RESEARCH ARTICLE

SABRAO Journal of Breeding and Genetics
53 (4) 620-631, 2021
http://doi.org/10.54910/sabrao2021.53.4.6
http://sabraojournal.org/
pISSN 1029-7073; eISSN 2224-8978

ANALYSIS OF THE GENETIC DIVERSITY OF *PHALAENOPSIS* ORCHIDS WITH SINGLE NUCLEOTIDE POLYMORPHISMS AND SNAP MARKERS DERIVED FROM THE *Pto* GENE

D. SUKMA[1,2*], J. ELINA[2], E. RAYNALTA[2], S.I. AISYAH[1,2], S.A. AZIZ[1], SUDARSONO[1,2], and M.T. CHAN[3]

[1]Department of Agronomy & Horticulture, IPB University, Bogor, Indonesia
[2]Plant Breeding and Biotechnology Study Program, Graduate School, IPB University, Indonesia
[3]Academia Sinica Biotechnology Center in Southern Taiwan, Tainan, Taiwan
*Corresponding author emails: dewi_sukma@apps.ipb.ac.id
Email addresses of co-authors: juanitaelina@gmail.com, erickraynalta@yahoo.com, syarifahiis@yahoo.com, sandraaziz@apps.ipb.ac.id, sudarsono_agh@apps.ipb.ac.id, mbmtchan@gate.sinica.edu.tw

SUMMARY

The *Pto* gene is a plant gene that has been reported to be involved in resistance to bacterial pathogens. A partial genomic sequence corresponding to *Pto* (~449 bp) was isolated from 16 species and four hybrids of *Phalaenopsis* during 2017 at the Department of Agronomy and Horticulture, IPB University, Bogor, Indonesia. Multiple sequence analysis was performed to find putative single nucleotide polymorphisms (SNPs) and design the corresponding single nucleotide-amplified polymorphism (SNAP) markers, which were in turn used to estimate the genetic diversity of 25 *Phalaenopsis* species. In total, 20 SNPs, of which 14 were nonsynonymous, were identified from the partial *Pto* sequences. Eighteen SNAP primers were then developed based on these 14 nonsynonymous and four synonymous SNPs. Validation results showed that 15 SNAP primers showed a polymorphism information content exceeding 0.3, suggesting the existence of more than two alleles for this locus. Upon their use, the SNAP markers described 86% of all interspecies variability. The *Pto* 52, *Pto* 349, *Pto* 229, and *Pto* 380 SNAP markers were very informative in the determination of genetic diversity. Notably, the existence of these nonsynonymous SNPs implied the possibility of functional changes within the amino acid sequence of the putative PTO protein. Thus, the resulting differences in the activity of the PTO protein may be used to breed tolerance to pathogen infection. Further work may be required to establish a functional link between tolerance to pathogens and the presence of *Pto*-SNAP markers in *Phalaenopsis* properly.

Keywords: *Phalaenopsis*, moth orchid, diversity, single nucleotide polymorphism, *Pto*, bacterial resistance

Key findings: *Phalaenopsis* orchids showed SNPs within the *Pto gene*. Population analysis suggested the existence of five haplotypes. The use of *Pto-derived* SNAP markers allowed differentiation among species, quantifying 86% of the variability and grouping the 25 investigated species into three main groups.

Manuscript received: October 10, 2021; Accepted: November 8, 2021.

Communicating Editor: Dr. Naqib Ullah Khan

INTRODUCTION

Phalaenopsis is one of the most famous orchid genera and perhaps the one with highest economic value within the Orchidaceae family. The name *Phalaenopsis* alludes to "*Phalaen*" (e.g., moth) and the moth genus "*Phalaena"* because these orchid species are believed to possess flowers whose shapes resemble flying moths (Gogol *et al*., 2012). The moth genus *Phalaena* comprises approximately 63 species, which are mostly distributed across Tropical Asia and with a small number in the temperate climates of Taiwan and China (Christenson, 2001; Tsai *et al*., 2011). The extensive breeding of *Phalaenopsis* has yielded more than 32 000 hybrids as shown by the records of the Royal Horticulture Society in the United Kingdom. These hybrids are mostly derived from 12 species: *Phalaenopsis amabilis*, *Phalaenopsis aphrodite* subsp. *formosana*, *Phalaenopsis schilleriana*, *Phalaenopsis stuartiana*, *Phalaenopsis equestris*, *Phalaenopsis sanderiana*, *Phalaenopsis lueddenmanniana*, *Phalaenopsis amboinensis*, *Phalaenopsis pulcherrima*, *Phalaenopsis fasciata*, *Phalaenopsis venosa*, and *Phalaenopsis gigantea* (Chung *et al.,* 2017). The high number of species, the ease of intercrossing, and intensive commercial breeding of this genus have allowed for great variability.

The analyses of biodiversity across *Phalaenopsis* species have mostly relied on the comparisons of floral morphology (Aziz *et al*., 2015; George *et al*., 2020) and biochemical content (Handini *et al.,* 2016). However, morphological and biochemical observations face constraints due to the limited number of species that flower at the same time and location and by environmental factors that affect sampling. Diversity at the molecular level has been reported mostly by using random amplified polymorphic DNA (Fu *et al*., 1997; Feng *et al*., 2003; Goh *et al*., 2005; Nicknejad *et al.,* 2009; Fu and Huang, 2011). Other markers that have been used are amplified fragment length polymorphisms, which have helped clarify the degree of similarity among 14 *Phalaenopsis* species (Chang *et al.,* 2009). Moreover, simple sequence repeat (SSR) markers were used to amplify and analyze DNA from 17 species of *Phalaenopsis* (Fatimah and Sukma, 2011). Genomic SSR markers were found to be sufficiently robust to differentiate between species and commercial hybrids and were also effective for simple molecular identification purposes (Chung *et al*., 2017). These markers have been mostly developed from genomic DNA that may belong to transcribed or nontranscribed regions of the genome (Varsney *et al*., 2007). Thus, the continuous development of markers is useful for the analysis of diversity and for marker-assisted breeding.

