Asia-Pacific Journal of Science and Technology: Volume: 26. Issue: 03. Article ID.: APST-26-03-07. Research Article

Asia-Pacific Journal of Science and Technology

https://www.tci-thaijo.org/index.php/APST/index

Published by the Research and Graduate Studies,
Khon Kaen University, Thailand

Genetic structure between wild and cultivated populations of *Carica papaya* in Thailand revealed by microsatellites

Nidchaya Aketarawong[1,*], Siriwan Isasawin[1] and Sujinda Thanaphum[1]

[1]Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand
[*]Corresponding author: nidchaya.akt@mahidol.ac.th

Received 1 March 2021
Revised 22 May 2021
Accepted 8 June 2021

Abstract

Evaluation of genetic variation and population structure between wilds and cultivars is important for understanding their origin and distribution, domestication, and genetic relationship. In this research, sixty-one papayas from eight wild populations and 52 additional samples from three cultivars in Thailand were studied using six polymorphic microsatellite loci. Higher genetic variation was observed in the wilds, especially from the South (mean number of alleles (N_a) = 2.333 to 2.833; mean observed heterozygosity (H_O) = 0.194 to 0.267; mean expected heterozygosity (H_E) = 0.481 to 0.526). This suggests that papayas from the southern part of Thailand could be a potential source population of the other local populations. Conversely, the cultivars have been maintained and selected over a long term. There were low levels of genetic variation (N_a = 1.167 to 1.500; H_O = 0.081 to 0.167; H_E = 0.086 to 0.183) with no private alleles. Cross-fertilisation was indirectly detected in both wilds and cultivars, possibly leading to retaining some polymorphisms. Lastly, genetic clustering of wilds and cultivars (K = 2) was identified; however, approximately 22.95% of wild samples (14 out of 61) were genetically admixed ($0.100 \leq Q_{\text{wilds}} \leq 0.900$) and 8.20% (five out of 61) belonged to the cultivar cluster ($Q_{\text{cultivars}} > 0.900$). This suggests that feral papayas could survive in a natural environment.

Keywords: Common papaya, Simple sequence repeats, Admixed structure, Gene flow, Risk assessment, Farm management

1. Introduction

The common papaya, *Carica papaya* Linn, is a fast-growing tree crop belonging to the family Caricaceae. Its breeding system can be either dioecious or gynodioecious (coexisting of female and hermaphrodite flowers) [1]. The papaya was introduced from a potential origin (East-Central or Southern America) to the Caribbean and Western India and arrived in the Philippines via the Portuguese and Spanish colonists in the 16th century. This crop was later introduced to Indonesia and nearby countries [1,2]. In Thailand, it is thought that the papaya was introduced via the east coast of the southern part during the late 18th century. Papaya is commonly found in both tropical and subtropical areas. Due to the beneficial qualities of papaya fruits and other substances (e.g., papain, pectin), papaya is widely cultivated throughout Southeast Asia and has become a key ingredient in famous and important dishes [2]. This plant is also cultivated on a large scale for local and international commerce as well as in backyard gardens and bordering paddy fields for small-scale consumption [3].

Papaya ringspot virus (PRSV) has been a major factor in reducing the yield of papaya plantations worldwide since 1949, and in Thailand since 1975 [3,4]. The virus reduces the photosynthetic ability of the upper papaya leaves. This infection results in a reduction of vegetative vigour, fruit quality and yields, and eventually the death of the plant. Naturally resistant varieties and insecticides were initially required for the control of the disease, but were proven ineffective [3-5]. An alternative approach using a genetic engineering technique was applied to develop a new papaya strain to defend against PRSV [4,5]. Although the papaya strain was approved for cultivation, and the output was appreciated in many areas such as the USA, Canada, China, and Japan [5], the genetically modified (GM) papaya and its product are not authorised to be imported into the European Union [6]. In Thailand, the GM issue is still being debated with regard to the economic promise and public acceptance because GM foods evoke environmental and agricultural concerns for farmers, consumers, and the environmental organisations [3-7]. One of the ecological and environmental concerns of GM crops is whether GM strains can establish themselves in the wild or outcross to their relative wild populations. Many studies have reported that transgenes probably escape to their wild relatives through the formation of transgenic hybrid populations [8-13]. This situation is riskier if they are in a temporal-spatial overlap or in close proximity [8,9].

Gene flow is an evolutionary mechanism to introduce genetic materials from one population of a species to another by breeding. This mechanism can spontaneously occur, at a rate of approximately 25%, between cultivars to wilds or their wild relatives through hybridisation, backcrossing, or selection [12,13]. These phenomena may lead to either a decrease or an increase in the degree of local adaptation in each population [9]. Gene flow plays an important role in addition to other evolutionary forces (e.g., selection, genetic drift) during cultivation and improvement. Farmers need to maintain desirable traits by human-mediated selection. It forces a genetic bottleneck, loss of genetic diversity, and has an impact on inbreeding depression and phenotypic plasticity of cultivars [14,15]. Outbreeding and/or multiple introductions between cultivars and wilds could maintain or increase genetic variability and subsequently reduce the risk of crop failure due to environmental changes [16]. At this point, understanding the genetic relationship and gene flow between wilds and cultivars would help to manage risks related to GM and agricultural practices. This information has been inadequately documented for papayas [17].

A better scientific understanding about the ecology and genetic interaction (i.e., genetic diversity, genetic structure, and gene flow) among plants is necessary for the risk management of environmental and agricultural problems. To address these issues, Simple Sequence Repeats (SSRs) or microsatellites are promising tool candidates. Microsatellites are widely distributed across the whole eukaryotic genome and their mutations cause a high proportion of polymorphisms. This marker system has been successfully applied to a wide variety of crops to infer genetic populations and detect natural hybridisation and introgression between cultivated crops and their wilds or their wild relatives (for examples in [18-22]).

In this study, we investigated two novel (from this study) and four additional highly polymorphic microsatellite loci [23] in order to study genetic viability, genetic structure, and demographical history among eight wild and three cultivated papaya populations from different geographical areas of Thailand. The results showed that the genetic variation of wilds was higher than cultivars as expected. Outbreeding was also indirectly detected in cultivars, not only in wilds. This would be a way to retain their polymorphisms. Lastly, evidence based on genetic clustering showed that genotypic cultivars differed from wilds. Also, they could survive in a natural environment as feral individuals. This genetic information could be used as a baseline for monitoring further farm management and risk assessment.

2. Materials and methods

2.1 Sampling collection

In this study, wild and cultivated samples were collected and analysed. A ‘wild’ was defined as a papaya tree that survived outside of a plantation. The habitats of these wild papayas included roadside ditches and abandoned paddy fields. Sixty-one samples from wild papaya trees were collected from a total of eight wild populations from four regions of Thailand: one from Phichit Province (PC-W) in the lower North, one from Nakhon Pathom Province (NK-W) in the Center, three from Khon Kaen Province (KK-W1, KK-W2, and KK-W3) in the Northeast, and three from Chumphon Province (CP-W1, CP-W2, and CP-W3) in the upper South (Table 1). Two wild populations (i.e., NK-W and KK-W1) had grown nearby (less than 3 km and 10 km, respectively) the previous experimental areas for plantations of genetically modified papayas. These two areas were closed due to a campaign by an environmental organisation in Thailand [6].

Fifty-two other papaya samples were collected from three popular cultivars (i.e., Khak Dam, Hawaii, and Eksotika) (Table 1). The cultivar Khak Dam was collected from Nakhon Pathom Province (NK-K), and the cultivars Hawaii and Eksotika were collected from the same orchard in Kanchanaburi Province (KC-H and KC-E, respectively).

