**BIODIVERSITAS**
Volume 20, Number 3, March 2019
Pages: 696-703

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d200312

# The genetic diversity and population structure of shallots (*Allium cepa* var. *aggregatum*) in Indonesia based on *R gene*-derived markers

**LINA HERLINA[1,2,♥], REFLINUR[1], SOBIR[2], AWANG MAHARIJAYA[2], SURYO WIYONO[3]**

[1]Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development. Jl. Tentara Pelajar No. 3A Cimanggu, Bogor 16111, West Java, Indonesia. Tel.: +62-251-8338820, 8333440, 8345975. Fax.: +62-251-8338820, ♥email: tydars@yahoo.com
[2]Department of Agronomy and Horticulture, Faculty of Agriculture, Institut Pertanian Bogor. Jl. Raya Dramaga, Bogor 16680, West Java, Indonesia
[3]Department of Plant Protection, Faculty of Agriculture, Institut Pertanian Bogor. Jl. Raya Dramaga, Bogor 16680, West Java, Indonesia

Manuscript received: 25 October 2018. Revision accepted: 17 February 2019.

**Abstract.** *Herlina L, Reflinur, Sobir, Maharijaya A, Wiyono S. 2019. The genetic diversity and population structure of shallots (Allium cepa var.* aggregatum*) in Indonesia based on R gene-derived markers. Biodiversitas 20: 696-703.* Molecular markers are very useful for determining plant genetic diversity, especially when they are associated with useful traits. In shallots, type of markers still very limited. Therefore, development of new molecular markers is challenging if the future demand for this crop is to be met. The present study reports the genetic diversity analysis of 36 accessions of shallots using molecular markers derived from partial *R* gene sequences. Six polymorphic *R* gene-derived markers (Acepa-1, Acepa-2, Acepa-3, Acepa-4, Acepa-5, and Acepa-6) were developed and applied to analyse the genetic diversity of shallots in Indonesia. These loci produced a total of 963 alleles with an average of 0.51 $\pm$ 0.77 alleles per marker. The polymorphic information content value ranged from 0.2784 to 0.5236 with an average of 0.36015. The markers were able to differentiate the shallot genotypes into three major clusters, suggesting that shallots are characterized by poor levels of genetic diversity in Indonesia. After further validation, the markers will be very useful, serving as tools to support the breeding of disease-resistant shallot varieties.

**Keywords:** *Allium cepa*, biotechnology, genetic diversity, developed marker, plant breeding, red onion

**Abbreviations:** PH: plant height, TBW: total bulb weight, NBP: number of bulbs per plant, WB: bulb weight per plant, PPI: percentage of plant invested FOC, NL: number of leaves; Tl: total loci, Neff: Σ allele effective, pl: polymorphic loci, Av-all: average allele per loci, Av-alm: average per marker

## INTRODUCTION

Shallot (*Allium cepa var. aggregatum* Don.), commonly called red onion, is a member of the Alliaceae family and is usually consumed as one of the primary culinary spices in Indonesia (MoA 2015). Shallots have become one of the most multi-functional vegetables in the world since 1600 BC (Messiaen and Rouamba 2004). The tubers are often used as medicines, processed into pickles, or eaten raw (Grubben et al. 2004; NOA 2011, Deviana et al. 2014). The skin of the tubers may be used as a dye, and the leaves are consumed as vegetables (Messiaen and Rouamba 2004). Shallots also show excellent economic prospects because they can be sold in processed form—such as shallot extract, powder, essential oil, and fried shallots (Grubben et al. 2004)—because their nutrient content (vitamins, minerals, and antioxidants) is low (Messiaen and Rouamba 2004; NOA 2011; Smith 2013).

One of the main constraints of shallot production in Indonesia is the incidence of *Fusarium oxysporum*, also known as bulb rot (Isniah and Widodo 2015; MoA 2015; Udiarto et al. 2015), which causes losses of shallot yield up to 100%. Development of disease-resistant crops is the most appropriate and prospective answer to this problem. Unfortunately, there is a scarcity in the source of shallot varieties which show high resistance to disease caused by *Fusarium* in field owing to the lack of information regarding the genetic diversity of this crop in Indonesia.

The rapid growth of genomic research has enabled the discovery and development of numerous promising new alternative molecular marker techniques in plant genetics. The rapid expansion of public genomic databases has simplified the development of DNA as functional markers (Poczai 2013). Numerous genes which are directly or indirectly associated with the regulation of plant defense and resistance mechanisms have been identified, such as the *R* gene, which plays an important role in induction of plant defense signaling, particularly to recognize specific pathogen effectors (Belkhadir et al. 2004). Those gene sequences isolated from crops have been identified, published, and are registered in the NCBI database. Rapid and remarkable advances in biotechnology have successfully characterized and introduced genes associated with plant defense and resistance mechanisms to create new resistant crop varieties. Of these, *RPM1*, *CYP79B2*, *CYP79B3*, *CES101*, *NP24*, and *L4* are among the most extensively studied genes.

It has been reported that the *RPM1* gene plays a role in the hypersensitive response mechanism of plants to pathogenic infections. RPM1 (CC-NB-LRR) acts as a signal receptor which recognizes molecules released by pathogens (Grant et al. 1995). NP24 is a thaumatin-like

protein isoform found in tomatoes with a size of 24 kDa that plays an important role in antifungal activity (Jia and Martin 1999). CES101 is a G-type lectin receptor kinase that regulates the expression of genes involved in plant photosynthesis and generation of callus tissue (Salanoubat et al. 2000). L4 is another NBS-LRR gene that is *R* gene-like (RGL) and is also associated with plant disease resistance (Aarts et al. 1998).

Crop improvement research programs have used genes associated with plant defense and resistance to develop disease-resistant varieties as an environmentally friendly and convenient alternative to protect crops from diseases. However, the introgression of resistance genes from one species to another often poses challenges, as it requires long periods of time to perform successful backcrossing (Gurunani 2012). Studies of the function, cloning, characterization, and genetic transformation of plant resistance genes could help researchers to overcome these problems in the near future.