The *Pto* gene corresponds to a *Pto*-type serine/threonine kinase protein that plays a vital role in the activation of plant resistance mechanisms during the first stages of infection (Wan *et al.,* 2009). In tomato, this gene confers resistance to *Pseudomonas syringae* pv. tomato (Pilowsky and Zutra, 1982; Martin *et al.,* 1993). It has been characterized in banana (Peraza-Echeverria *et al*., 2007) and *Hevea brasieliensis* (Zhai *et al.,* 2014). The comparison of *Pto* sequences from 20 *Phalaenopsis* genotypes revealed that the highest sequence identity corresponds to *Pto* from *Musa acuminata* and that their translation may result in 149 amino acids that correspond to a proper kinase, *PTO*-type catalytic domain (Elina *et al*., 2017). The results of this study suggest that diversity in *Phalaenopsis* can be feasibly identified through the identification of SNPs within the *Pto* locus sequence and with the subsequent design of SNAP markers.

MATERIALS AND METHODS

Plant material

This study was carried out during 2017 at the Department of Agronomy and Horticulture, IPB University, Bogor, Indonesia. The *Pto* fragments in 20 genotypes of *Phalaenopsis* (obtained from 16 species and four hybrids) were isolated and characterized, and the same was also reported in past studies (Elina *et al*., 2017). The sequence of the fragments was used for SNP analysis and SNAP marker design was based on SNPs in the *Pto* sequences. The genotypes of the 25 *Phalaenopsis* species were used for genetic diversity analysis by using the *Pto*-SNAP marker (Table 1).

Sequence analysis with SNPs and evaluation of SNAP markers

On the basis of the *Pto* sequences that were retrieved from 20 genotypes of *Phalaenopsis* (16 species and four hybrids), several putative loci harboring SNPs were identified through the alignment of the *Pto* files deposited in GenBank under the accession number AAM979914.1 by using Geneious Pro-5.6.6 software (Biomatters, USA). Nonsynonymous SNPs were selected to construct haplotype variation networks by using Network ver. 4.6.1.3 software (http://www.fluxus-engineering.com/sharenet.htm). Nucleotide mutation pattern analyses were carried out with the median-joining algorithm.

Table 1. Phalaenopsis species used for the analysis of genetic diversity with SNAP markers.

Species	Code	Section
P. amabilis 'Jawa Barat'	PAJ	*Phalaenopsis*
P. amabilis 'Kalimantan'	PAB	*Phalaenopsis*
P. amabilis Papua'	PAP	*Phalaenopsis*
P. amabilis 'Pelaihari'	PPL	*Phalaenopsis*
P. amboinensis	PAM	*Amboinenses*
P. aphrodite	PRO	*Phalaenopsis*
P. bellina	PBE	*Amboinenses*
P. celebensis	PCE	*Stauroglottis*
P. corningiana	PCO	*Zebrinae*
P. cornu-cervi	PCC	*Polychilos*
P. cornu-cervi f. Sanguinea	PCCR	*Polychilos*
P. fimbriata	PFI	*Amboinenses*
P. floresensis	PFL	*Amboinenses*
P. gigantea	PG	*Amboinenses*
P. javanica	PJA	*Amboinenses*
P. lamelligera	PLA	*Polychilos*
P. lueddemanniana	PLU	*Amboinenses*
P. modesta	PMO	*Amboinenses*
P. pantherina	PPA	*Polychilos*
P. pulcherrima	PPU	*Esmeralda*
P. schilleriana	PSC	*Phalaenopsis*
P. stuartiana	PST	*Phalaenopsis*
P. tetraspis	PTE	*Zebrinae*
P. violacea Mentawai'	PVM	*Amboinenses*
P. violacea	PVI	*Amboinenses*
P. viridis	PVD	*Fuscatae*
P. zebrina	PZE	*Zebrinae*

On the basis of 18 SNPs (14 nonsynonymous and four synonymous) that were identified within all the *Pto* sequences, SNAP markers were designed by using WebSNAPER software (http://ausubellab.mgh.harvard.edu/). For the analysis of genetic diversity, total genomic DNA was isolated from the leaves of 25 *Phalaenopsis* species (*P. amabilis* includes three ecotypes, namely, *P. amabilis* 'Jawa Barat', *P. amabilis* 'Kalimantan' [PAK], and *P. amabilis* 'Papua') as shown in Table 1 by using the standard CTAB method with a few minor modifications developed for orchid species and perennial crops (Doyle and Doyle, 1987; Sutanto *et al*., 2014; Elina *et al*., 2017; Pesik *et al*., 2017; Sukma *et al*., 2017; Sudarsono *et al*., 2018; Raynalta *et al*., 2018; Sukma *et al*., 2020). The SNAP primers were used for PCR with genomic DNA. Each PCR reaction (total of 12.5 µl) consisted of total genomic DNA (4 µl), a set of three primers (0.25 µl each), 10× PCR Ready Mix (25 µl, KAPA Biosystem), and ddH_2O (6.5 µl). The reaction cycles involved predenaturation (95 °C for 3 min.), followed by 35 amplification cycles that consisted of denaturation (95 °C for 15 s), primer annealing (47.9 °C–59.8 °C for 15 s in accordance with the appropriate primer *Ta*), primer extension (72 °C for 1 s), and a final extension cycle (72 °C for 10 min) as recommended by the KAPA Biosystem PCR kit.

The resulting PCR amplicons were fractionated in 1× sodium borate agarose (2%) through gel electrophoresis at 50 V for 35 min. The gel was stained by using GelRedTM (Biotium Inc.), visualized with a UV transilluminator, and photographed with a digital camera. The data obtained from PCR product visualization by electrophoresis were converted into a binary dataset based on the band's appearance at each locus. Each allele was scored manually by using the electropherogram, and the observed alleles were recorded as the genotype for each species or hybrid. The binary data were then converted into allelic data for further analysis. Data were analyzed by utilizing CERVUS 2.0 software (Kalinowski *et al*., 2007) to determine the values for polymorphic information content (PIC), and GenAlEx 6.502 software (Peakall and Smouse, 2006) was used to determine heterozygosity values. PIC was calculated to identify the informative markers. An unrooted weighted neighbor-joining phylogenetic tree was constructed based on genetic dissimilarity by using Darwin 6.0.14 software (Perrier and Jacquemoud-Collet, 2014).