2.2 Genomic DNA preparation, primer screening, and DNA genotyping

Total genomic DNA was extracted from 300 mg of fresh young papaya leaf using SabaiKit (Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand). Genomic DNA was prepared from *C. papaya* from the cultivar Khak Dam in Bangkok, Thailand. The method to enrich microsatellite DNA was described in [24]. Briefly, *Rsa*I-digested genomic DNA were ligated to the adaptor A (5’CTCTTGCTTACGCGTGGACTA) and the adaptor B (5’TAGTCCACGCGTAAGCAAGAGCAC), and used as a polymerase chain reaction (PCR) template. The PCR products were hybridised to a biotinylated

$(dCA)_{10}$ probe and selectively retained using streptavidin-paramagnetic particles (Promaga, Wisconsin, USA). The enriched DNA by PCR were ligated into the pGEM®-T Easy vector (Promega, Wisconsin, USA) and transformed into *Escherichia coli* strain DH5α competent cells. Sixty putative clones were sequenced using the ABI PRISM 3730XL automate DNA sequencer from Macrogen, Seoul, South Korea.

Fourteen microsatellite loci (m*Cp*1, m*Cp*2, m*Cp*3, m*Cp*4, m*Cp*5, m*Cp*6, m*Cp*7, m*Cp*8, m*Cp*9, m*Cp*10, m*Cp*11, m*Cp*12, m*Cp*13, and m*Cp*14) isolated from an enriched library of *C. papaya* in Thailand (this study) and five additional loci (m*Cp*CIR2, m*Cp*CIR3, m*Cp*CIR4, m*Cp*CIR7, and m*Cp*CIR9) isolated from *C. papaya* in Caribbean [23] were preliminary screened for four wild populations. Seven polymorphic loci (m*Cp*3, m*Cp*4, m*Cp*CIR2, m*Cp*CIR3, m*Cp*CIR4, m*Cp*CIR7, and m*Cp*CIR9) were subsequently used in analyses for wild and cultivated papayas.

DNA amplicon was carried out using a 15 µl final volume containing 100 ng of genomic DNA, 1 × buffer, 2.5 mM $MgCl_2$, 25 µM dNTP, 1 U *Taq* DNA polymerase (Promega, Wisconsin, USA) and 5 µM of each primer. PCR analyses were performed using a FlexCycler PCR machine (Analytik Jena, Thuringia, Germany) with the following conditions: 5 min at 95°C, 29 cycles of 30 s at 95°C, 90 s at T_a (see Table 2) and 90 s at 72°C, and 5 min of final elongation at 72°C. PCR products were electrophoresed following the method from [25]. Each gel included three lanes of 25 base pair DNA ladder (Promega, Wisconsin, USA) and a DNA fragment of known sequence. Allelic data were scored and the genotype of each sample was determined. An individual was declared a null (non-amplifying) allele at a locus only after at least two amplification failures.

Table 1 Wild and cultivated samples of *Carica papaya* L. collected from different geographical locations in Thailand.

Region	Province/ District	Type of sample	Variety	Population name	Latitude	Longitude	Sample sizes
Northern	Pichit/Sak Lek	Wild	Unidentified	PC-W	16°34'N	100°38'E	12
Central	Nakhon Pathom/ Phutthamonthon	Wild	Unidentified	NK-W	13°47'N	100°19'E	16
Northeastern	Khon Kaen/ Mueang Khon Kaen	Wild	Unidentified	KK-W1	16°26'N	102°50'E	5
	Khon Kaen/ Nong Ruea	Wild	Unidentified	KK-W2	16°30'N	102°32'E	6
	Khon Kaen/ Mancha Khiri	Wild	Unidentified	KK-W3	16°05'N	102°37'E	4
Southern	Chumphon/ Mueang Chumphon	Wild	Unidentified	CP-W1	10°29'N	99°10'E	7
	Chumphon/Tha Sae	Wild	Unidentified	CP-W2	10°39'N	99°10'E	6
	Chumphon/ Sawi	Wild	Unidentified	CP-W3	10°15'N	99°05'E	5
Central	Nakhon Pathom/ Phutthamonthon	Cultivar	Khak Dam	NK-K	13°48'N	100°15'E	18
Western	Kanchanaburi/ Sai Yok Noi	Cultivar	Hawaii	KC-H	14°12'N	99°03'E	16
	Kanchanaburi/ Sai Yok Noi	Cultivar	Malaysia	KC-M	14°12'N	99°03'E	18

2.3 Data analyses

2.3.1 Genetic variability

To avoid the contribution of selection and adaptation in shaping genetic variations among samples, candidate loci under selection were determined using two different programs: Lositan [26] and BayeScan 2.1 [27,28]. The Lositan program identifies outlier loci by evaluation of the relationship between F_{ST} and the expected heterozygosity (H_E) in an island model. The loci with excessively low or high F_{ST} compared to the neutral expectations were then subjected to selection. On the other hand, the BayeScan program identifies loci under natural selection using differences in allele frequencies between populations. The program uses a reversible-jump Markov Chain Monte Carlo (MCMA) approach to estimate posterior probabilities. It was set up as follows: 20 pilot runs with 5,000 iterations each, 50,000 steps of the initial burn-in, 5,000 iterations of the chains, and a thinning interval of 10. The threshold of a posterior P of > 0.99 and the logarithm (base 10) of the posterior odds ($\log_{10}$ PO) of ≥ 2 were used as decisive for accepting a selection model. Loci that were identified by both approaches were determined as probable directional outliers [29].

The genetic variability of each wild and cultivated population was estimated using mean number of alleles (N_a), mean number of private alleles (N_p), mean observed and expected heterozygosities (H_O and H_E), and inbreeding coefficient (F_{IS}) with GenAlEx v.6.501 [30]. Outcrossing rate (t) was estimated from the inbreeding coefficient using the equation: $t = (1-F_{IS})/(1+F_{IS})$ [31]. The number of effective alleles was estimated at

1/(1-H_E). The frequency of null alleles (r) was estimated according to [32]. Deviation from the Hardy-Weinberg equilibrium and linkage disequilibrium, together with their critical levels after the sequential Bonferroni correction [33], were tested using GENEPOP version 3.3 [34].

Table 2 Variation[1] at seven microsatellite loci in wild and cultivated *Carica papaya* L. collected from Thailand.