The present study aims to analyze the genetic diversity of shallots (*Allium cepa var. aggregatum*) using markers derived from partial *R* gene sequences to support the breeding program of *Fusarium*-resistant shallot varieties in Indonesia.

## MATERIALS AND METHODS

### Plant materials

Thirty-six shallot genotypes used for molecular characterization in the present study are listed in Table 1. Sumenep was the only cultivar previously known to be resistant to several kinds of diseases and expected to have a high resistance to *Fusarium oxysporum* (Baswarsiati et al. 2010). Experiments were conducted at Laboratorium of Indonesian Center of Agricultural Biotechnology and Genetic Resource Research and Development (ICABIOGRAD), Bogor, Indonesia.

### Procedures

*Agro-morphological characterization*

Characterization was applied to 36 shallot genotypes planted in the screen house of the ICABIOGRAD, at Cikeumeuh, Bogor Indonesia, started from May 2016 to August 2017. The research was arranged in a Randomized Complete Group Design with 3 replicates, each containing 20 plants. Quantitative observations were conducted according to the *Calibration Book Onion and Shallot* (Naktuinbouw 2010), with adjustments applied as needed. Quantitative traits were plant height (PH), total bulb weight (TBW), number of bulbs (NBP), weight of bulb plant$^{-1}$ (WB), number of leaves (NL), and percentage of plant infested by FOC (PPI). The results of the quantitative characterization were used to perform the correlation analysis with several molecular marker indices.

*Primer design and marker development*

The primers were designed based on partial sequences of six genes: *CYP79B2* ( (NCBI ID 830154)) (primer set Acepa-1), *CYP79B3* (NCBI ID: 816765) (primer set Acepa-2), *RPM1* (NCBI ID: 819889) (primer set Acepa-3), *NP24 (NCBI* ID: 543979) (primer set Acepa-4), *CES101* (NCBI ID: 820848) (primer set Acepa-5), and *L4* (NCBI ID AAC14559.1) (primer set Acepa-6). Sequences from amplicons were searched using BLAST (Basic Local Alignment Tool, developed by the NCBI-NIH, and freely accessible at https: //www.ncbi.nlm.nih.gov/BLAST/ and carefully selected; only those primers sharing sequence similarity to any gene sequence found in *Allium* sp. were considered as a marker. The parameters for primers were as follows: product size of 100 to 300 bp; primer size of 18 to 22 bp, with optimal length of 20 bp; primer melting temperature (Tm) of 50°C to 65°C, with an optimum at 55°C. The markers were designed using Primer3Plus (Table 2).

**Table 1**. Shallot genotypes use in the present study

| Code of genotype | Name of genotypes | Site of collection | No | Code of genotype | Name of genotypes | Site of collection |
|---|---|---|---|---|---|---|
| A-01 | BM Bandung | West Java | 19 | BM-15 | BM Bogor-2 | West Java |
| BM-2 | BM Yuwono | Central Java | 20 | BM-16 | BM Demak | Central Java |
| BM-3 | BM Maja | West Java | 21 | A-21 | BM Sumenep | East Java |
| BM-4 | BM Kramat-1 | Central Java | 22 | BM-17 | BM Bogor-3 | West Java |
| BM-5 | BM Boyolali | East Java | 23 | BM-18 | BM Bogor-4 | West Java |
| A-06 | BM Tajuk | East Java | 24 | BM-19 | BM Brebes | Central Java |
| BM-6 | BM Kramat-2 | West Java | 25 | BM-20 | BM-20 | West Java |
| BM-7 | BM AG-1 | West Java | 26 | BM-21 | BM-21 | West Java |
| BM-8 | BM Trisula | West Java | 27 | BM-22 | BM-22 | West Java |
| A-10 | BM Sumenep-I | East Java | 28 | BM-23 | BM-23 | West Java |
| A-11 | BM Bali Karet | East Java | 29 | A-29 | BM Sumenep-II | West Java |
| BM-9 | BM AG-2 | West Java | 30 | BM-24 | BM-24 | West Java |
| BM-10 | BM Sembrani | East Java | 31 | BM-25 | BM BT Ijo | East Java |
| A-14 | BM Manjung | West Java | 32 | BM-26 | BM Biru Lancor | East Java |
| BM-11 | BM Pikatan | West Java | 33 | BM-27 | BM Bauji | East Java |
| BM-12 | BM Mentes | Central Java | 34 | BM-28 | BM SuperPhillips | East Java |
| BM-13 | BM Katumi | East Java | 35 | A-35 | BM Bima Curut | West Java |
| BM-14 | BM Bogor-1 | West Java | 36 | A-36 | BM Bombay | West Java |

*DNA extraction*

The total genomic DNA was extracted from 36 shallot genotypes freshly harvested from a greenhouse at ICABIOGRAD. The bulbs were ground into a fine powder in liquid nitrogen using a mortar and pestle. The DNA was extracted using the *Cetyl Trimethyl Ammonium Bromide* (CTAB) method as described by Doyle and Doyle (1987). The quality and concentration of extracted DNA were estimated using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Walthum, MA). The DNA was then separated on a 1% agarose gel (Thermo Scientific, 2012), stained with SYBR Safe DNA gel stain, and visualized under ultraviolet light (UVP BioImaging Systems, Upland, CA). DNA was stored at-20°C until further use.

*PCR amplification and genotyping*

PCR amplification of the markers was performed using standard PCR to determine the appropriate annealing temperature. PCR amplifications were performed in a total volume of 25 uL containing 10 uL KAPA FAST 2G PCR premix with dye (Thermoscie Genetic Science), 8 uL ddH2O, 0.05 to 0.15 uM forward primer, 0.05 to 0.15 uM reverse primer, and 2 uL template DNA (20 ng $uL^{-1}$). Amplification was performed in a 96-well GeneAmp_PCR System 9700 (Applied Biosystems) under the following cycle program: initial denaturation step for 4 min at 94°C, followed by 40 cycles at 94°C for 30 s (denaturation), 46°C-56°C for 45 s (annealing), and 72°C for 120 s (extension), followed by a final extension step at 72°C for 7 min. This cycle refers to Martanti et al (2015) with some modifications. To reveal polymorphism and allele identification, PCR products were separated on a 1.5% agarose gel containing 1% TBE (45 mM Tris-borate, 1 mM EDTA) and 0.5 mg $ml^{-1}$ aqueous solution of ethidium bromide. The gel was then visualized with a UV trans-illuminator at 300 nm. To confirm the reproducibility of the banding patterns, the PCR experiments were repeated twice.

