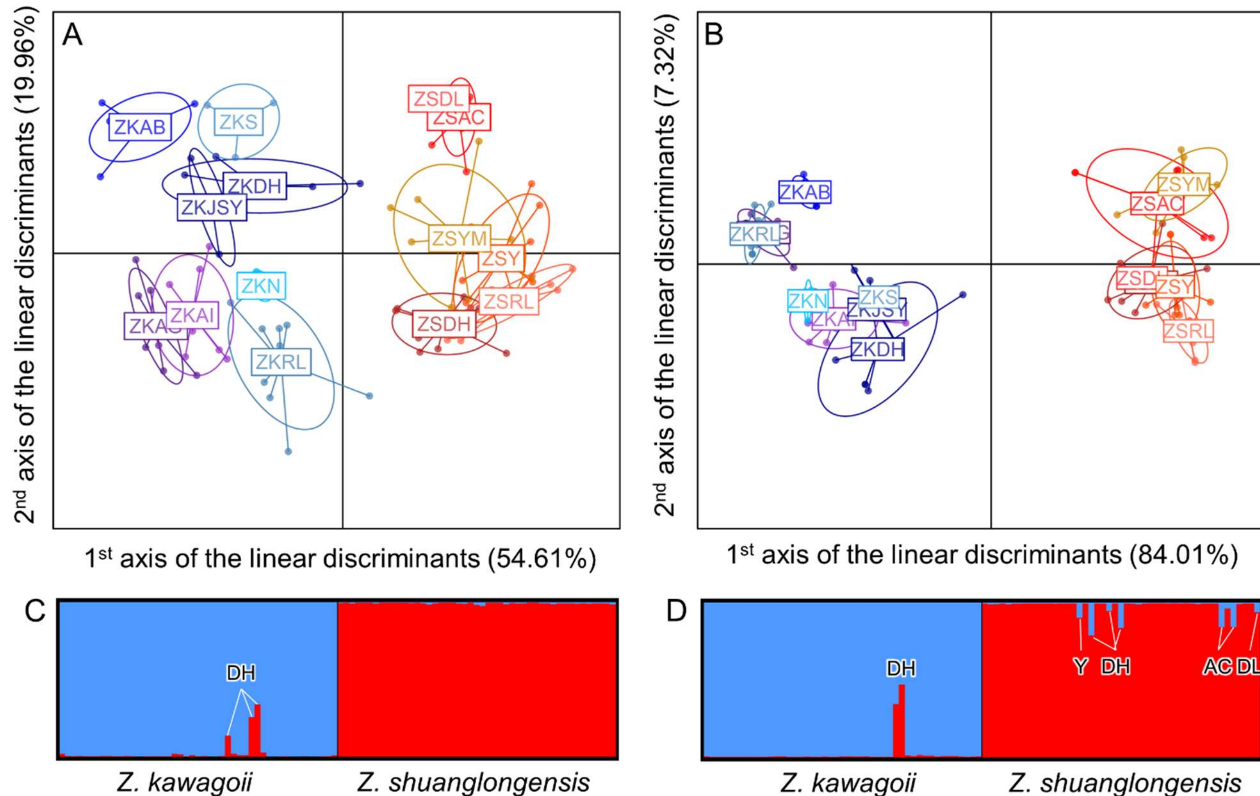


**Fig. 2** Genetic assignment between ***Zingiber shuanglongensis*** (red-series color) and ***Z. kawagoii*** (blue-series color) by discriminant analysis of principal components (DAPC) and Bayesian clustering analysis (BCA). **(A, B)** Scatterplots of DAPC revealing the grouping patterns of the sampled populations by neutral (**A**) and positive outlier loci (**B**). **(C, D)** Component plots of BCA drawing by the individual Q-matrix (indQ) estimated by neutral (**C**) and positive outlier loci (**D**).

source of EST-SSR, is shaped by both demographic change and selection (Blankers *et al.*, 2018). Recent demographic fluctuations may partly explain the high proportion of outliers. Although the selective pressure these outlier loci reflect remains unknown, these outlier loci are still useful for species delimitation.

### *Species delimitation by DAPC and BCA*

The neutral and positive outliers were used for both DAPC and BCA. In DAPC, six PCs were retained for two-axis DA based on the optimal a-score values in both the neutral and positive-outlier datasets. The first two axes of linear discriminants of DAPC explained 74.57% and 91.33% of the variations in neutral and positive-outlier loci, respectively (Fig. 2A and 2B). Unsurprisingly, the positive outliers had better discrimination than neutral loci in DAPC, although these two species could still be distinguished by neutral loci. However, the results of BCA were somewhat different. The best grouping number (K) of genetic components was two for both neutral and positive-outlier markers according to ΔK. These two genetic groups fit to the taxonomic units (species) well, but some introgression events were detected (Fig. 2C and 2D); by contrast, the introgression phenomenon was not obvious in DAPC, probably because the low explanatory PC dimensions (axes) were eliminated in DAPC. The elimination of these dimensions will reduce some effects that interfere with grouping, such as common ancestral polymorphisms and introgressions. By contrast, no genetic variables were eliminated in the calculation of the proportions of genetic composition in each individual (indQ) in BCA. Therefore, the inference of introgression detected by BCA was more obvious than that detected by DAPC.