Locus		Size	T_a	Total	Wild (N = 61)			Cultivar (N = 52)			
[Repeat motif]	Primer (5'-3')	range	(°C)	N_a	N_a	H_O	H_E	N_a	H_O	H_E	F_{ST}
m*Cp*3	F: ATGGTAGGTGAAATCGTGGAGG	90-96	69	4	4	0.464	0.443	2	0.656	0.441	0.145
[GAA(GA)$_4$]	R: CGGACCTCTCTCACTCACTG										
m*Cp*4	F: GTTGTTAGATTTGTGCTTG	155-159	59	3	3	0.085	0.615	1	0.000	0.000	0.286
[TGTC(TG)$_7$]	R: AAACAAGAGTATCAGCAAG										
m*Cp*CIR2[2]	F: GTCTATCTACCTCCCA	282-318	52	10	10	0.000	0.702	2	0.000	0.453	0.185
[(TC)$_{24}$]	R: GAGTGTTATCATAGTCTACA										
m*Cp*CIR3[2*]	F: GAACTCACCTACACGAACT	178-202	50	5	5	0.016	0.602	1	0.000	0.000	0.463
[(TC)$_{14}$]	R: ACTTCTACCACCGGC										
m*Cp*CIR4[2]	F: AAGCCAAGAACAGCAA	162-172	51	5	5	0.085	0.098	2	0.077	0.473	0.178
[(TC)$_{10}$]	R: AATGCTTGAAGTAAACACC										
m*Cp*CIR7[2]	F: CCAACACATTCATCCAC	227-231	50	3	3	0.200	0.555	2	0.098	0.369	0.025
[(TC)$_{18}$]	R: CTGAAGCATTACCGAGA										
m*Cp*CIR9[2]	F: TAAAACCCTAACGAGCA	130-142	50	3	3	0.091	0.643	1	0.000	0.000	0.316
[(CT)$_9$]	R: CAAAGACAGACTTGGA										
Overall				33	33	0.134	0.522	11	0.119	0.248	0.228
SE						0.060	0.077		0.091	0.089	0.053

Note: [1] N (number of samples), N_a (number of alleles), H_O (observed heterozygosity), H_E (expected heterozygosity), F_{ST} (fixation index). [2] Primer from [23], [*] Locus was determined to be under directional selection using Lositan [26] and BayeScan [27,28].

2.3.2 Genetic structure

The degree of genetic differentiation among populations in terms of pairwise F_{ST} values [35] was estimated using MICROSATELLITE ANALYSIS (MSA) [36]. The statistical significance of each value with the values obtained in 10,000 matrix permutations was used.

A Bayesian Markov chain Monte Carlo (MCMC) approach, using STRUCTURE version 2.2 [37], was used to estimate the proportions of genetic admixture among the hypothetical clusters at the individual level. The program was run for the values of K ranging from 1 to 9. To determine the most likely number of clusters (K) underlying the 61 given samples, the log-likelihood probability [37] and the Delta K method [38] were used. The genetic admixture between wild and cultivated papayas was also estimated using prior population information in order to infer introgression between them. Five independent simulations were run for each value of K using the following settings: burn-in period = 100,000 iterations; number of MCMC replicates after burn-in = 500,000 repetitions; admixture model using the same alpha for all populations and a uniform prior; allele frequencies assumed to be correlated among populations.

A principal component analysis (PCoA) done in GenAlEx v.6.501 [30] was used to display genetic divergence among samples in multidimensional space using the allele frequency data. Analysis of molecular variance (AMOVA), as implemented in ARELEQUIN 3.5 [39], was performed to partition the observed genetic variability among the populations of wild papayas, as well as between wilds and cultivars.

2.3.3 Demographic analyses

The isolation by distance based on $F_{ST}/1\text{-}F_{ST}$ was analysed to infer the relationship between geographical and genetic distance using GENEPOP version 3.3 [34].

GENECLASS 2 software [40] was used to assign individuals to their original population or other populations. The parameters were set using the following criteria: The Bayesian method [41], the simulation algorithm using [42], number of simulation individuals of 10,000, and type I error of 0.05. In addition, the same software was used to detect the first-generation migrants (recent migration). The parameter for likelihood ratio was set as *L_home*/*L_max* and the other criteria as above.

2.4 Evaluation of risk assessment

The PRSV-resistant papaya DNA construct contained the CaMV 35s promoter for controlling the expression of the coat protein gene and the beta-glucuronidase (GUS) gene [4]. The designed primer sequences specific to the CaMV 35s promoter were used to test all papaya samples: forward 5'-GATTCCATTGCCCAACTATC-3' and reverse 5'-TACAGGACGTAACATAAGGG-3'. Each PCR mixture was carried out with a 15 μl final volume containing 100 ng of genomic DNA, 1 × buffer, 2.5 mM $MgCl_2$, 25 μM dNTPs, 1 U *Taq* DNA polymerase (Promega, Wisconsin, USA) and 5 μM of each primer. PCR analyses were performed using the FlexCycler PCR machine (Analytik Jena, Thuringia, Germany) with the following conditions: 5 min at 95°C, 29 cycles of 1 min at 95°C, 2 min at 47°C and 2 min at 72°C, and 5 min of final elongation at 72°C. PCR products were electrophoresed using 1% agarose gel and stained with ethidium bromide. Amplicons were compared with the 100 base pair DNA ladder (Promega, Wisconsin, USA) and positive control (template DNA is a DNA construct contained in the CaMV promoter).

3. Results

3.1 The enrichment of microsatellite DNA and polymorphic loci screening

Twenty-four of the 60 total clones (40%) contained microsatellite sequences. After characterisation, only fourteen loci (Accession number MZ484098-MZ484111) presented expected amplicons. However, only two loci (m*Cp*3 and m*Cp*4) provided polymorphism in four wild populations. For the rest of 12 microsatellite loci, seven (m*Cp*6 to m*Cp*10, m*Cp*12, m*Cp*13) were monomorphic in all populations; two loci (m*Cp*5 and m*Cp*11) were polymorphic in only one population, three microsatellite loci (m*Cp*1, m*Cp*2, and m*Cp*14) were database searched to be linked with the Y chromosome.

3.2 Microsatellite variation in wild and cultivated papayas

Two microsatellite loci isolated from this study and five additional loci [23] were primarily used to test genetic variation (Table 2). In total, 33 and 11 alleles were detected from seven microsatellite loci in wild and cultivated papayas, respectively. The values for number of alleles and heterozygosity of wild papayas were relatively higher than cultivated papayas. N_a in wild papayas ranged from three (m*Cp*4, m*Cp*CIR7, and m*Cp*CIR9) to 10 (m*Cp*CIR2) while in cultivated papayas, it ranged from one (m*Cp*4, m*Cp*CIR3, and m*Cp*CIR9) to two (m*Cp*3, m*Cp*CIR2, m*Cp*CIR4, and m*Cp*CIR7). No specific alleles were detected in the cultivated papayas. H_E averaged over seven loci was 0.522 ± 0.077 (ranging from 0.098 to 0.702) in wild papayas, and was 0.248 ± 0.089 (ranging from 0.000 to 0.473) in cultivated papayas. F_{ST} between wild and cultivated papayas ranged from 0.025 (m*Cp*CIR7) to 0.463 (m*Cp*CIR3).

The majority (86.05%; 37 out of 43 combinations) of the wild populations conformed to Hardy-Weinberg Equilibrium (HWE) at most loci after the sequential Bonferroni correction [33]. The locus m*Cp*CIR3 was subjected to be under selection using Lositan and BayeScan. Such a locus was therefore excluded from further analyses. Another potential explanation is a null allele and/or inbreeding. No significant linkage disequilibrium was detected between genotypes at different loci.

3.3 Intrapopulation variation and breeding system

Table 3 shows the overall level of genetic variability, according to the six microsatellite loci, in the 11 analysed papaya populations. Wild populations from Chumphon Province (CP-W) provided a relatively high level of genetic variability (i.e., N_a, N_e, H_O, and H_E). Private alleles were only detected in four wild populations, varying from one (PC-W, NK-W, and CP-W2) to three (CP-W3), with frequencies ranging from 0.071 (NK-W) to 0.167 (CP-W2). Null alleles appeared to be present in all samples (r: 0.01 to 0.25), except NK-K. This indicates that null alleles could contribute to the observed heterozygote deficiency.