**Data analysis**

Developed markers: *R* gene sequence-derived markers were used to analyze the genetic diversity of 36 shallot genotypes. Amplified bands from each marker were scored as present (1) or absent (0). Only those bands that amplified consistently were considered; smeared and weak bands were excluded from the analysis. The basic statistics, such as number of amplified loci (n), the maximum (Max) and minimum (Min) band size, the percentage of polymorphic loci (Ppl), the average of observed (na) and effective (ne) alleles, loci and their standard deviations (SD) were determined using POPGENE version 1.31 (Yeh et al. 1999). The percentage of polymorphism (Pp) was given as the number of polymorphic loci/number of total loci, regardless of allele frequencies. Polymorphic information content (PIC) values were calculated using the PICcalc online program (liverpool.ac.uk/~kempsj/pic.html*)*. The dendogram was constructed using Unweighted Neighbor-Joining in DARwin version 6.0.010 (http: //darwin.cirad.fr) (Perrier et al. 2003). The resulting tree was bootstrapped under 10,000 replicates (Felsenstein 1985). The pattern of population structure and detection of admixture were inferred using Bayesian model-based clustering algorithm implemented in STRUCTURE version 2.3.4. (Pritchard et al. 2000; Falush et al. 2003). Analysis was run without prior information about the population by allowing the STRUCTURE software to assign each individual to a population (Pritchard et al. 2000). The admixture model with correlated allele frequencies was used as suggested in the software manual. To determine the most appropriate number of populations (K), a burn-in period of 5,000 was used in each run, and data were collected over 100,000 Markov Chain Monte Carlo replications from K = 1 to K = 10 with 3 iterations. The K value was determined by the log probability of data (LnP (D)) based on the rate of change in LnP (D) between successive Ks. The optimum K value was predicted following the simulation method of Evanno et al. (2005) using the web-based software STRUCTURE HARVESTER version 0.6.92 (Earl and vonHoldt 2005).

## RESULTS AND DISCUSSION

**Molecular marker characterization**

During the initial step, we explored 6 genes which are known to be associated with defense or resistance mechanisms in several crops using the NCBI database (www.ncbi.nlm.nih.gov/gene). We designed primers using Primer3Plus (bioinformatics.nl/primer3plus) according to these genes sequences. Each primer was designed to amplify gene fragments between 150 and 250 bp in size, because shorter gene fragments will more easily yield many amplicons and increase the probability of getting polymorphic bands. These primers were named Acepa-1, Acepa-2, Acepa-3, Acepa-4, Acepa-5, and Acepa-6 respectively (Table 2). In the second step, we amplified the genomic DNA of the shallot genotype using the primers listed in Table 2. The amplicons were then detected using 1% agarose gel electrophoresis.

**Molecular marker characterization**

All the primers revealed marker polymorphisms in shallot genotypes, with the maximum percentage of polymorphic loci obtained from the Acepa-1 (75%) and the minimum percentage obtained from Acepa-5 (33.33%). All the markers were polymorphic, and the PIC ranged from 0.2562 to 0.5236, with an average of 0.36015. The highest PIC value was obtained from marker Acepa-2, and these loci produced a total of 963 alleles with an average of 0.51 (Table 3).

**Genetic variability of *R* gene-derived markers**

The highest percentage of polymorphic loci was obtained in shallot BM-4, whereas the lowest was in shallot BM-35 (Table 4). In agreement with the polymorphism results, BM-35 showed the lowest average allele per marker, whereas the highest was generated by BM-4.

**Table 2**. List of molecular markers for assessing genetic diversity of shallots in Indonesia

| Primer name | Sequence | Target-band size (bp) |
|---|---|---|
| Acepa-1 | TCCCACATTTTCCTCACTCC//AAAGGACGATGTTTCGGTTG | 226 |
| Acepa-2 | CGTGGCACTCTCTGATACGA//GCAGCACATCCTCTCTTTCC | 241 |
| Acepa-3 | GTATATGTGGGCGAGGCACT//CTTAGAAGCCGTCCGATGAG | 173 |
| Acepa-4 | TGTACCACGTTTGGAGGACA//ATCTCCAAGGGGAAATTTGG | 152 |
| Acepa-5 | GTGGTAGACCAAGGGAACGA//GCCTCCTCCGTATGTAACCA | 241 |
| Acepa-6 | TTACCCGGAAGAATGTGTGG//GGGTGGAAATATTGGCTTGA | 210 |

**Table 3.** Diversity indices of the *R* gene markers loci used in this study

| Primer | Neff | pl | n | Ppl | Max -b | Min -b | Σ band | Fb | pi | PIC value |
|---|---|---|---|---|---|---|---|---|---|---|
| Acepa-3 | 2 | 5 | 7 | 71.43 | 1500 | 100 | 33 | 0.042 | 16,50 | 0.2784 |
| Acepa-5 | 2 | 4 | 12 | 33.33 | 1500 | 70 | 44 | 0.056 | 22,00 | 0.3703 |
| Acepa-4 | 2 | 2 | 3 | 66.67 | 1500 | 100 | 27 | 0.034 | 13,50 | 0.2562 |
| Acepa-6 | 2 | 2 | 4 | 50.00 | 900 | 100 | 46 | 0.058 | 23,00 | 0.3574 |
| Acepa-1 | 2 | 6 | 8 | 75.00 | 900 | 100 | 42 | 0.053 | 21,00 | 0.5236 |
| Acepa-2 | 3 | 4 | 7 | 57.14 | 1000 | 100 | 52 | 0.066 | 17,33 | 0.3750 |
| Average | 2.17 | 3.836,83 | 58.93 | 1216.6 | 95 | 40.67 | 0.05 | 18.89 | 0.36015 | |