RESULTS

Determination of the location of SNPs and the haplotype network

On the basis of *Pto* sequence analysis, 18 putative SNP sites were identified (14 nonsynonymous and four synonymous). The predicted corresponding amino acids for the nonsynonymous SNPs are shown on Table 2. On average, a total of 18 SNPs per 449 nucleotides suggested the presence of roughly one SNP per 24 nucleotides, and the ratio of nonsynonymous to synonymous SNPs was 0.28. The presence of 14 nonsynonymous SNPs implied that amino acid variation was the result of selection pressure on the activity of the PTO kinase. Differences in protein activity could be a meaningful source of resistance to pathogen infection and may perhaps be useful in plant breeding.

The mutation patterns for nucleotides among haplotypes and the corresponding network analysis were developed by using the median-joining algorithm (Figure 1). Twenty *Pto* sequences were used in the experiment. As a result, five putative haplotypes were identified. A haplotype refers to the inheritance of a cluster of SNPs in which variation at a single position in the DNA is retained by all individuals in a population (Andersen and Lubberstedt, 2003). Most genotypes (16 species) harbored haplotype one, which is represented by PAK (Figure 1), whereas haplotypes 2, 3, 4, and 5 consisted of one species each, namely, *Phalaenopsis lamelligera*,

Table 2. Position of 14 nonsynonymous substitution mutations in the predicted amino acid sequence of the *Pto* fragment as obtained from 20 *Phalaenopsis* genotypes (16 species and four hybrids).

Species	SNP *non-synonymous position*													
	8	27	29	39	41	49	53	57	64	67	69	71	91	112
PAJ	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PAK	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PAP	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PCC	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PFI	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PGG	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PHA	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PHK	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PHM	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PHV	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PLU	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PMO	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PPU	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PSC	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PST	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PTE	K	R	R	D	R	E	K	K	D	N	N	E	S	M
PLA	K	H	R	N	G	D	N	N	N	N	N	E	S	M
PVI	K	R	H	D	R	E	K	K	D	N	N	E	S	M
PAM	K	R	R	D	R	E	K	K	N	N	N	E	S	M
PBE	K	R	R	D	R	E	K	N	D	N	N	E	S	M
M. acuminata	T	R	R	D	K	E	K	K	N	P	S	K	Q	L

PAJ = *P.amabilis* 'Jawa Barat', PAK = *P. amabilis* 'Kalimantan', PAP = *P. amabilis* 'Papua', PCC = *P. cornu-cervi*, PFI = *P. fimbriata*, PGG = *P. gigantea*, PHA = *P.* 'AMP17', PHK = *P.* 'KHM0421', PHM = *P.* hibrida 'MKW002', PHV = P. 'V3', PLU = *P. lueddemanniana*, PMO = *P. modesta*, PPU = *P. pulcherrima*, PSC=*P. schilleriana*, PST = *P. stuartiana*, PTE = *P. tetraspis*, PLA=*P. lamelligera,* PVI = *P. violacae*, PAM = *P. amboinensis*, PBE = *P. bellina*, and *M. acuminata* as an outgroup.

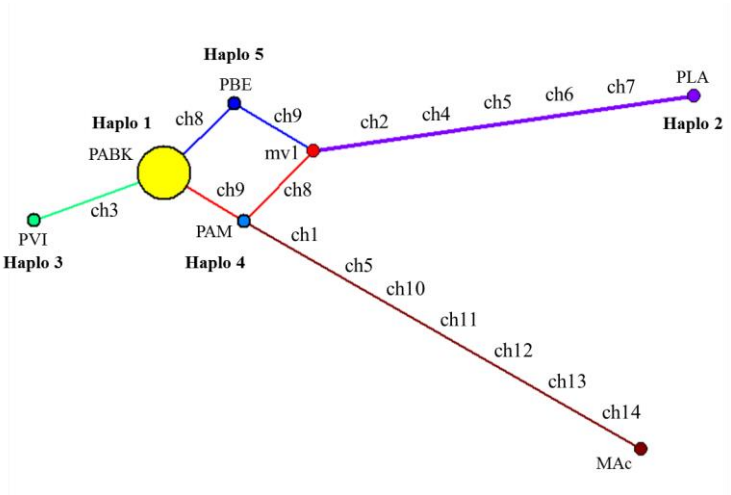

Figure 1. Haplotype network based on the median-joining algorithm. This network shows the mutation/substitution on the *Pto* sequences in *Phalaenopsis* and *Musa acuminata (Mac)* as an outgroup. Each circle stands for a unique haplotype, and the circle size shows the number of individuals harboring the haplotype. Crosshatches show the number of nucleotide differences between haplotypes.

Phalaenopsis violaceae, *Phalaenopsis amboinensis*, and *Phalaenopsis bellina*, respectively.

For each haplotype, the key mutation corresponded to a nonsynonymous SNP that was linked to an amino acid substitution within the PTO catalytic domain (Figure 1, Table 2). One SNP differentiated haplotype 1 from haplotypes 3, 4, and 5, whereas seven SNPs differentiated haplotype 1 from 2. *P. amabilis*, *P. violaceae*, *P. amboinensis*, and *P. bellina* were all members of the Amboinenses section (haplotype 1), while *P. lamelligera* (haplotype 2) was a member of the Polychilos section (Wiersema 2019). Therefore, our results suggested that species in the Amboinenses section may show a different evolutionary selection pattern for *Pto*. This pattern may be related to resistance to disease, including soft-rot disease. In fact, Sukma *et al*. (2017) reported that *P. amboinensis* is indeed resistant to bacterial disease, whereas *P. amabilis* and *P. bellina* are susceptible to infection by the pathogen *Dickeya dadantii* (also known as *Erwinia chrysanthemi*). However, detailed reports regarding the resistance of *P. violaceae* and *P. lammeligera* to *D. dadantii* do not exist. The development of genotyping tools based on *Pto*-derived SNAP markers and the phenotyping of all species for their response to *D. dadantii* constitute an opportunity for the identification of a useful selection trait in *Phalaenopsis*.