The mean outcrossing rate (t) was estimated from the inbreeding coefficient of all populations (Table 3). These values ranged from 15.71% (NK-W) to 47.53% (CP-W3) for the wilds. Six of eight wild populations had an average outcrossing rate estimated at more than 20%; among three cultivars, the t values were extremely varied from zero (complete inbreeding) in NK-K to 100% (complete outcrossing) in KC-H. These populations came from three regions of Thailand: North (1), Northeast (2), and South (3).

Table 3 Genetic variability over six microsatellite loci for 11 papaya populations.

Population	Sample size	N_a	N_e	P (%)	N_p	A_p	H_O	H_E	F_{IS}	t (%)	r
1 PC-W	12	2.833	2.129	83.33	1	0.083	0.088	0.443	0.657	20.68	0.246
2 NK-W	16	2.167	1.537	66.67	1	0.071	0.073	0.271	0.729	15.71	0.156
3 KK-W1	5	2.333	1.929	83.33	0	0.000	0.181	0.477	0.667	19.98	0.201
4 KK-W2	6	2.333	1.899	66.67	0	0.000	0.200	0.393	0.388	44.12	0.138
5 KK-W3	4	2.333	2.106	83.33	0	0.000	0.194	0.535	0.570	27.36	0.221
6 CP-W1	7	2.333	1.858	100.00	0	0.000	0.224	0.481	0.441	38.82	0.174
7 CP-W2	6	2.500	2.151	100.00	1	0.167	0.194	0.551	0.492	34.04	0.230
8 CP-W3	5	2.833	2.404	83.33	3	0.133	0.267	0.526	0.356	47.53	0.170
9 NK-K	18	1.167	1.167	16.67	0	0.000	0.167	0.086	-1.000	n/a	n/a
10 KC-H	16	1.333	1.314	33.33	0	0.000	0.167	0.183	0.000	100.00	0.014
11 KC-E	18	1.500	1.294	50.00	0	0.000	0.081	0.179	0.616	23.79	0.083

Note. N_a (mean number of alleles), N_e (number of effective alleles: $N_e = 1/[1\text{-}H_E]$), P (percentage of polymorphic loci), N_p (number of private alleles), A_p (mean frequency of private alleles), H_O (mean observed heterozygosity), H_E (mean expected heterozygosity), F_{IS} (mean inbreeding coefficient), t (outcrossing rate: $t = [1\text{-}F_{IS}]/[1+F_{IS}]$) [31], r (mean frequency of null alleles: $(H_E - H_O)/(H_E + 1)$) [32], n/a not available.

3.4 Population structure

The genetic differentiations among populations were measured using F_{ST} values [35] (Table 4). Significant differences in pairwise F_{ST} values ($P < 0.05$) were concentrated among pairs of NK-W populations and the other wilds, except between NK-W and KK-W1 ($F_{ST} = 0.290$) (Table 4). The genetic differentiation between PC-W and KK-W1 was also significantly different. Between the wilds and cultivars, moderate ($0.30 < F_{ST} < 0.50$) and high ($F_{ST} > 0.50$) levels of genetic differentiation were typically present, ranging from 0.425 (KK-W1 and KC-E) to 0.686 (NK-W and KC-H). Three cultivars also showed moderate to high genetic differentiation, ranging from 0.378 (KC-H and KC-E) to 0.764 (NK-K and KC-H).

The genetic structure of the eight wild populations was considered using the Bayesian cluster analysis. The ancestry of each individual was referred to one of the likely hypothetical original populations defined as a cluster (K), using STRUCTURE [37]. The method of [38] indicated five as the optimal K value ($K = 5$). At $K = 5$, all eight populations presented heterogeneous genetic structures (admixed populations) (Figure 1).

For AMOVA analyses, papaya populations were grouped into three scenarios: among wilds (four geographical areas), between wilds and cultivars, and among wilds and cultivars (five geographical areas) (Table 5). For the first scenario, the molecular variances among four different geographical areas and within populations were non-significantly different ($P = 0.129$ and $P = 0.922$, respectively). Between wilds and cultivars, the second scenario, an AMOVA indicated that variation between groups (20.12%, $P = 0.145$) was lower than that among populations within groups (42.63%, $P < 0.000$) and within populations (37.25%, $P < 0.000$). This pattern was also observed when wilds and cultivars were grouped according to the five geographical areas (the last scenario): variation between groups (0.80%, $P = 0.566$), among populations within groups (58.40%, $P < 0.000$) and within populations (40.80%, $P < 0.000$).

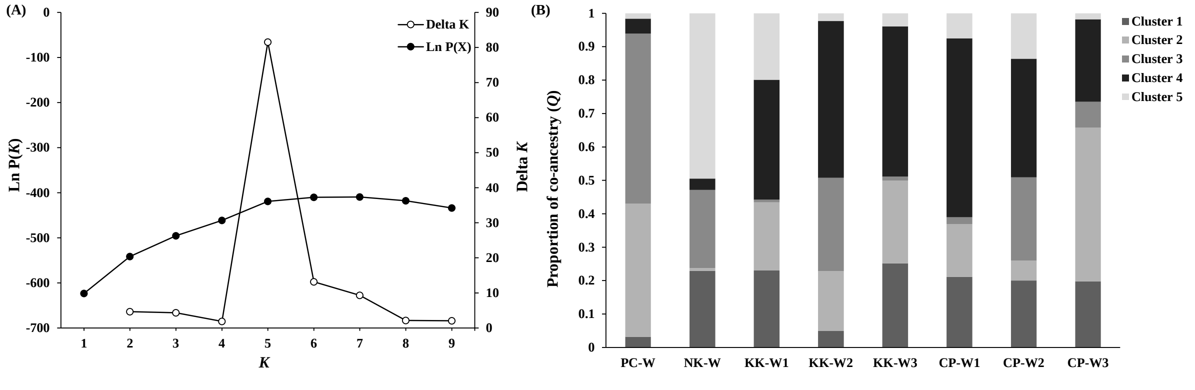

Figure 1 Inferring population structure for the eight wild papaya populations using STRUCTURE software: (A) The two curves indicate the number of inferred clusters (K) according to the log-likelihood probability [37] and the Delta K method [38]. Five is indicated to be the optimal K value ($K = 5$) using the admixture model, (B) The clustering outcomes represent the coefficients of ancestry obtained from a STRUCTURE run with $K = 5$. Each colour represents the average proportion of co-ancestry (Q) with regard to each hypothetical genetic cluster.

3.5 Demographic history

Isolation by distance (IBD) analysis, examined in 11 populations, revealed a non-significant negative relationship between genetic and geographical distances (R^2 = 0.051; P = 0.862, $F_{\mathrm{ST}}/(1-F_{\mathrm{ST}})$ = -0.0013 (geographical distance) + 1.3909). It indicated no limit to dispersal among populations.

The probability values of self-assignment ranged from 0.638 (KK-W1) to 0.921 (CP-W2) for wilds, and 0.815 (KC-E) to 1.000 (NK-K) for cultivars (Table 6). The populations from the southern (CP-W1, CP-W2, CP-W3) and northeastern parts of Thailand (KK-W2, KK-W3) contributed asymmetric migration to all wild populations (varying from m = 0.189 (from KK-W2 to PC-W) to m = 0.816 (from CP-W2 to CP-W1)) and two cultivars (varying from m = 0.110 (from KK-W2 to KC-E) to m = 0.985 (from CP-W2 to NK-K)). Within the same region of Thailand, populations significantly received genetic composition from their neighbours; Migration rates range from 0.403 to 0.657 within the northeastern region and from 0.246 to 0.816 within the southern region. Among cultivars, population NK-K significantly received genetic composition from wilds (from the North, Northeast, and South) and cultivar KC-E; however, population NK-K only shared its genetic composition to KC-E (m = 0.269). Similarly, cultivar KC-E significantly received genetic composition from all wilds (except NK-W) and two cultivars, but asymmetrically shared with only the cultivar NK-K (m = 0.538). Only cultivar KC-H did not receive and share its genetic composition with any populations.