Note: Neff: Σ allele effective, pl: Σ loci polimorfix, n: Σ Total loci, Ppl: percentage of polymorphic loci, Max-b: Maximal band size, Min-b: Minimal band size, Fb: Frequency of band, pi: Allele frequency

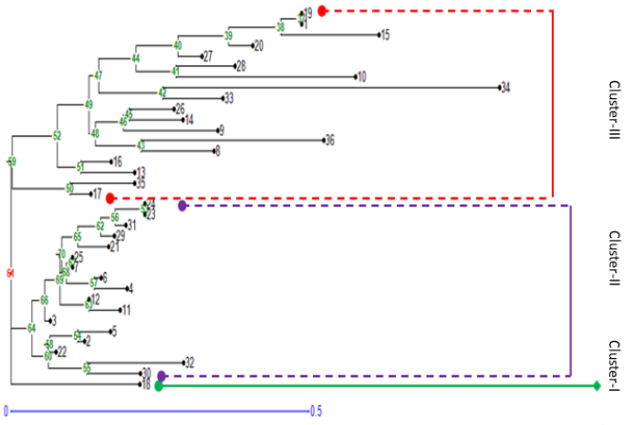


**Figure 1**. Dendrogram for clustering shallot genotypes based on Unweighted Neighbor-Joining

## Genotype clustering

Unweighted Neighbor-Joining was used to carry out the clustering analysis, and to generate a dendrogram showing the relationship among the genotypes (Figure 1). The resulting image showed that there were three main branching nodes, with shallot genotypes separated into three main clusters.

## Population structure of Indonesian shallot

Six markers designed based on the *R* gene sequence were distributed throughout the shallot genome to assess genetic architecture and population structure. It was observed that delta K declined after K = 3, and was continuously significant (Figure 2), suggesting the presence of three subgroups (Figure 3). At K = 3, most genotypes were classified into three groups. The degree of admixture (alpha) generated in this study was close to zero (0.03), meaning that most individuals are essentially derived from one population.

**Table 4**. Genetic diversity indices of shallot genotypes based on *R* gene-*based* marker in this study

| Accession | Total loci | Σ allele effective (*Neff*) | % polimorphic loci (pl) | Average allele per loci ± SD | Average allele per marker ± SD |
|---|---|---|---|---|---|
| A-01 | 33 | 10 | 3.43 | 0.77+ 0.82 | 1.67+0.14 |
| BM-2 | 20 | 3 | 2.08 | 0.23+ 0.84 | 0.50+0.14 |
| BM-3 | 30 | 8 | 3.12 | 0.62+ 0.82 | 1.33+0.14 |
| BM-4 | 39 | 10 | 4.05 | 0.77+ 0.52 | 1.67+0.09 |
| BM-5 | 22 | 7 | 2.28 | 0.54+ 0.75 | 1.17+0.13 |
| A-06 | 22 | 7 | 2.28 | 0.54+ 0.98 | 1.17+0.16 |
| BM-6 | 30 | 8 | 3.12 | 0.62+ 0.52 | 1.33+0.09 |
| BM-7 | 35 | 9 | 3.63 | 0.69+ 0.55 | 1.50+0.09 |
| BM-8 | 24 | 5 | 2.49 | 0.38+ 0.75 | 0.83+0.13 |
| A-10 | 27 | 6 | 2.80 | 0.46+ 0.63 | 1.00+0.11 |
| A-11 | 19 | 3 | 1.97 | 0.23+ 0.84 | 0.50+0.14 |
| BM-9 | 31 | 9 | 3.22 | 0.69+ 1.05 | 1.50+0.17 |
| BM-10 | 38 | 10 | 3.95 | 0.77+ 0.82 | 1.67+0.14 |
| A-14 | 25 | 7 | 2.60 | 0.54+ 0.75 | 1.17+0.13 |
| BM-11 | 26 | 5 | 2.70 | 0.38+ 0.75 | 0.83+0.13 |
| BM-12 | 20 | 2 | 2.08 | 0.15+ 0.82 | 0.33+0.14 |
| BM-13 | 22 | 7 | 2.28 | 0.54+ 0.75 | 1.17+0.13 |
| BM-14 | 32 | 8 | 3.32 | 0.62+ 0.52 | 1.33+0.09 |
| BM-15 | 27 | 6 | 2.80 | 0.46+ 1.10 | 1.00+0.18 |
| BM-16 | 17 | 3 | 1.77 | 0.23+ 0.84 | 0.50+0.14 |
| A-21 | 18 | 4 | 1.87 | 0.31+ 0.82 | 0.67+0.14 |
| BM-17 | 37 | 10 | 3.84 | 0.77+ 1.03 | 1.67+0.17 |
| BM-18 | 37 | 9 | 3.84 | 0.69+ 0.55 | 1.50+0.09 |
| BM-19 | 28 | 7 | 2.91 | 0.54+ 0.98 | 1.17+0.16 |
| BM-20 | 25 | 7 | 2.60 | 0.54+ 0.98 | 1.17+0.16 |
| BM-21 | 26 | 9 | 2.70 | 0.69+ 0.55 | 1.50+0.09 |
| BM-22 | 27 | 7 | 2.80 | 0.54+ 0.98 | 1.17+0.16 |
| BM-23 | 22 | 5 | 2.28 | 0.38+ 0.75 | 0.83+0.13 |
| A-29 | 32 | 7 | 3.32 | 0.54+ 0.98 | 1.17+0.16 |
| BM-24 | 32 | 8 | 3.32 | 0.62+ 0.82 | 1.33+0.14 |
| BM-25 | 26 | 8 | 2.70 | 0.62+ 1.03 | 1.33+0.17 |
| BM-26 | 30 | 8 | 3.12 | 0.62+ 1.03 | 1.33+0.17 |
| BM-27 | 25 | 8 | 2.60 | 0.62+ 1.03 | 1.33+0.17 |
| BM-28 | 22 | 5 | 2.28 | 0.38+ 0.75 | 0.83+0.13 |
| A-35 | 8 | 1 | 0.83 | 0.08+ 0.41 | 0.17+0.07 |
| A-36 | 19 | 6 | 1.97 | 0.46+ 0.00 | 1.00+0.00 |
| BM-30 | 10 | 2 | 1.04 | 0.15+ 0.52 | 0.33+0.09 |