Expectedly, introgression events should be less obvious in the BCA of positive outliers because of the higher FST of outlier loci. However, genetic admixture was more obvious in outlier loci than in neutral loci in BCA in this case, particularly for *Z. shuanglongensis* (Table 1, Fig. 2C and 2D). As a whole, *Z. kawagoii* has a higher proportion of genetic admixture than *Z. shuanglongensis* in neutral loci, but the main admixture comes from the Dahan Trail. However, in several populations of *Z. shuanglongensis*, the proportion of introgression was several times higher at the adaptive loci than that at the neutral loci (Table 1). When 5% heterospecific genetic components was considered as the threshold for defining introgression, two and seven introgressive individuals of *Z. kawagoii* and *Z.*

*shuanglongensis* were detected in the BCA of outlier loci, respectively, whereas only three samples of *Z. kawagoii* were detected in neutral loci (Fig. 2C and 2D). These introgression individuals were all sampled in southern Taiwan (Da-Han Trail, Li-Jia Trail, Mt. Du-Lan, and Teng-Jhih), where there are denser contact zones.

The higher frequency of genetic introgression in positive outliers than in neutral loci may indicate that the introgression is adaptive. Genetic introgression is the transfer of some genes to another species through hybridization and backcrossing. Adaptive introgression indicates the acquisition of higher adaptability through introgression (cf. Taylor and Larson, 2019). When the environment changes rapidly, species with smaller gene pools (such as island species) may withstand more pressure (Stork, 2010). The narrow but changeable terrain of Taiwan allows endemic plants, such as *Z. kawagoii* and *Z. shuanglongensis*, to survive stronger extinction threats than their continental relatives. Interspecific gene flow may help supplement genetic variation to increase the adaptive width and reduce the stress for survival (Alcala *et al.*, 2013, Bell, 2013). For example, Iris brevicaulis acquired flooding tolerance, which increased its survival rate, through introgression with I. fulva (Martin *et al.*, 2005, Martin *et al.*, 2006). Such a phenomenon of adaptive introgression represents genotype-dependent and condition-dependent dispersion rather than occurring comprehensively in the genome (Edelaar and Bolnick, 2012). Adaptive introgression may be beneficial to recently divergent species. Although it is still unclear what advantages *Z. kawagoii* and *Z. shuanglongensis* acquired via introgression, the discovery of adaptive introgression opens another avenue for the evolutionary study of these island-endemic gingers.

## CONCLUDING REMARKS

Although the number of populations and the sample size used for this preliminary test were small, we still selected seven highly interspecific-divergent loci from 20 transferable, polymorphic EST-SSR loci for rapid genetic discrimination between two morphologically similar ginger species, *Z. kawagoii* and *Z. shuanglongensis*. These markers revealed obvious introgressions in these samples, indicating partial gamete compatibility between these two species, and these codominant markers will be useful for related experiments on introgression. The highly variable EST-SSR markers developed in this study will be helpful for future studies of population genetics, interspecific differentiation, and hybridization/introgression.

## ACKNOWLEDGEMENTS

We thank Hsin-Pei Lu, Chien-Ti Chao, and Jui-Tse Chang for assistance with sampling. We deeply appreciate technical and bioinformatics support from the National Center for Genome Medicine of the National Core Facility Program for Biotechnology, Ministry of Science and Technology (MOST), Taiwan. This research was financially supported by a grant from MOST, Taiwan (MOST 105-2628-B-003-002-MY3), to PCL. This article was also subsidized by the National Taiwan Normal University.

## LITERATURE CITED

**Alcala, N., D. Streit, J. Goudet and S. Vuilleumier** 2013. Peak and persistent excess of genetic diversity following an abrupt migration increase. Genetics **193(3)**: 953-971.

**Antao, T., A. Lopes, R.J. Lopes, A. Beja-Pereira and G. Luikart** 2008. LOSITAN: A workbench to detect molecular adaptation based on a $F_{st}$-outlier method. BMC Bioinformatics **9(1)**: 323.

**Barrett, R.D. and H.E. Hoekstra** 2011. Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. **12(11)**: 767-780.

**Beaumont, M.A. and D.J. Balding** 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. **13(4)**: 969-980.

**Beaumont, M.A. and R.A. Nichols** 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. London Ser. B: Biol. Sc. **263(1377)**: 1619-1626.

**Bell, G.** 2013. Evolutionary rescue and the limits of adaptation. Philos. T. R. Soc. B **368(1610)**: 20120080.

**Bierne, N., J. Welch, E. Loire, F. Bonhomme and P. David** 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol. Ecol. **20(10)**: 2044-2072.

**Blankers, T., S.T. Vilaca, I. Waurick, D.A. Gray, R.M. Hennig, C.J. Mazzoni, F. Mayer and E.L. Berdan** 2018. Demography and selection shape transcriptomic divergence in field crickets. Evolution **72(3)**: 553-567.