SNAP marker development, genetic diversity, and phylogenetic analyses

The identification of 18 SNP sites resulted in the design of primer sets for SNAP markers (reference and alternate). The corresponding PCR products ranged in length from 205 base pairs to 297 base pairs (Table 3). The genetic diversity of the 27 genotypes is shown in Table 4. In general, the corresponding PIC values for the *Pto*-SNAP markers were in the range of 0.200 to 0.469. The PIC value indicates the expected fraction of informative offspring from a given type of pedigree (Hildebrand *et al*., 1992). In this case, the values suggested that the polymorphic markers *Pto* 52, *Pto* 349, *Pto* 229, and *Pto* 380 were informative. Factorial analysis showed that the variability that can be evaluated with these 18 SNAP markers was 85.57% and can be fully assessed by using only 15.

The phylogenetic analysis of 25 *Phalaenopsis* species genotypes based on 18 *Pto*-SNAP markers is shown in Figure 2. *Phalaenopsis* species were divided into three main groups. Groups I and II were split into two subgroups (a, b). Group 1 had 15 species, group II included 10 species, and group III had two species. Most of the *Phalaenopsis* species were in group II, and species from section Amboinensis were placed in groups I, II, and III. Species in the Polychilos section were found in groups I and II, while all of the species of the Zebrina section were recognized in group I. These results suggested that *Pto*-SNAP markers could differ among *Phalaenopsis* sections.

In the present research, based on SNP occurrence in the *Phalaenopsis Pto*-sequence, the 18 SNAP markers were developed from 14 nonsynonymous SNP and four synonymous SNP. The markers were used to evaluate and estimate genetic variability in 25 *Phalaenopsis* genotypes. Four markers (*Pto* 52, *Pto* 349, *Pto* 229, and *Pto* 380) with high PIC values were identified as informative markers (Table 4). *Pto 52* revealed the highest value for the expected heterozygosity (He), inertia, and accessible variability. The present results showed that the 18 markers were sufficient for describing genetic variability (85.57%).

Table 3. SNAP markers developed from 18 SNPs located within *Pto* sequences from *Phalaenopsis* and may be used for genetic diversity analysis.