SD: standard deviation

**Table 5.** Clustering based on molecular analysis

| Cluster | Discriminate value at node | Code of genotypes | Name of genotype | Number |
|---|---|---|---|---|
| I | 59 | 17, 35, 13, 16, 8, 36, 9, 14, 26, 33, 34, 10, 28, 27, 20, 15, 1, 19 | Katumi, Bima Curut, Sembrani, Demak, AG-1, Bombay, Trisula, Tajuk, Manjung, Bauji, SuperPhillips, Sumenep-1, B23, B22, Demak, Pikatan, Bandung, Bogor-2 | 18 |
| II | 64 | 3, 11, 12, 4, 6, 7, 25, 21, 29, 31, 23, 24, 22, 2, 5, 30, 32 | Maja, Bali Karet, AG-2, Kramat-1, Tajuk, Kramat-2, BM-20, Sumenep, Sumenep-II, Batu Ijo, BM-18, Bima Brebes, BM17, Yuwono, Boyolali, BM24 | 17 |
| III | 64 | 18 | BM-14 | 1 |

Note: *Significant at $p < 0.001$

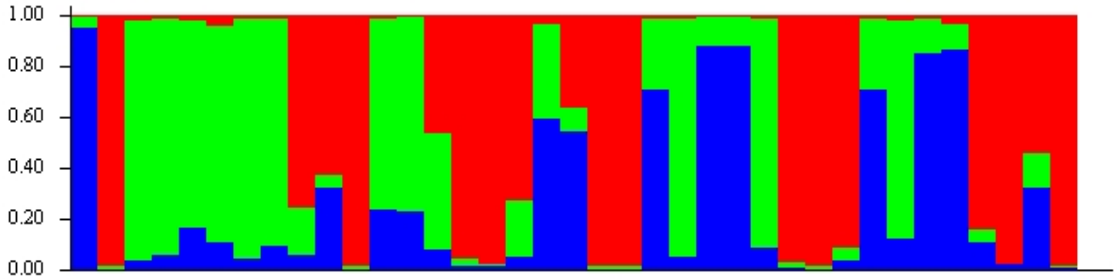


**Figure 3**. Bayesian model-based estimation of population structure (K = 3) for the 36 shallot genotypes. Groups are separated into different colors. Numbers on the y-axis show coefficient of membership/assignment

**Table 6.** Matrix correlation between quantitative traits and molecular indices

| | PH | TWB | NBP | WB | PPI | NL | Tl | Neff | pl | Av-all |
|---|---|---|---|---|---|---|---|---|---|---|
| **TWB** | 0.685*<br>0.000 | | | | | | | | | |
| **NBP** | 0.233<br>0.223 | 0.623*<br>0.000 | | | | | | | | |
| **WB** | 0.809*<br>0.000 | 0.803*<br>0.000 | 0.500<br>0.006 | | | | | | | |
| **PPI** | -0.204<br>0.290 | 0.123<br>0.526 | 0.177<br>0.359 | -0.287<br>0.131 | | | | | | |
| **NL** | 0.487<br>0.009 | 0.635*<br>0.000 | 0.565<br>0.002 | 0.492<br>0.008 | 0.306<br>0.113 | | | | | |
| **Tl** | 0.258<br>0.176 | 0.314<br>0.097 | -0.040<br>0.836 | 0.361<br>0.054 | 0.052<br>0.789 | 0.165<br>0.400 | | | | |
| **Neff** | 0.347<br>0.065 | 0.309<br>0.103 | -0.060<br>0.759 | 0.249<br>0.193 | 0.130<br>0.502 | 0.190<br>0.333 | 0.828*<br>0.000 | | | |
| **pl** | 0.258<br>0.176 | 0.314<br>0.097 | -0.040<br>0.836 | 0.361<br>0.054 | 0.052<br>0.789 | 0.165<br>0.400 | 1<br>0.000 | 0.828*<br>0.000 | | |
| **Av-all** | 0.344<br>0.068 | 0.307<br>0.105 | -0.063<br>0.744 | 0.245<br>0.200 | 0.129<br>0.506 | 0.189<br>0.334 | 0.826*<br>0.000 | 0.999*<br>0.000 | 0.826*<br>0.000 | |
| **Av-alm** | 0.345<br>0.066 | 0.309<br>0.103 | -0.059<br>0.760 | 0.246<br>0.194 | 0.130<br>0.500 | 0.188<br>0.338 | 0.828*<br>0.000 | 0.999*<br>0.000 | 0.828*<br>0.000 | 0.999*<br>0.000 |

Note: numbers in first line indicate Pearson correlation value; whereas in second line indicate *p*-value. *PH*: plant height, *TBW*: total bulb weight, NBP: number of bulbs per plant, *WB*: bulb weight per plant, *PPI*: percentage of plant invested FOC, *NL*: number of leaves; *Tl:* total loci, *Neff*: Σ allele effective, *pl*: polymorphic loci, *Av-all*: average allele per loci, *Av-alm*: average per marker. * Significant at p-value $\leq 0.05$