**Comeault, A. A.** 2018. The genomic and ecological context of hybridization affects the probability that symmetrical incompatibilities drive hybrid speciation. Ecol. Evol. **8(5)**: 2926-2937.

**Earl, D.A. and B.M. Vonholdt** 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4(2)**: 359-361.

**Edelaar, P. and D.I. Bolnick** 2012. Non-random gene flow: an underappreciated force in evolution and ecology. Trends Ecol. Evol. **27(12)**: 659-665.

**Evanno, G., S. Regnaut and J. Goudet** 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14(8)**: 2611-2620.

**Excoffier, L. and H.E.L. Lischer** 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. **10(3)**: 564-567.

**Falush, D., M. Stephens and J.K. Pritchard** 2003. Inference of population structure using multilocus genotype data:

Linked loci and correlated allele frequencies. Genetics **164(4)**: 1567-1587.

**Garza, J.C. and E.G. Williamson** 2001. Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. **10(2)**: 305-318.

**Huang, B.-H., Y.-C. Lin, C.-W. Huang, H.-P. Lu, M.-X. Luo and P.-C. Liao** 2018. Differential genetic responses to the stress revealed the mutation-order adaptive divergence between two sympatric ginger species. BMC Genomics **19(1)**: 692.

**Hubisz, M. J., D. Falush, M. Stephens and J.K. Pritchard** 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. **9(5)**: 1322-1332.

**Jombart, T.** 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24(11)**: 1403-1405.

**Jombart, T., S. Devillard and F. Balloux** 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. **11(1)**: 94.

**Khodwekar, S. and O. Gailing** 2017. Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations. Am. J. Bot. **104(7)**: 1088-1098.

**Kofler, R., C. Schlotterer and T. Lelley** 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics **23(13)**: 1683-1685.

**Lin, Y.-C.** 2017. Taxonomic Study of *Zingiber* Mill. (Zingiberaceae) in Taiwan. Department of Forestry. **Master Degree:** 136, National Chung Hsing University. Taichung, Taiwan.

**Lotterhos, K.E. and M.C. Whitlock** 2014. Evaluation of demographic history and neutral parameterization on the performance of $F_{\mathrm{ST}}$ outlier tests. Mol. Ecol. **23(9)**: 2178-2192.

**Manel, S., O.E. Gaggiotti and R.S. Waples** 2005. Assignment methods: matching biological questions techniques with appropriate. Trends Ecol. Evol. **20(3)**: 136-142.

**Martin, N.H., A.C. Bouck and M.L. Arnold** 2005. Loci affecting long-term hybrid survivorship in Louisiana irises: Implications for reproductive isolation and introgression. Evolution **59(10)**: 2116-2124.

**Martin, N.H., A.C. Bouck and M.L. Arnold** 2006. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics **172(4)**: 2481-2489.

**Ohta, T. and M. Kimura** 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22(2)**: 201-204.

**Osozawa, S., R. Shinjo, A. Armid, Y. Watanabe, T. Horiguchi and J. Wakabayashi** 2012. Palaeogeographic reconstruction of the 1.55 Ma synchronous isolation of the Ryukyu Islands, Japan, and Taiwan and inflow of the Kuroshio warm current. Int. Geol. Rev. **54(12)**: 1369-1388.

**R Core Team** 2015. R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, Austria.

**Ravinet, M., R. Faria, R.K. Butlin, J. Galindo, N. Bierne, M. Rafajlovic, M.A.F. Noor, B. Mehlig and A.M. Westram** 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. **30(8)**: 1450-1477.

**Schuelke, M.** 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotech. **18(2)**: 233-234.

**Sibuet, J.-C. and S.-K. Hsu** 1997. Geodynamics of the Taiwan arc-arc collision. Tectonophysics **274(1-3)**: 221-251.

**Stork, N. E.** 2010. Re-assessing current extinction rates. Biodivers. Conserv. **19(2)**: 357-371.

**Taylor, S. A. and E. L. Larson** 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat. Ecol. Evol. **3(2)**: 170-177.

**Yeh, C.-L., S.-W. Chung, Y.-W. Kuo, T.-C. Hsu, C.-S. Leou, S.-J. Hong and C.-R. Yeh** 2012. A new species of *Zingiber* (Zingiberaceae) from Taiwan, China, based on morphological and molecular data. J. Syst. Evol. **50(2)**: 163-169.

**Yuan, X.-Y., Y.-W. Sun, X.-R. Bai, M. Dang, X.-J. Feng, S. Zulfiqar and P. Zhao** 2018. Population structure, genetic diversity, and gene introgression of two closely related walnuts (*Juglans regia* and *J. sigillata*) in Southwestern China revealed by EST-SSR markers. Forests **9(10)**: 646.

**Zhang, D., T. Xia, M. Yan, X. Dai, J. Xu, S. Li and T. Yin** 2014. Genetic introgression and species boundary of two geographically overlapping pine species revealed by molecular markers. PLOS ONE **9(6)**: e101106.

**Zhang, L. Y., M. Bernard, P. Leroy, C. Feuillet and P. Sourdille** 2005. High transferability of bread wheat EST-derived SSRs to other cereals. Theor. Appl. Genet. **111(4)**: 677-687.