Locus No.	Primer Code	Primer sequence[a]	TM	L	PCR Product size (bp)
52	CN*Pto*#52_L_REF_2	CAAGGTCTCAATGAATTGCAC	54.46	21	267
	CN*Pto*#52_L_ALT_1	CAGCAAGGTCTCAATGAATTTAAT	55.58	24	270
	CN*Pto*#52_L_ALT_1_REV	ATCGCGATGAATGATGGC	56.5	18	
72	CN*Pto*#72_L_REF_4	CGGAAATTGAATTGCGTTC	55.39	19	297
	CN*Pto*#72_L_REF_4_REV	CAGCAACCTTTGCCATGA	55.18	18	
	CN*Pto*#72_L_ALT_4	CGGAAATTGAATTGCGTTT	54.99	19	250
	CN*Pto*#72_L_ALT_4_REV	TTGACATCGCGATGAATGAT	56.12	20	
79	CN*Pto*#79_L_REF_6	AAATTGAATTGCTTTCGAGACTT	55.32	23	250
	CN*Pto*#79_L_REF_6_REV	GACTTGACATCGCGATGAAT	54.78	20	
	CN*Pto*#79_L_ALT_1	TGAATTGCTTTCGAGGGTC	54.91	19	253
	CN*Pto*#79_L_ALT_1_REV	ATTTGCAGACTTGACATCGC	55.04	20	
94	CN*Pto*#94_L_REF_5	GCTTCGTCATCGTCAGCTC	55.61	19	250
	CN*Pto*#94_L_REF_5_REV	ATCCAGAAGGATATTTGCAGACT	54.98	23	
	CN*Pto*#94_L_ALT_7	GGCTTCGTCATCGTCATCTT	56.41	20	290
	CN*Pto*#94_L_ALT_7_REV	CGAAAGACCAAAATCAGC	50.26	18	
181	CN*Pto*#181_L_REF_5	GGGAACTCTGAAGAGTCAGCTC	55.99	22	251
	CN*Pto*#181_L_ALT_5	GGGAACTCTGAAGAGTCAGCTT	55.59	22	251
	CN*Pto*#181_L_ALT_5_REV	TAACTGCGGTGCTCACGT	54.84	18	
220	CN*Pto*#220_L_REF_3	CCTCAACTGGGAGCAGAGG	56.97	19	218
	CN*Pto*#220_L_ALT_5	CTCAACTGGGAGCAGCGA	57.12	18	217
	CN*Pto*#220_L_ALT_5_REV	AGCTCCCTTTAACTGCGG	54.28	18	
223	CN*Pto*#223_L_REF_1	AACTGGGAGCAGCGGTTC	57.41	18	205
	CN*Pto*#223_L_ALT_1	AACTGGGAGCAGCGGATG	58.35	18	205
	CN*Pto*#223_L_ALT_1_REV	TAACTGCGGTGCTCACGT	54.84	18	
229	CN*Pto*#229_L_REF_2	GAGCAGCGGCTCGAGATC	58.33	18	208
	CN*Pto*#229_L_ALT_4	GAGCAGCGGCTCGAGATT	57.82	18	208
	CN*Pto*#229_L_ALT_4_REV	AGCTCCCTTTAACTGCGGT	55.32	19	
241	CN*Pto*#241_R_REF_1	TAGTGAAGCCCTCGTCCG	55.93	18	257
	CN*Pto*#241_R_ALT_4	GATAGTGAAGCCCTCGTTCC	55.06	20	259
	CN*Pto*#241_R_ALT_4_REV	GCTGTCAAGCGTGGCAAT	56.93	18	
292	CN*Pto*#292_R_REF_6	GCAGACTTGACATCCCGA	53.89	18	260
	CN*Pto*#292_R_ALT_7	TGCAGACTTGACATCTCGG	54.22	19	261
	CN*Pto*#292_R_ALT_7_REV	CACACGGAAATTGAATTGCT	54.9	20	
340	CN*Pto*#340_R_REF_1	AGACCAAAATCAGCAACGTTT	55.24	21	260
	CN*Pto*#340_R_ALT_4	GACCAAAATCAGCAACGTTC	54.43	20	259
	CN*Pto*#340_R_ALT_4_REV	CTTATCGGATACTGCGACGA	55.19	20	
349	CN*Pto*#349_R_REF_6	TGTCTTCGAAAGACCAAAATCA	56.17	22	262
	CN*Pto*#349_R_ALT_7	GTCTTCGAAAGACCAAAGTCG	55.36	21	261
	CN*Pto*#349_R_ALT_7_REV	ATACTGCGACGAGCGAAA	54.13	18	
380	CN*Pto*#380_R_REF_2	TGCTCACGTGAGTCTGATACAA	55.66	22	293
	CN*Pto*#380_R_ALT_2	GTGCTCACGTGAGTCTGATCTAG	55.49	23	294
	CN*Pto*#380_R_ALT_2_REV	ATACTGCGACGAGCGAAA	54.13	18	
424	CN*Pto*#424_R_REF_2	CGGAAGTATTCAGGATCAAGCTAT	56.87	24	291
	CN*Pto*#424_R_ALT_3	CGGAAGTATTCAGGATCAAGGTAC	57.05	24	291
	CN*Pto*#424_R_ALT_3_REV	GAAGGGAACTCTGAAGAGTCATC	54.96	23	
37	CN*Pto*#37_L_REF_2	CCGAAATCCCAGCAAGAC	55.19	18	265
	CN*Pto*#37_L_ALT_2	CCGAAATCCCAGCAGGGT	59.46	18	265
	CN*Pto*#37_L_ALT_2_REV	TTGCCTTTGCAGAACCAG	54.43	18	
64	CN*Pto*#64_L_REF_3	GAATTTCACACGGAAATCGAG	55.96	21	261
	CN*Pto*#64_L_ALT_4	AATTTCACACGGAAATCGAA	54.4	20	260
	CN*Pto*#64_L_ALT_4_REV	CTTGACGTCGCGGTGAAT	56.76	18	
127	CN*Pto*#127_L_REF_1	TACTGCGACGAGCGCAAT	57.72	18	258
	CN*Pto*#127_L_ALT_7	GATACTGCGACGAGCGTAAC	54.83	20	260
	CN*Pto*#127_L_ALT_7_REV	CTTCGAAAGACCGAAATCG	54.69	19	
355	CN*Pto*#355_R_REF_6	CCCGTCTTCGAAAGTCCG	57.83	18	
	CN*Pto*#355_R_REF_6_REV	CGTCATCGTCACCTCGTG	55.4	18	293
	CN*Pto*#355_R_ALT_7	CCCGTCTTCGAAAGACCA	55.72	18	
	CN*Pto*#355_R_ALT_7_REV	TTCGTCATCGTCACCTCG	54.98	18	295

Table 4. Profiles of 18 *Pto*-SNAP markers used for the diversity analysis of 27 *Phalaenopsis* genotypes.

No.	Marker	Na[a]	PIC	Ne	Ho	He	Factorial analysis for the level of accessible variabilities		
							Number of markers	Inertia (%)	Variabilities accessible (%)
1	*Pto*-52	3	0.469	2.30	0.59	0.57	1	19.05	19.05
2	*Pto*-72	2	0.372	1.98	0.89	0.49	2	14.52	33.57
3	*Pto*-79	2	0.366	1.93	0.74	0.48	3	12.29	45.86
4	*Pto*-94	2	0.352	1.84	0.70	0.46	4	8.16	54.02
5	*Pto*-181	2	0.366	1.93	0.30	0.48	5	6.57	60.59
6	*Pto*-220	2	0.256	1.43	0.15	0.30	6	5.95	66.54
7	*Pto*-223	2	0.200	1.29	0.19	0.23	7	4.98	71.52
8	*Pto*-229	2	0.375	2.00	1.00	0.50	8	3.58	75.1
9	*Pto*-241	2	0.372	1.98	0.81	0.49	9	3.43	78.53
10	*Pto*-292	3	0.357	1.72	0.33	0.42	10	2.57	81.1
11	*Pto*-340	2	0.310	1.62	0.37	0.38	11	2.13	83.23
12	*Pto*-349	3	0.468	2.29	0.33	0.56	12	1.37	84.6
13	*Pto*-380	2	0.375	2.00	0.81	0.50	13	0.64	85.24
14	*Pto*-424	2	0.366	1.93	0.74	0.48	14	0.25	85.49
15	*Pto*-37	3	0.294	1.47	0.22	0.32	15	0.08	85.57
16	*Pto*-64	2	0.372	1.98	0.67	0.49	16	0.00	85.57
17	*Pto*-127	2	0.330	1.72	0.59	0.42	17	0.00	85.57
18	*Pto*-355	2	0.352	1.84	0.63	0.46	18	0.00	85.57

N = number of plant species, Na = number of alleles, PIC = polymorphism information content, Ne = number of effective alleles, Ho = observed heterozygosity, He = expected heterozygosity.