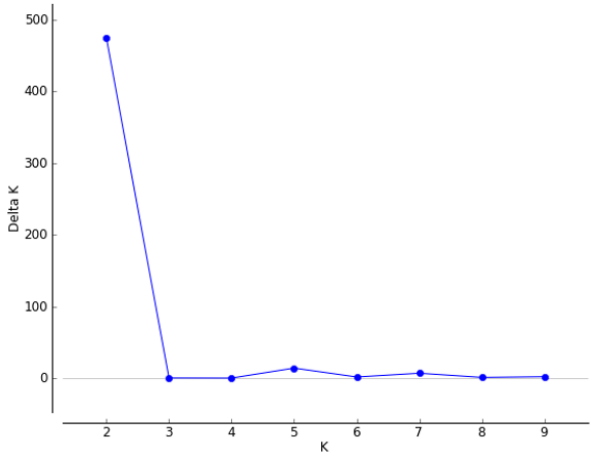


**Figure 2**. The presence of K = 3 and Delta K = mean (|L'' (K)|)/sd (L (K)), K ranged from 2-9

**Correlation analysis between quantitative traits and R-gene based marker**

We investigated the correlation between 6 quantitative traits of shallots and four *R* gene marker indices to reveal the relationship between the defense attribute (indicated by the existence of allele diversity of *R* gene-derived markers) and the plant performance. Based on matrix correlation, a total of 40 pairwise combinations were formed among ten variables, of which nine combinations were found to be very significant ($p < 0.01$). Most combinations were positively correlated, and five were negatively correlated.

**Discussion**

Research and development of molecular markers has experienced rapid progress because of advances in genomics. This advance has been accompanied by data and information disclosure and ease of worldwide access to genomic data, especially from various plant species. Molecular markers are preferred tools in plant molecular breeding because of their polymorphic nature, co-dominance, selective neutral behavior. In addition, the assays for these markers are easy and fast with high reproducibility, and data may be easily exchanged between laboratories (Joshi et al. 1999). The outstanding progress of bioinformatics tools and databases has also contributed to the advanced development of gene-based markers (Varshney 2010).

The information on shallot genetic diversity and population structure are important to elucidate the potential for advanced breeding of this crop, especially to create varieties resistant to *Fusarium* bulb rot. The present study is the first report on the genetic variability of shallots of Indonesia-based on molecular markers derived from several *R* gene sequences.

According to BLAST results, we found six genes which were candidates for markers: *RPM1*, *CES101*, *CYP79B2*, *CYP79B3*, *NP24*, and *L4*. As mentioned above, RPM1 (CC-NB-LRR) acts as a signal receptor that recognizes molecules released by pathogens. When inactive, signal competent RPM1 is a protein associated with the plasma membrane (Aarts et al. 1998), and is thought to play a role in cytosolic $Ca^{2+}$ in plasma membranes, particularly as it relates to the hypersensitive response mechanism of plants to pathogenic infections. The *RPM1* gene plays an identified role in the disease caused by *Pseudomonas* sp.; RPM1 interacts with the plasma membrane, and localizes another protein, RIN4. RPM1 recognizes RIN4-mediated effector modification in the presence of an effector type III bacterium protein AvrRpm1 or AvrB (Aarts et al. 1998). AvrRpm1 and AvrB are also localized to the host plasma membrane by acylation; RIN4 is subsequently modified, thereby activating RPM1 (Aarts et al. 1998).

As been mentioned before NP24, CES101, L4, and CYPB79 are proteins associated with plant defense mechanism. NP24 is a thaumatin-like protein isoform found in tomatoes with size 24kDa which played an important role in antifungal activity (Jia and Martin 1999), CES101 is involved in plant photosynthesis and generation of callus tissue (Salanoubat et al. 2000) and L4 is *R* gene-like which associated with plant disease resistance (Aarts et al. 1998). In *Arabidopsis* Col and Ler, genomic DNA isolation using degenerate primers allows the isolation of the L4 fragment from chromosome 1. CYP79B2 and CYP79B3 monooxygenases are responsible for the redundant conversion of Trp to indole-3-acetaldoxime, which is indispensable for the synthesis of camalexin, indole glucosinolates, and subsequently, PEN2-related metabolites (Zhao et al. 2002; Bednarek et al. 2009). In *Arabidopsis*, the synthesis of Trp-derived metabolites is essential for post-invasive resistance to *Colletotrichum gloeosporioides* at the non-adapted hemibiotrophic stage (Hiruma et al. 2013).

There are six *R* gene-derived markers which were successfully developed during the present study (Acepa-1, Acepa-2, Acepa-3, Acepa-4, Acepa-5, and Acepa-6), and were used to assay the genetic diversity of shallots in Indonesia. Using genes that play roles in plant disease resistance as a reference for designing molecular markers, we aimed to increase the probability of obtaining markers associated with disease resistance. In this case, the markers developed require further testing for additional validation by implementing their use in a larger population. In the present study, the association of markers with plant disease resistance was not carried out because the number of shallot genotypes used was relatively limited, as was the source of resistance genes (in this case against *Fusarium* tuber rot disease), which were obtained only from the Sumenep shallot genotype.

During the molecular analysis of shallot genetic indices, a total of 963 well-resolved band classes were observed. The amplified fragments ranged from 70 bp (marker Acepa-5) to 1500 bp (marker Acepa-3, Acepa-5 and Acepa-4) in size (Table 3). The number of bands obtained for each primer varied from 27 (marker Acepa-4) to 52 (marker Acepa-2), with an average of 40.67 bands per primer. All primers revealed marker polymorphisms in shallot genotypes, with the maximum percentage of polymorphic loci obtained from the Acepa-1 (75%) and the minimum percentage obtained from Acepa-5 (33.33%). Marker Acepa-5 also showed the highest number of observed and effective alleles (2.00 and 1.834,

respectively), implying that this primer is most appropriate for genetic diversity analysis. The polymorphisms identified were used to generate genetic dissimilarity. These loci produced a total of 963 alleles, with an average of 0.51 $\pm$ 0.77 alleles per marker. Selection of *R* gene-derived markers with a range of polymorphisms may reduce the risk of overestimating genetic variability. This is in agreement with a study on genetic diversity using microsatellite markers in sheep as reported by Sharma et al. (2016).