Table 4 Pairwise-F_{ST} among 11 papaya populations.

Population	Wild								Cultivar		
	1	2	3	4	5	6	7	8	9	10	11
1 PC-W											
2 NK-W	0.291										
3 KK-W1	0.237	0.290^{ns}									
4 KK-W2	0.192^{ns}	0.401	0.047^{ns}								
5 KK-W3	0.130^{ns}	0.286	-0.199^{ns}	-0.040^{ns}							
6 CP-W1	0.135^{ns}	0.234	-0.080^{ns}	0.048^{ns}	-0.178^{ns}						
7 CP-W2	0.062^{ns}	0.169	-0.035^{ns}	0.068^{ns}	-0.138^{ns}	-0.154^{ns}					
8 CP-W3	0.027^{ns}	0.415	0.017^{ns}	0.083^{ns}	-0.160^{ns}	-0.130^{ns}	-0.129^{ns}				
9 NK-K	0.543	0.569	0.552	0.678	0.589	0.592	0.549	0.674			
10 KC-H	0.631	0.686	0.574	0.685	0.607	0.600	0.564	0.631	0.764		
11 KC-E	0.541	0.615	0.425	0.592	0.428	0.474	0.434	0.532	0.378	0.631	

ns: non-significant

Table 5 The Analysis of Molecular Variance Analysis (AMOVA).

Criteria	Number of groups	Among groups				Among populations within groups				Within populations			
		V_a	%	P	F_{CT}	V_b	%	P	F_{SC}	V_c	%	P	F_{ST}
Among wilds													
Geographical areas	four	0.044	11.33	0.129	0.113	0.011	2.86	0.922	0.032	0.335	85.82	0.033	0.142
Among wilds and cultivars													
Cultivation	two	0.120	20.12	0.145	0.201	0.255	42.63	<0.000	0.534	0.223	37.25	<0.000	0.627
Geographical areas	five	0.004	0.800	0.566	0.008	0.319	58.40	<0.000	0.589	0.223	40.80	<0.000	0.592

Note. V_a (among groups variance), V_b (among populations within groups variance), V_c (Within populations variance), % (percentage of variation), P (p-value), F_{CT} (the fixation index among groups), F_{SC} (the fixation index among populations within groups), F_{ST} (the fixation index among groups)

Table 6 Migration analysis of the 11 papaya populations.

	Reference populations										
Assignment	Wilds								Cultivars		
	1	2	3	4	5	6	7	8	9	10	11
1 PC-W	***0.675***	0.035	0.070	**0.189**	**0.227**	**0.303**	**0.530**	**0.598**	0.000	0.000	0.000
2 NK-W	**0.126**	***0.675***	**0.446**	**0.353**	**0.492**	**0.670**	**0.791**	**0.380**	0.026	0.063	0.007
3 KK-W1	0.004	**0.196**	***0.638***	**0.520**	**0.657**	**0.699**	**0.695**	**0.603**	0.000	0.000	0.011
4 KK-W2	**0.162**	0.038	**0.403**	***0.885***	**0.563**	**0.489**	**0.583**	**0.655**	0.000	0.000	0.002
5 KK-W3	0.037	0.034	**0.556**	**0.549**	***0.713***	**0.698**	**0.578**	**0.662**	0.000	0.000	0.003
6 CP-W1	0.024	0.092	**0.605**	**0.663**	**0.754**	***0.881***	**0.816**	**0.746**	0.000	0.000	0.006
7 CP-W2	0.032	0.079	**0.385**	**0.429**	**0.463**	**0.519**	***0.921***	**0.532**	0.000	0.000	0.002
8 CP-W3	0.054	0.001	0.089	**0.277**	**0.222**	**0.246**	**0.287**	***0.663***	0.000	0.000	0.027
9 NK-K	**0.179**	0.063	**0.725**	**0.456**	**0.656**	**0.877**	**0.985**	**0.609**	***1.000***	0.000	**0.538**
10 KC-H	0.004	0.000	0.000	0.000	0.010	0.006	0.039	0.023	0.000	***0.998***	0.016
11 KC-E	**0.203**	0.068	**0.306**	**0.110**	**0.443**	**0.545**	**0.778**	**0.534**	**0.269**	**0.109**	***0.815***

Significant migration rates ($m \geq 0.100$) are in bold. Assignment to original population is indicated by the italic value.

3.6 Detection of putative hybrids between wilds and cultivars

Admixture analysis was also performed using prior population information on the status of cultivation (wild and cultivated papayas) in order to infer introgression. The results revealed that 113 samples formed two hypothetical genetic clusters (Figure 2). All 52 individuals of cultivars were apparently assigned to Cluster 2 (with a major proportion of co-ancestry, $Q_{\text{Cultivars}} > 0.900$). Admixed genetic structure ($0.100 \leq Q_{\text{wilds}} \leq 0.900$) was detected in 14 of the 61 wild papayas (or 22.95%). These 14 admixed samples were from populations PC-W (5), NK-W (4), KK-W2 (2), and CP-W2 (3). In addition, the other five out of the 61 wild individuals (8.20%) were grouped into the same cultivar cluster ($Q_{\text{Cultivars}} > 0.900$). They were from NK-W, KK-W1, KK-W3, CP-W1, and CP-W3.

The PCoA analysis provided a graphical ordination of the first two axes in the plane for 113 wilds and cultivars as presented in Figure 2. The first axis, accounting for 25.10% of the total variation, mainly separated wild papayas from a group of cultivars. However, the second axis, accounting for 17.56% of the total variation, did not clearly reflect any differentiation. These 14 admixed samples and five suspected cultivated samples were scattered in the middle of the wild and cultivar groups.

Nineteen papaya samples were subsequently analysed with an individual assignment test for identification of first-generation migrants. Eleven of 19 individuals were assigned to the Chumphon populations. The rest of the samples were assigned to the Pichit (2) or Khon Kean (5) populations. One sample from NK-W was assigned to belong to both wild and cultivar. Regarding first-generation migrant detection, only 12 individuals were significantly identified from wilds (i.e., NK-W (1), PC-W (1), KK-W1 (2), KK-W3 (1), CP-W1 (1), CP-W3 (2)) and the cultivar NK-K (4)).

3.7 No evidence for escape of transgenic papaya

Based on molecular genetic tools, no expected amplified product of the CaMV 35s promoter was detected on tested papaya samples (Figure 3). Even though one of KK-W samples showed an amplified PCR product close to an expected product, its sequence did not show a significant value matching the CaMV 35s promoter sequence.