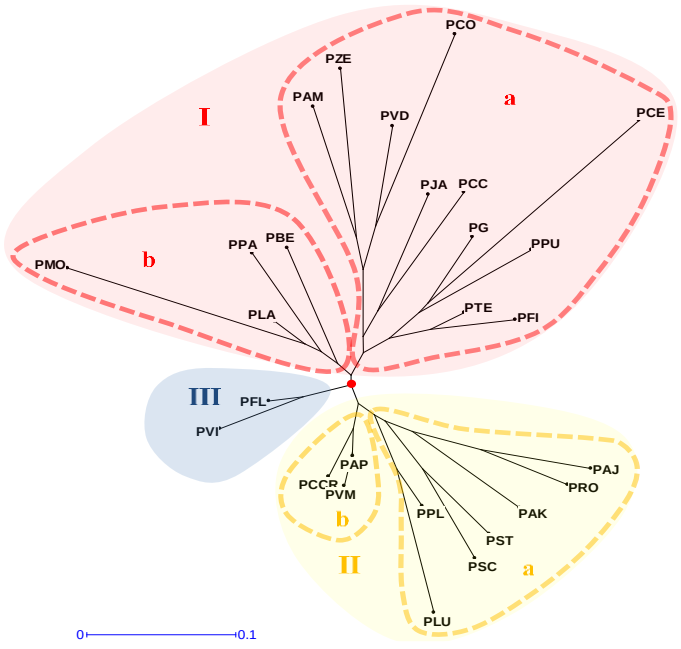

Figure 2. Phylogenetic network of 27 *Phalaenopsis* species based on 18 Pto-SNAP markers using the weighted neighbor-joining method: *P. amabilis* 'Jawa Barat' (PAJ), *P. amabilis* 'Kalimantan' (PAK), *P. amabilis* 'Papua' (PAP), *P. amabilis* 'Pelaihari' (PPL), *P. amboinensis* (PAM), *P. aphrodite* (PRO), *P. bellina* (PBE), *P. celebensis* (PCE), P. *corningiana* (PCO), *P. cornu-cervi* (PCC), *P. cornu-cervi* f. Sanguinea (PCCR), *P. fimbriata* (PFI), *P. floresensis* (PFL), *P. gigantea* (PG), *P. javanica* (PJA), *P. lamelligera* (PLA), *P. lueddemanniana* (PLU), *P. modesta* (PMO), *P. pantherina* (PPA), *P. pulcherrima* (PPU), *P. schilleriana* (PSC), *P. stuartiana* (PST), and *P. zebrina* (PZE).

DISCUSSION

The isolation and characterization of *Pto* loci with degenerate primers allowed the retrieval of 40 *Pto* sequences from 20 different *Phalaenopsis* species harboring the catalytic domain (Elina *et al*., 2017). The catalytic *Pto*-type serine/threonine kinase domain is conserved in most plant species (Oh and Martin, 2011). The *Pto* kinase has been documented to play a vital role in the activation of plant defense responses to bacterial pathogens (Lehti-Shiu and Shiu, 2012).

The number of SNPs nested within *Pto s*equences in *Phalaenopsis* was higher (1 SNP per 24 nucleotides) than the previously reported number of 1 SNP per 100–300 bp (Xu, 2010) or 1 SNP per 31–124 bp for the *WRKY* gene (Eulgem *et al.,* 2000; Zhang and Wang, 2005). In addition, the ratio of nonsynonymous SNPs to synonymous SNPs was 0.28 (<1), suggesting high nucleotide variability and hinting at the presence of selective pressure (Xie *et al.,* 2019), and purifying selection (Xiao *et al*., 2017).

The PIC shows the informativeness of a marker in terms of genetic diversity. Heterozygosity is a parameter providing information about the level of genetic variability. Therefore, low heterozygosity shows limited genetic variability. Usually, markers with high PIC values, especially PIC values greater than 0.5 (Botstein *et al*., 1980), are better at distinguishing species or cultivars (Feng *et al*., 2016) than those with low PIC values. Most of the loci in our report showed a PIC value of approximately 0.3 with the highest at *Pto* 52 (PIC = 0.468) and *Pto* 349 (PIC 0.469), suggesting that both were informative markers, whereas *Pto 52* showed the highest heterozygosity in agreement with a prior report that suggested its use in differentiating between clonal plantlets of *P. amabilis* (Raynalta *et al*., 2018).

The level of heterozygosity was estimated under the assumption that alleles were under Hardy–Weinberg equilibrium. Most of the markers showed high values of observed heterozygosity (Ho) than He except for *Pto* 181, *Pto* 220, *Pto* 223, *Pto* 292, *Pto* 349, and *Pto* 37. Low Ho values may suggest inbreeding, and high He values may suggest speciation (Yun *et al.,* 2020). In our case, He was usually lower than Ho, suggesting excess heterozygosity (Yun *et al.,* 2020).

Pto genes have been extensively reported for their crucial role in activating plant resistance mechanisms (Pilowsky and Zutra, 1982; Martin *et al.,* 1993; Thilmony *et al.,* 1995; Riely and Martin, 2001; Wan *et al.,* 2009). The *Phalaenopsis Pto* gene may also be considered to play a role. For example, a survey in Taiwan (1987–1988) found that the most devastating diseases of *Phalaenopsis* are soft-rot (*D. dadantii*), brown spot (*Pseudomonas cattleyae*), black rot (*Phytophthora palmivora*, Pnicotianae var. *parasitica*), petal blight (*Botrytis cinerea*), and cymbidium mosaic potexvirus. Pathogens, *D. dadantii, P. cattleyae*, *P. palmivora*, and *P. nicotianae* var. *parasitica* can cause the death of orchid plants (Wey *et al*., 1988).

Niu *et al.* (2016) have also reported the presence of disease resistance (R) genes and NBS-encoding gene families in the assembled transcriptomes of *P. equestris.* However, information about the actual role of R genes and resistance to pathogens in *Phalaenopsis* is limited. By contrast, the current study built upon early evidence regarding the response of *Phalaenopsis* species to infection by *D. dadantii* (Sukma *et al.,* 2017; Sanjaya *et al.,* 2020). Nonetheless, further research is required to dissect further the functional relationship between *Pto* and the response to *D. dadantii* in live plants. Our results revealed that SNAP markers, such as *Pto* 52, may be used for such a project.