In accordance with molecular analysis of the shallot genetic indices, marker characterization showed similar results; all the markers were polymorphic, with an average PIC value of = 0.36015 (Table 3). One marker obtained from Acepa-1 was highly informative, with a PIC value reaching 0.5236. The PIC value provides an estimate of the discriminatory power of a locus by taking into account the number and the relative frequencies of the alleles. The higher the PIC value, the more effective the marker in discriminating individuals within population.

The Unweighted Neighbor-Joining dendrogram grouped the 36 shallot genotypes into three major clusters (Figure 1). Of the 36 genotypes, 1, 17, and 18 genotypes were grouped together in Clusters I, II, and III respectively (Table 5). The overall topology of the dendrogram indicated that the shallots of Indonesia may be divided into three lineages associated with six functional genes that may contribute to plant defense or resistance functions, i.e. *RPM1*, *CES101*, *CYP79B2*, *CYP79B3*, *NP24*, and *L4*. The pattern of clustering was similar to that reported by Evanno et al. The method using STRUCTURE outputs predicted K = 3 to be the most likely number of clusters (Figure 2), although each cluster was composed of different shallot genotypes.

Based on the dendrogram (Figure 1), it is quite interesting that the BM-14 occupies a separate cluster from the other genotypes. This genotype, based on the characteristics of genetic diversity indices has a higher polymorphism and total number of loci (Table 4). Interestingly, based on these characteristics, this genotype almost similar to BM-3 in terms of high polymorphism, high total number of loci and Neff. However, when we carefully examined in more detail (data not shown), it turns out that both of them have differences in terms of markers that contribute to the polymorphism. The contribution of high polymorphism and the highest total loci in genotype BM-14 are the result of alleles contribution encoded by CES101; whereas BM-3 does not have any allele from CES101. This difference in the arrangement of polymorphisms, therefore, places BM-14 in a separate cluster from others. Allelic diversity is an important component especially in the response of plants to adapt to changes in the environment (Caballero and Garcia-Dorado 2013). In this case, even though it requires further confirmation and testing, BM-14 has the potential to be a prospective parent for breeding new variety of shallot which resistance to Fusarium disease.

According to correlation analysis, the Nbp (number of bulbs plant$^{-1}$) was negatively correlated with all marker indices (Table 6), although the correlation was not significant. However, from these results, we may draw an interesting conclusion that the developed markets are not significantly correlated with quantitative traits. Overall, the genetic diversity indices of shallot genotypes based on *R* gene-derived markers in the present study showed poor diversity among the studied genotypes.

There are several arguments that may explain the low genetic diversity obtained from the clustering using the developed markers. First, the number of shallot accessions used in this study is limited, as all accessions were collected exclusively from Java Island. The limitation of the germplasm source use in this study may contribute to the limited genetic diversity observed. Second, shallots in Indonesia are generally propagated vegetatively (although the species belongs to a cross-pollinated plant). Cultivation using vegetative propagation does not require the flowering process to produce planting material for the next growing season; therefore, opportunities for natural hybridization are very limited. Most likely, the Indonesian farmer uses shallot tubers (bulbs) as "seedlings" for planting the crop. According to Syukur (2012), cultivating plants using vegetative propagation will almost certainly produce generations of plants that are phenotypically similar to their vegetative parent, and the apparent differences in plant morphology do not necessarily reflect genetic differences. Those morphological differences usually result from variations that occur during the plants' adaptation to their environment, which in this case are not genetically inherited. This also offers an explanation for the problems which often occur in these crops: the highly contagious *Fusarium* diseases spread easily between shallot crops in Indonesia and are difficult to control. As phenotypically observed, the low genetic variations in this crop indicate poor genetic diversity, especially in plant defense genes composition.

One important requirement in a strategy for pyramiding genes to create resistant crops is the availability of wide genetic diversity/variations in germplasm. This ensures the availability of the targeted genes as donors for resistance to facilitate introgression into the genomes of commercial varieties in the breeding program. If genetic diversity of the crop is low, efforts to assemble resistant plants will pose a greater challenge.

This study developing molecular markers in shallots is an initial study that requires additional evaluation for further development. However, based on the results of this research (i.e., the analysis of genetic diversity in shallots in Indonesia using these developed markers), we suggest that these markers are a suitable tool to determine genetic diversity in shallots. Further, the results showed that *R* gene-derived markers were able to classify shallot genotypes into specific patterns. Genetic analysis of shallot germplasm using molecular markers will help in understanding the extent of genetic diversity and varietal identification. Markers linked to defense mechanisms will be useful in the functional analysis of these traits, and further improvement of the cultivars and recent sequencing efforts will accelerate the process.

In conclusion, *R* gene-derived markers were successfully designed and applied to analyze the genetic

diversity and population structure of 36 shallot genotypes from Java Island, Indonesia. The results of the analysis showed low genetic diversity in the shallot crops cultivated in Indonesia. Information obtained from the present study was useful as initial research on marker development using partial *R* gene sequences, and further evaluation and development in this field are required.

## ACKNOWLEDGEMENTS

This work was supported by "The Development of Molecular Markers related to Resistance Genes against *Fusarium* Disease on Shallot" project, funded by the Indonesia Agency for Agriculture Research and Development, Ministry of Agriculture of Republic Indonesia.

## REFERENCES

Aarts MGM, Hekkert BL, Holub EB, Beynon JL, Stiekema WJ, Pereira A. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant Microbe Interact 11 (4): 251-258.

Baswarsiati, Rahmawati D, and Abu. 2010. Membangun industri perbenihan di Jawa Timur. Cakrawala 5 (1): 42-52. [Indonesian]

Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M. 2009. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. Science 323: 101-106. DOI: 10.1126/science.1163732.

Belkhadir Y, Subramaniam R, Dangl JL. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr Opin Plant Biol. 7: 391-399.

Caballero A, Garcia-Dorado A. 2013. Allelic Diversity and Its Implications for the rate of adaptation. Genetics 195: 1373-1384.