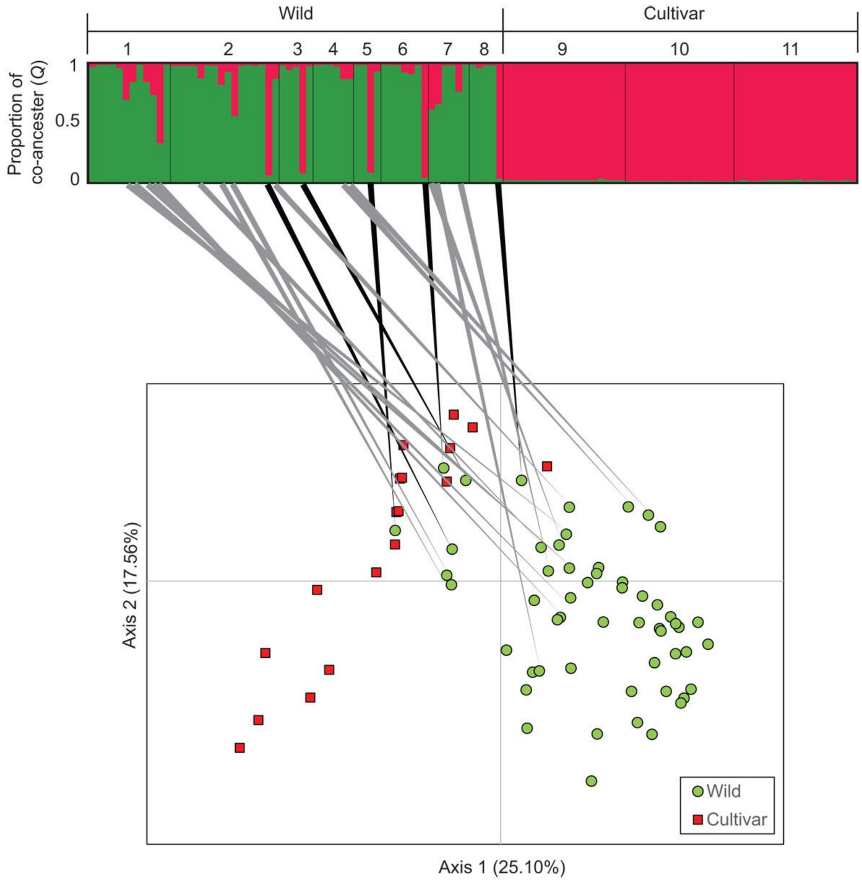

Figure 2 Detection of migrants among 113 papayas, compared between STRUCTURE [37] when $K = 2$ (Bar graph) and PCoA analyses [30] (Scatter plot). Fourteen wild papayas show an admixed structure ($0.100 < Q_{\text{Wilds}} < 0.900$; grey arrow), and the other five belong to the cultivar cluster ($Q_{\text{Cultivars}} > 0.900$; black arrow). Each arrow links the same papaya individual between two analyses. 1: PC-W; 2: NK-W; 3: KK-W1; 4: KK-W2; 5: KK-W3; 6: CP-W1; 7: CP-W2; 8: CP-W3; 9: NK-W; 10: KC-H; 11: KC-M.

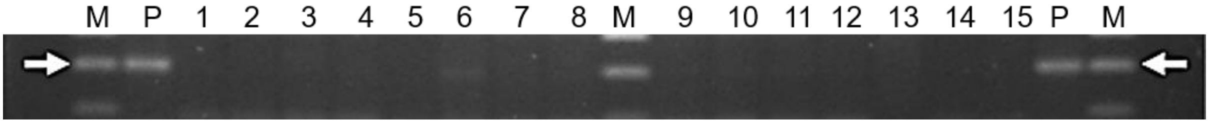

Figure 3 Evaluation of risk assessment by amplification with primers of the CaMV 35S promoter. The amplicons were analysed in 1% agarose gel stained with ethidium bromide. Arrows indicate the expected amplified product (338 bp). M: 100 bp ladder marker, P: DNA construct containing the CaMV 35S promoter (positive control), 1 to 3: PC-W, 4 to 8: KK-W, 9 to 10: NP-K, 11: KC-H, 12 to 15: KC-M.

4. Discussion

The analysis of microsatellite variation in *Carica papaya* samples collected in Thailand aimed to unravel the genetic relationship, population structure, and potential gene flow in and among wilds and cultivars. The results illustrate adequate allelic variation allowing population analyses and discussion of the following: (i) genetic diversity of wild and cultivated papayas in Thailand, (ii) the impacts of domestication, and (iii) implication of gene flow on risk assessment.

4.1 Genetic diversity of wild and cultivated papayas in Thailand

Papaya populations grown in Thailand are undoubtedly introduced, although the route of distribution is unclear. During introduction, both natural and artificial selection were forces contributing to the reduction of population size; PRSV outbreak is a strong selection factor causing significant reduction in quality and quantity of fruit production [3,4]. These forces may be a reason that the six microsatellite loci provided relatively moderate genetic polymorphism in the wilds and low polymorphism in the cultivars, even though four out of the six loci were previously proven to have relatively high polymorphism in papayas from South America and the Caribbean [23]. However, the moderate variation levels presented here sufficiently reveal private alleles with a frequency 0.1-0.2 in wild populations, present high genetic differentiation (F_{ST}: 0.459 to 0.702), and two genetic clusters between wild and cultivated papayas. This implies that during establishment, the wild populations may maintain their genetic variation under complex conditions (e.g., multiple introductions, outcrossing) as supported by gene flow analyses.

On the other hand, in cultivars, no specific microsatellite allele was observed. Several detected alleles (eight out of 11 total alleles) were present at a higher frequency than in wild papayas. Moreover, gene flow analyses indicated asymmetric migration from wilds to cultivars. These findings strongly support that cultivars used to be a part or parts of the wild ancestor gene pool. Even though the genetic makeup of cultivars significantly differed from that of wilds, cultivation practices might not be a strong enough factor to drive cultivars to adapt and establish their own adventive populations. Artificial selection, during domestication and improvement, is generally considered to be an evolutionary force. Although this process has an impact on the genetic background of the cultivars and results in significant differentiation from wilds, the domestication can be affected by several factors, for example, the strength of selection, number of human-mediated migrations, and mating system [15]. The phenomenon of genetic differentiation between wilds and cultivars is similarly observed in papayas from Costa Rica [21] and other crop species that are commonly cultivated such as soybean [18], cassava [20], and eggplant [43]. This suggests that cultivation has an impact on genetic relationship between wilds and cultivars.

Wild samples collected from the southern areas (Chumphon Province) are probably a potential source population of the local papayas in this study. This hypothesis is supported by the consistent evidence based on population genetics (this study), history, and ecology. First, samples from the South provided higher estimated polymorphism (N_a, N_e, N_p, H_O, and H_E) than from the other areas and showed asymmetric migration to the other wilds and cultivars. Second, the papaya probably came ashore in the South and was subsequently distributed nationwide. Lastly, according to ecological data, the papaya requires abundant rainfall (or irrigation) with good drainage to avoid root rot.

Papayas collected from the Northeast (Khon Kaen Province) also showed relatively high genetic variety. However, migration from these populations to the others was observed less frequently than that of papayas from southern area. The observed high level of genetic variation may be because papaya is a traditional and important staple. The papaya is therefore planted or grown in wild conditions for household use. Based on the current analyses (i.e., genetic structure, migration), these papayas were probably introduced from the South. Their genetic structures are still similar to the Chumphon populations, suggesting that continuous dispersal of seeds or plants among farmers can increase the distribution of alleles, regardless of geographical origin and growth habitat [44]. For the rest of the wild papaya populations, their genetic patterns suggest that they were introduced populations from other southern and northeastern areas and subsequently established their own populations.

4.2 The impacts of domestication

Three cultivars formed a fairly homogeneous group based on distinct genetic variability, genetic structure, and molecular variance compared with wild populations. This implies a limitation of genetic diversity among cultivars as observed in a previous study using AFLPs [2]. However, the observed level of genetic variation among three cultivars varied; the cultivar Khak Dam (NK-K) revealed the lowest level while the cultivar Eksotika (KC-E) showed the highest level of genetic variation. This seems to indicate that natural conditions and/or agricultural practices (artificial factor) may influence the differences of each cultivar.