CONCLUSIONS

Partial *Pto* sequences (449 bp) isolated from *Phalaenopsis* revealed the existence of 20 SNP loci, 14 of which were nonsynonymous. These loci were then used to design their corresponding *Pto*-

SNAP markers. The evaluation of these markers allowed the identification of 15 *Pto*-SNAP loci showing a PIC value of at least 0.3 and describing 86% of the existing variability among the 25 *Phalaenopsis* genotypes under study. The results indicated that *Pto*-SNAP52*, Pto*-SNAP349*, Pto*-SNAP229, and *Pto*-SNAP380 markers are informative and could be very valuable *Phalaenopsis* genetic analysis in the future.

ACKNOWLEDGEMENTS

Part of this research was supported by a Competitive Grant (*Hibah Kompetensi*) entitle "Variety, In Vitro Clonal Propagation Technology, and Molecular Marker Development for Phalaenopsis Orchids (*Pengembangan Varietas, Teknologi Propagasi Klonal* In Vitro *dan Marka Molekuler pada Anggrek* Phalaenopsis)" obtained from the Ministry of Research and Technology and Higher Education (RISTEK-DIKTI), Republic of Indonesia, Project No. 079/SP2H/LT/DRPM/II/2016, under the coordination of Dewi Sukma. This publication was also supported by the Overseas International Seminar Support Program (*Program Bantuan Seminar Luar Negeri*) from RISTEK-DIKTI, in 2019.

REFERENCES

Andersen J, Lubberstedt T (2003). Functional markers in plants. *Trends Plant Sci*. 8: 554–560.

Aziz SA, Sukma D, Romeida A (2015). Morphological characterization of *Phalaenopsis* spp. and hybrids from Indonesia. *Acta Hort*. 1078: 139-148.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

Chang YK, Veilleux RE, Iqbal MJ (2009). Analysis of genetic variability among *Phalaenopsis* species and hybrids using amplified fragment length polymorphism. *J. Am. Soc. Hort. Sci.* 134: 58-66.

Christenson EA (2001). *Phalaenopsis.* Portland (US): Timber Press. pp. 330.

Chung YL, Kuo YT, Wu WL (2017). Development of SSR Markers in *Phalaenopsis* orchids, their characterization, cross-transferability, and application for identification, In: Chen WH, Chen HH (eds). Orchid Biotechnology II. Singapore (SG): World Sci.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. 19: 11-15.

Elina J, Sukma D, Giyanto, Sudarsono (2017). Isolasi dan karakterisasi gen *Pto* asal 20 aksesi anggrek *Phalaenopsis* (Isolation and characterization of *Pto* gene from 20 *Phalaenopsis* orchid genotypes). *J. Agron. Indonesia* 45: 204-211.

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5: 199-206.

Fatimah, Sukma D (2011). Development of sequence-based microsatellite markers for *Phalaenopsis* orchids. *HAYATI J. Biosci.* 18: 71-76.

Feng M, Yuguang D, Yuxia L, Jingwen W, Bin L, Mingming Y, Daling S (2003). RAPD analysis on genetic diversities of phalaenopsis varieties with different flower colors. *Acta Agric. Shanghai* 19: 44-47.

Feng S, He R, Lu J, Jiang M, Shen X, Jiang Y, Wang Z, and Wang H (2016). Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. *Front. Genet*. 7: 113. doi: 10.3389/fgene.2016.00113

Fu S, Huang HJ (2011). Molecular characterization of the early response of orchid *Phalaenopsis amabilis* to *Erwinia chrysantemi* infection. In: Chen WH, Chen HH, editor. Orchid Biotechnology II. Singapura (SG): *World Sci.* pp. 283-308.

Fu YM, Chen WH, Tsai WT, Lin YS, Chyou MS, Chen YH (1997). Phylogenetic studies of taxonomy and evolution among wild species of Phalaenopsis by random amplified DNA markers. *Report of the Taiwan Sugar Research Institute* 157: 27-42.

George, RM, Thomas B, M Unni A (2020). Genetic variability studies in *Phalaenopsis* orchids. *J. Pharmacog. Phytochem.* 9: 377-379.

Gogol K, Raju D, Yonzone R (2012). Diversity and distribution of the genus *Phalaenopsis* Blume (*Orchidaceae*) in Assam, India. *Sci. Res. Report* 2: 260-264.

Goh MWK, Kumar PP, Lim SH, Tan HTW (2005). Random amplified polymorphic DNA analysis of the moth orchids, *Phalaenopsis* (Epidendroideae: Orchidaceae). *Euphytica* 141: 11-22.

Handini AS, Sukma D**,** Sudarsono (2016). Analisis keragaman morfologi dan biokimia pada Anggrek Phalaenopsis (Orchidaceae). *J. Agron. Indonesia* 44: 62-67.

Hildebrand CE, Torney DC, Wagner RP (1992). Informativeness of polymorphic DNA markers. *Los Alamos Sci.* 20(20): 100-102.

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099-1106. http://dx.doi.org/10.1111/j.1365-294x.2007.03089.x.

Lehti-Shiu MD, Shiu SH (2012). Diversity, classification, and function of the plant protein kinase superfamily. *Phil. Trans. R. Soc*. B 3672619-2639.

Martin GB, Devicente MC, Tanksley SD (1993). High-resolution linkage analysis and physical characterization of the *Pto* bacterial resistance locus in tomato. *Mol. Plant Pathol.* 6: 26-34.

Meera TM, Louis V, Beena S (2016). Diseases of *Phalaenopsis*: symptoms, etiology, and management. *Int. J. Agric. Innov. Res.* 5: 2319-1473.

Niknejad A, Kadir MA, Kadzimin SB, Abdullah NAP, Sorkheh K (2009). Molecular characterization and phylogenetic relationships among and within species of Phalaenopsis (Epidendroideae: Orchidaceae) based on RAPD analysis. *Afr. J. Biotechnol*. 8: 5225-5240.