Develey-Rivière MP, Galiana E. 2007. Resistance to pathogens and host developmental stage: a multifaceted relationship within the plant kingdom. New Phytol 175 (3): 405-416.

Deviana W, Meiriani, Silitonga S. 2014. Pertumbuhan dan produksi bawang merah (*Allium ascalonicum* L.) dengan pembelahan umbi bibit pada beberapa jarak tanam. Jurnal Online Agroekoteknologi 2014; 2 (3): 1113-1118. [Indonesian]

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

Earl DA, von Holdt BM. 2012. Structure harvester; a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 4: 359-361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14: 2611-2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW. and Dangl JL.1995.Structure of the *Arabidopsis* RPM1 gene enabling dual specificity disease resistance. J Sci 269 (5225): 843-846.

Grubben GJH, Denton OA. 2004. Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen.

Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. 2012. Plant disease resistance genes: Current status and future directions. Physiol Mol Plant Pathol 78: 51-65.

Heath MC. 2000. Nonhost resistance and nonspecific plant defenses. Curr Opin Plant Biol 3: 315-319.

Hiruma K, Fukunaga S, Bednarek P, Takano Y. 2013. Glutathione and tryptophan metabolites are key players in *Arabidopsis* nonhost resistance against *Colletotrichum gloeosporioidesi.* Plant Signal Behav 8 (9): e25603. DOI: 10.4161/psb.25603

Isniah US, Widodo. 2015. Exploration of nonpathogenic *Fusarium* for the control of basal rot disease on shallot. Jurnal Fitopaltologi Indonesia 11 (1): 14-22.

Jia Y, Martin GB.1999. Rapid transcript accumulation of pathogenesis-related genes during an incompatible interaction in bacterial speck disease-resistant tomato plants. Plant Mol Biol 40 (3): 455-465.

Joshi SP, Ranjekar PK, Vidya SG. 1999. Molecular markers in plant genome analysis. Curr Sci 77: 230-240.

Martanti D, Widyastuti U, Poerba SY, Megia R. 2015. Identification of Gene Candidate of Nucleotide Binding Site (NBS) from Banana *Musa acuminata* Colla var. *Malaccensis* (Riddl.) Nasution and Musa, AAA, Cavendish Sub-group. Pakistan J Biol Sci 18 (3): 99-106.

Messiaen CM, Rouamba A. 2004. *Allium cepa* L. In: Grubben GJH, Denton OA (eds.). PROTA: Plant Resources of Tropical Africa / Ressources végétales de l'Afrique tropicale. PROTA Foundation, Wageningen, Netherlands.

MoA. 2015. Kinerja Perdagangan Komoditas Pertanian Vol. 6 No. 1 Tahun 2015. Pusat Data dan Informasi Pertanian Kementerian Pertanian. Jakarta. [Indonesian]

Naktuinbouw. 2010. Naktuinbouw Calibration Book: *Allium cepa* (Cepa group), *Allium cepa* (Aggregatum group), *Allium oschaninii* O, Fedtsch and hybrids between them, Onion, Echalion, Shallot, Grey Shallot. Version 1. Naktuinbouw, Nederland.

Nei M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, USA.

NOA. 2011. History of onions. US National Onion Association, Greely, CO, US. https://www.onions-usa.org/. [6 January 2018].

Passardi F, Cosio C, Penel C, Dunand C. 2005. Peroxidases have more functions than a Swiss army knife. Plant Cell Rep 24: 255-265.

Perrier X, Flori A, Bonnet F. Data analysis methods. In: Hamon P, Sequin M, Perrier X, Glaszmann JC (eds.). Genetic Diversity of Cultivated Tropical Plants. Science Publishers, Enfield, Montpellier.

Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JPT, Hyvönen J. 2013. Advances in plant gene-targeted and functional markers: a review. Plant Methods 9 (6): 1-32. DOI: 10.1186/1746-4811-9-6.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, et al. 2000. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408 (6814): 820-822.

Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, Fartmann B, Valle G, Blocker H, Perez-Alonso M, Obermaier B, Delseny M, et.al. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408 (6814): 820-822.

Savary S, Ficke A, Aubertot JN, Hollier C. 2012. Crop loses due to diseases and their implications for global food production losses and food security. Food Secur 4: 519-537.

Sharma R, Kumar B, Arora R, Ahlawat S, Mishra AK, Tantia MS. 2016. Genetic diversity estimates point to immediate efforts for conserving the endangered Tibetan sheep of India. Meta Gene 8: 14-20.

Smith SE. 2013. "What is onion powder". WiseGeek. Conjecture Corporation. https://www.wisegeek.com/what-is-onion-powder.htm [6 January 2018].

Syukur M, Sujiprihati S, Yunianti R. 2012. Teknik Pemuliaan Tanaman. Penebar Swadaya, Jakarta. [Indonesian]

Udiarto BK, Setiawati W and Suryaningsih E. 2005. Pengenalan Hama dan Penyakit pada Tanaman Bawang Merah dan Pengendaliannya. Panduan Teknis PTT Bawang Merah No.2 tahun 2005. Balai Penelitian Tanaman Sayuran. Badan Litbang Kementerian Pertanian, Jakarta. [Indonesian]

Varshney RK. 2010. Gene-based marker system in plant: High throughput approaches for marker discovery and genotyping. In: Jain SM, Brar DS (eds.). Molecular Techniques in Crop Improvement. Springer, New York.

Yeh FC, Yang RC, Boyle T. 1999. Popgene Version 1.31 "Microsoft Windows-based freeware for population genetic analysis". www.ualberta.ca/_fyeh/popgene.pdf [16 April 2018].

Zhang Y, Yu C, Lin J, Liu J, Liu B, Wang J. 2017. *OsMPH1* regulates plant height and improves grain yield in rice. PLoS One 12 (7): e0180825. DOI: 10.1371/journal.pone.0180825.

Zhao Y, Hull AK, Gupta NR, Goss KA, Alonso J, Ecker JR. 2002. Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. Genes Dev 16: 3100-3112. DOI: 10.1101/gad.1035402.