In general, the mating system of papayas can be either dioecious or gynodioecious; however, only female and hermaphrodite papayas are desired for the market. In common cultivars (e.g., Khak Dam, Solo, Eksotika), fruits developed from hermaphrodite flowers are of better quality and have a higher yield than those from female flowers [45]. Based on a conventional breeding approach, farmers maintain their papaya orchards by selection of only hermaphrodite papayas; gynodioecious cultivated papayas generally produce two-thirds hermaphrodites and one-third female plants. Such hermaphrodite papayas predominately self-pollinate, at a rate of approximately 95% [44]. This supports low genetic variation due to inbreeding and subsequently pure lines [1].

The cultivar Khak Dam is a native variety that has been planted for over 20 years (counted from the collection date); it is a popular cultivar in Thailand. Farmers have selected only hermaphrodite plants to raise. Our genetic data revealed that five of six microsatellite loci (83.33%) were monomorphic. Only one polymorphic locus showed excess heterozygosity, resulting in the calculation of completely negative assortative mating. Similar circumstances were observed in the other gynodioecious cultivars Hawaii and Eksotika that were introduced to Thailand in approximately 2002 (less than five years before the collection date). Both cultivars had been continuously planted using introduced seeds from abroad. Genetic polymorphisms of these two cultivars are relatively more numerous than in the native cultivar Khak Dam. Four (66.57%) and three

(50.00%) of six microsatellite loci were monomorphic in the cultivars Hawaii and Eksotika, respectively. Their inbreeding coefficients were based on only two and three polymorphic microsatellite loci, respectively. The incidence of a low level of heterozygosity (approximately 0.06%) was similarly observed in the cultivar Sunup after 25 generations of inbreeding [44]. Niklas and Marler (2007) [45] examined papaya cultivation practices following commercial guidelines and fertiliser management methods in the cultivars Sunrise and Tainung. They found that the consequence of these practices was the reduction of critical functional traits involving papaya vegetation and reproduction. For that reason, agricultural practices and length of application can strongly affect not only the papaya's genetic composition, but also its phenotype.

The detection of genetic variation, even at low levels, in given cultivars may imply that some amount of outcrossing may happen in those populations. Pollen carried by the wind and/or other pollinators such as honeybees or skipper butterflies [1] can increase the rate of outcrossing, especially if domestication is not maintained in closed conditions. Moreover, environmental and climatic factors can affect the outcrossing rate. At high temperatures (> 35°C), hermaphrodites probably form functional male flowers with poorly developed and non-functional female parts. This tendency may vary within a cultivar as well as between cultivars. In addition, a sex reversal and morphological change can happen in varying degrees between male and hermaphrodite trees [1]. On the other hand, at cooler temperatures, the number of stamens in hermaphrodite flowers may be reduced. With regard to our study, we cannot observe a spatial pattern between the cultivar NK-K (a favourable commercial stain) and the nearby (within approximately 10 km) Nakhon Pathom-wild (NK-W). Similarly, the cultivar Hawaii shows lower genetic variability than Eksotika although they were planted adjacently under the same conditions. This implies that outcrossing more likely occurs between populations in very close proximity to each other. The other potential explanations include the differences of demographic history of the cultivar's background (i.e., origin, migration route, breeding lines) and/or limitations of reproduction (i.e., outbreeding depression).

Balance between the introduction of new variations (e.g., mutation, outcrossing) in cultivation and agricultural practices (e.g., artificial selection) may assist in maintaining genetic variation as well as conserving preferred phenotypes. Chan (2009) [1] showed that pollination with pollen from a few trees can slow down genetic drift and loss of varietal identity. The current research did not examine time-based populations. Study over several generations (or different periods of time) may gain knowledge on the mating system of these papaya cultivars in Thailand. Moreover, the collection of information from farmers (or breeders) is useful for risk evaluation during the domestication and improvement of cultivars.

4.3 Implication of gene flow on risk assessment

The current study aims to analyse the potential gene flow within and between eight wild and three cultivated papaya populations in Thailand. The results from several methods (i.e., STRUCTURE, PCoA, individual assignment analyses, and identification of first-generation migrants) show a congruent pattern. This implies that cultivated genotypes could survive in a natural environment as feral individuals. Additionally, the IBD analysis also infers that human-mediated dispersal can influence the movement of seeds over long distances among different geographical areas. For the mating system, the relatively high breeding coefficient observed in most of the wilds ($F_{IS} > 0.300$) supports the possibility that the breeding system of these populations is predominantly inbreeding or a mixed system. This pattern was similarly observed in papaya populations in Costa Rica [21]. The papaya can have a dioecious breeding system with pollen and eggs from different plants pollinated under either natural or artificial conditions. Accordingly, the possibility of mating between cultivars and wilds may be increased if seeds and/or pollen disperse within a sympatric habitat, flowering times are synchronous, and pollinators are similar [8,9,26]. However, if mating between wild and cultivated papayas (whether GM or non-GM) occurs in nature, the fitness and genetic transmissibility of the hybrid, F_2, and backcross progeny still need to be evaluated. It was reported that in GM squash and a wild relative, a low density of GM pollen can compete with wild pollen to produce viable hybrids carrying the transgene, although the hybrid's performance was reduced in terms of vegetation and reproduction compared to wilds. This performance can be further improved by crossing F_1 GM and wilds [13]. An occurrence of PRSV, a strong natural pressure, can also improve the survival rate of hybrids in natural conditions. However, in the meantime, it can cause wild populations to break down.

The incorporation of the mentioned concerns and knowledge from this study suggests that if cultivation of GM papaya is necessary in Thailand, the southern and northeastern parts might not be suitable places for this in terms of genetic conservation. Moreover, plantations of GM crops in greenhouses or confined areas following strict critical practices is required in order to prevent natural and artificial dispersal. A two- to three-kilometer buffer between wilds and GM plantations is recommended to preclude foreign pollen contamination [1], although as little as 400 meters was proven to be enough for prevention [6]. Also, a male-sterility system or removing the transgene from pollen is an alternative method to control or limit the risk of transgenes [9]. However, ecological interaction between papayas (whether GM or non-GM) and other organisms is complex.

More study on this aspect is needed, and implementation should only be done on a case-by-case basis for better understanding and proper risk assessment of GM papaya escape [9,46].

This research provides data based on only microsatellite DNA markers. This is one approach for evaluation and validation of hybridisation and introgression between wild and cultivated papayas. Other evidence such as morphological, biochemical, and cytological, as well as mating studies incorporating more samples in different geographical areas are still required for further assessment.

5. Conclusion

The genetic variation of wild papayas, especially from southern areas, was higher than cultivars. Cross-fertilisation was possibly a way to retain their polymorphisms. Observation of wild samples with genetic admixtures and genetic material belonging to cultivars suggests that feral papayas could survive in a natural environment. Human-mediated dispersal may influence the movement of seeds over long distances. In addition, cultivation practice and length of application may affect the genetic variation and structure of each cultivar. Therefore, information from farmers (or breeders) is useful for risk evaluation during the domestication as well as for improvement of cultivars.

6. Acknowledgements

This research project was partially supported by Faculty of Science, Mahidol University. The authors sincerely thank for the anonymous crop growers to provide their papaya samples. Also, we sincerely thank Mr. Robert Bachtell Eastland for his English editing service.