Niu SC, Xu Q, Zhang GQ, Zhang YQ, Tsai WC, Hsu JL, Liang CK, Luo YB, Liu ZJ (2016). De novo transcriptome assembly databases for the butterfly orchid *Phalaenopsis equestris*. *Sci. Data* 3: 160083. DOI: 10.1038/ sdata.2016.83.

Oh C, Martin GB (2011). Effector-triggered immunity mediated by the *Pto* kinase. *Trends Plant Sci*. 16: 1360-1385.

Peakall R, Smouse PE (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol*. 6: 288-295.

Peraza-Echeverria S, James-Kay A, Canto-Canche B, Castillo-Castro E. (2007). Structural and phylogenetic analysis of *Pto*-type disease resistance gene candidates in bananas. *Mol. Genet. Genomics* 278: 443-453.

Perrier X, Jacquemoud-Collet JP (2014). DARwin software. Available online at http://darwin.cirad.fr/darwin.

Pesik A, Efendi D, Novarianto H, Dinarti D, Sudarsono S (2017). Development of SNAP markers based on nucleotide variability of *WRKY* genes in coconut and their validation using multiplex PCR. *Biodiversitas* 18: 465-475.

Pilowsky M, Zutra D (1982). Screening wild tomatoes for resistance to bacterial speck pathogen (*Pseudomonas tomato*). *Plant Dis.* 66: 46-47.

Raynalta E, Elina J, Sudarsono S, Sukma D (2018). Clonal fidelity of micro-propagated *Phalaenopsis* plantlets based on assessment using eighteen Ph-*Pto* SNAP Marker Loci. *Agrivita J.* 40: 390-402.

Riely BK, Martin GB (2001). The ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. *Proceed. Nat. Acad. Sci. USA* 98: 2059-2064.

Sanjaya IPW, Sukma, D, Sudarsono, Chan MT (2020). Effect of genotype, concentration, and timing of salicylic acid application to *Phalaenopsis* against *Dickeya dadantii* infection. *Biodiversitas* 21: 4317-4324.

Sudarsono S, Elina J, Giyanto, Sukma D (2018) Pathogen causing *Phalaenopsis* soft rot disease – 16S rDNA and virulence characterization. *Plant Prot. Sci.* 54: 1-8.

Sukma D, Elina J, Giyanto, Sudarsono (2017). Disease resistance breeding of *Phalaenopsis* spp. for the tropical environment and molecular marker development for plant selection. *Acta Hort*. 1167: 237-243.

Sukma D, Handini AS, Sudarsono (2020). Isolation and characterization of chalchone synthase (CHS) gene from *Phalaenopsis* and *Doritaenopsis* orchids. *Biodiversitas* 21: 5054-5064.

Sutanto A, Sukma D, Hermanto C, Sudarsono (2014). Isolation and characterization of resistance gene analogue (RGA) from *Fusarium* resistant banana

cultivars. *Emirates J. Food Agric.* 26: 508-518.

Thilmony RL, Chen Z, Bressan RA, Martin GB (1995). Expression of the tomato *Pto* Gene in tobacco enhances resistance to *Pseudomonas syringae* pv *tabaci* expressing avr*Pto*. *The Plant Cell* 7: 1529-1536.

Tsai CC (2011). Molecular phylogeny and biogeography of *Phalaenopsis* species. In: Chen WH, Chen HH (eds). *Orchid Biotechnology II.* World Scientific. Singapore. pp. 1-24.

Varsney RK, Aggarwal RK, Thudi M, Borner A (2007) Genic molecular markers in plants: development and applications. *In:* Varsney RK and Tuberosa R. Genomics-Assisted Crop Improvement: Vol. 1: *Genomics Approaches and Platforms*, 13-29. Springer.

Wan H, Qian C, Malik AA, Zhao Z, Chen J (2009). Isolation, phylogeny, and evolutionary analysis of *Pto*-type disease resistance gene analogs from a *Cucumis hystrix* introgression line of cucumber (*C. sativus*). *Funct. Plant Biol.* 3: 513-523.

Wey GC (1988). Occurrence and investigation of important diseases on *Phalaenopsis* in Taiwan. *Reports of Taiwan Sugar Research Institutes* 122: 31-41. https://www.cabi.org/ISC /abstract/19901142252.

Wiersema JH (2019). *Phalaenopsis lamelligera* H.R.Sweet. GRIN Taxonomy. US National Plant Germplasm System. Checklist dataset https://doi.org/10.15468/ao14pp accessed via GBIF.org on 2021-11-06.

Xiao L, Quan M, Du Q, Chen J, Xie J, Zhang D (2017). Allelic interactions among *Pto*-MIR475b and its four target genes potentially affect growth and wood properties in *Populus*. *Front. in Plant Sci.* 8: 1055. doi: 10.3389/fpls.2017.01055.

Xie DF, Yu HX, Price M, Xie C, Deng YQ, Chen JP, Yu Y, Zhou SD, He XJ (2019). Phylogeny of Chinese Allium species in section *Daghestanica* and adaptive evolution of *Allium* (Amaryllidaceae, Allioideae) species revealed by the chloroplast complete genome. *Front. Plant Sci.* https://doi.org/10.3389/ fpls.2019.00460.

Xu Y (2010). Molecular plant breeding. Wallingford: CABI.

Yun SA, Son HD, Im HT, Kim SC (2020). Genetic diversity and population structure of the endangered orchid *Pelatantheria scolopendrifolia* (Orchidaceae) in Korea. *PLoS ONE* 15(8): e0237546. https://doi.org/ 10.1371/journal. pone.0237546

Zhai W, Zhao Y, Zhang LX, Li XJ (2014). Structural and phylogenetic analysis of *Pto*-type disease resistance gene candidates in *Hevea brasiliensis*. *Genet. Mol. Res.* 13: 4348-4360.

Zhang Y, Wang L (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* 5 doi:10.1186/1471-2148-5-1.