7. References

[1] Chan Y-K. Breeding papaya (*Carica papaya* L.). In: Mohan Jain S, Priyadarshan PM, editors. Breeding plantation tree crops: tropical species. New York: Springer; 2009. p. 121-159.

[2] Ratchadaporn J, Sureeporn K, Khumcha U. An analysis on DNA fingerprints of thirty papaya cultivars (*Carica papaya* L.), grown in Thailand with the use of amplified fragment length polymorphisms technique. Pak J Biol Sci. 2007;10:3072-3078.

[3] Napasintuwong O, Traxler G. Ex-ante impact assessment of GM papaya adoption in Thailand. Ag Bio Forum. 2009;12(2):209-217.

[4] Davidson SN. Power, progress and prevarication. GM Crops & Food. 2012;3(2):104-110.

[5] Gonsalves D, Gonsalves C, Ferreira S, Pitz K, Fitch M, Manshardt R, Slightom JL. Transgenic virus resistant papaya: from hope to reality for controlling papaya ringspot virus in Hawaii. APS*net* Features. 2004;Online. doi:10.1094/APSnetFeature-2004-0704

[6] Stokstad E. GM papaya takes on ringspot virus and wins. Science. 2008;320(5875):472.

[7] Pattanapomgthorn J, Sutduean J, Keohavong B. Impact of genetically modified food knowledge, environmental, and food safety concerns on purchase intention of genetically modified food in mediating role of perceived risk: an empirical study in Thailand. World Food Policy. 2020;6:23-41.

[8] Dale PJ, Clarke B, Fontes EMG. Potential for the environmental impact of transgenic crops. Nat Biotechnol. 2002;20:567-574.

[9] Stewart CN Jr, Halfhill MD, Warwick SI. Transgene introgression from genetically modified crops to their wild relatives. Nature Rev Genet. 2003;4:806-817.

[10] Poppy GM. Gene flow from GM plants-towards a more quantitative risk assessment. Trends Biotechnol. 2004;22:436-438.

[11] Warwick SI, Légère A, Simard MJ, James T. Do escaped transgenes persist in nature? The case of an herbicide resistance transgene in a weedy *Brassica rapa* population. Mol Ecol. 2008;17:1387-1395.

[12] Ellstrand NC, Meirmans P, Rong J, Bartsch D, Ghosh A, de Jong TJ, et al. Introgression of crop alleles into wild or weedy populations. Annu Rev Ecol Evol Syst. 2013;44:325-345.

[13] Cruz-Reyes R, Ávira-Sakar G, Sánchez-Montoya G, Quesada M. Experimental assessment of gene flow between transgenic squash and a wild relative in the center of origin of cucurbits. Ecospere. 2015;6(12): 1-13.

[14] Vaughan DA, Balázs E, Heslop-Harrison JS. From crop domestication to super-domestication. Ann Bot. 2007;100(5):893-901.

[15] Chen YH, Gols R, Benrey B. Crop domestication and its impact on naturally selected trophic interactions. Annu Rev Entomol. 2015;60:35-58.

[16] Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plant and its recent advance: an overview of the analytical perspective. Genet Res Int. 2015;2015:1-14.

[17] Chávez-Pesqueira M, Núñez-Farfán J. Domestication and genetics of papaya: a review. Front Ecol Evol. 2017;5:155.

[18] Kuroda Y, Kaga A, Tomooka N, Vaughan DA. Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. Mol Ecol. 2006;15:959-974.

[19] Scarcelli N, Tostain S, Vigouroux Y, Agbangla C, Daïnou O, Pham JL. Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. Mol Ecol. 2006;15: 2421-2431.

[20] Duputié A, David P, Debain C, Mckey D. Natural hybridization between a clonally propagated crop, cassava (*Manihot esculenta* Crantz) and a wild relative in French Guiana. Mol Ecol. 2007;16:3025-3038.

[21] Brown JE, Bauman JM, Lawrie JF, Rocha OJ, Moore RC. The Structure of morphological and genetic diversity in natural populations of *Carica papaya* (Caricaceae) in Costa Rica. Biotropica. 2012;44:179-188.

[22] Aizawa M, Yoshimaru H, Takahashi M, Kawahara T, Sugita H, Saito H, et al. Genetic structure of Sakhalin spruce (*Picae glehnii*) in northern Japan and adjacent regions revealed by nuclear microsatellites and mitochondrial gene sequences. J Plant Res. 2015;128:91-102.

[23] Pérez JO, Dambier D, Ollitrault P, Deeckenbrugg GP, Brottier P, Froelicher Y, et al. Microsatellite markers in *Carica papaya* L.: isolation, characterization and transferability to *Vasconcellea* species. Mol Ecol Notes. 2006;6(1):212-217.

[24] Aketarawong N, Bonizzoni M, Malacrida AR, Gasperi G, Thanaphum S. Seventeen novel microsatellite markers from an enriched library of the pest species *Bactrocera dorsalis* sensu stricto. Mol Ecol Notes. 2006;6:1138-1140.

[25] Aketarawong N, Isasawin S, Thanaphum S. Evidence of weak genetic structure and recent gene flow between *Bactrocera dorsalis* s.s. and *B. papayae*, across Southern Thailand and West Malaysia, supporting a single target pest for SIT applications. BMC Genetics. 2014;15:70.

[26] Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics. 2008;9:323.

[27] Foll M, Gaggiotti O. Agenome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180(2):977-993.

[28] Foll M. BayeScan v2. 1 user manual. Ecology. 2012;20:1450-1462.

[29] Nayfa MG, Zenger KR. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). Mar Genomics. 2016;28:99-106.

[30] Paekall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012;28(19):2537-2539.

[31] Weir BS. Genetic data analysis II: methods for discrete population genetic data. Sunderland: Sinauer Associates, Inc.; 1996.

[32] Brookfield JFY. A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol. 1996;5:453-455.

[33] Rice WR. Analysis tables of statistical tests. Evolution. 1989;43(1):223-225.

[34] Raymond M, Rousset F. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. J Hered. 1995;86:248-249.

[35] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358-1370.

[36] Dieringer D, Schlötterer C. MICROSATELLITE ANALYSER (MSA)-a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes. 2003;3:167-111.

[37] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotypic data. Genetics. 2000;155:945-959.

[38] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005;14:74-75.

[39] Excoffier L, Lischer HE. Arlequin suie ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564-567.

[40] Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. GeneClass2: A software for genetic assignment and first-generation migrant detection. J Hered. 2004;95:536-539.

[41] Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA. 1997;94:9197-9221.

[42] Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct, real time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol. 2004; 13:55-65.

[43] Mutegi E, Snow AA, Rajkumar M, Pasquet R, Ponniah H, Daunay M, et al. Genetic diversity and population structure of wild/weedy eggplant (*Solunum insanum*, Solananceae) in Southern India: implications for conservation. Am J Bot. 2015;102(1):140-148.

[44] Matos ELS, Oliveira EJ, Jesus ON, Dantas JLL. Microsatellite markers of genetic diversity and population structure of *Carica papaya*. Ann Appl Biol. 2013;163:298-310.

[45] Niklas KJ, Marler TE. *Carica papaya* (Caricaceae): a case study into the effects of domestication on plant vegetative growth and reproduction. Am J Bot. 2007;94:999-1002.

[46] Celis C, Scurrah M, Cowgill S, Chumbiauca S, Green J, Franco J. et al. Environmental biosafety and transgenic potato in a center of diversity for this crop. Nature. 2004;432:222-225.