Potential SNPs related to microspore culture in *Raphanus sativus* based on a single-marker analysis

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Plant Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CJPS-2017-0333.R3</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>28-Feb-2018</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Chung, Yong Suk  
|                         | Lee, Yun Gyeong                                      |
|                         | Silva, Renato                                         |
|                         | Park, Suhyoung                                        |
|                         | Park, Min Young                                       |
|                         | Lim, Y.                                               |
|                         | Choi, Sang Chul                                       |
|                         | Kim, Changsoo                                         |
| Keywords:            | Genotype-by-sequencing, Single marker analysis, Regeneration, Microspore, Radish |
Potential SNPs related to microspore culture in *Raphanus sativus* based on a single-marker analysis

Yong Suk Chung$^{1,2}$, Yun Gyeong Lee$^2$, Renato Rodrigues Silva$^3$, Suhyoung Park$^4$, Min Young Park$^{4,5}$, Yong Pyo Lim$^5$, Sang Chul Choi$^2$, and Changsoo Kim$^{2,*}$

$^1$Department of Plant Resources and Environment, College of Applied Life Sciences, Jeju National University, Jeju, 63243, Republic of Korea

$^2$Department of Crop Science, College of Life Sciences, Chungnam National University, Daejeon, 34134, Republic of Korea

$^3$Institute of Mathematics and Statistics, Federal University of Goiás, Goiânia, Brazil

$^4$National institute of Horticultural & Herbal Science, Rural Development Administration, Wanju, 55365, Republic of Korea

$^5$Department of Horticulture, College of Life Sciences, Chungnam National University, Daejeon, 34134, Republic of Korea

Yong Suk Chung and Yungyeong Lee have contributed equally to this work.

*Corresponding author

Email: changsookim@cnu.ac.kr
Abstract

Radish (Raphanus sativus) is an economically important crop grown for its edible roots and leaves. It is a self-incompatible, outcrossing species making the production of homozygous lines and the development of breeding populations difficult. However, this can be overcome with haploids production techniques using isolated microspores providing the rapid production of homozygous lines for breeding. Thus, it would be useful to identify radishes with a high regeneration rate from microspore culture. In the current study, 96 radish cultivars or germplasms were evaluated for high regeneration rates. Also, a single-marker analysis (SMA) was applied to detect single nucleotide polymorphisms (SNPs) potentially associated with this trait using genotype-by-sequencing (GBS) technology. The regeneration rate from microspore culture of 96 lines showed a wide range from 0 to 269.5 percent. From the SMA, fifty-two markers were detected at the $p$-value of 0.001, and a total of 11 physically nearby genes with high levels of similarity in various species were identified as candidates for high regeneration rates. This result could be used for clarifying the genetic basis underlying these traits and developing molecular markers associated with regeneration rates and would be beneficial to generating homozygous inbred lines.

Keywords genotype-by-sequencing (GBS) · regeneration rate · doubled haploid (DH) · single nucleotide polymorphism · phylogeny tree · microspore
Introduction

Radish (Raphanus sativus) belongs to the Brassicaceae family which includes many economically important plants, such as Chinese cabbage, cabbage, and oilseed rape. It is a diploid plant (2n = 2x = 18) with an estimated genome size of 530 mega base pairs (Mb) (Marie and Brown 1993). It is an economically important crop grown for its edible roots and leaves. The cultivated radish is distributed worldwide and contains large morphological and agro-ecological variations (Yamagishi and Terachi 2003).

Radish is a self-incompatible, outcrossing species (Sampson 1957). In consequence, it is difficult to produce homozygous lines to develop breeding populations. However, this can be bypassed using doubled haploids (DHs) by virtue of technologies like anther culture or isolated microspores, providing the rapid production of homozygous lines for breeding self-incompatible outcrossing lines, which are typical in Brassica vegetables (Cardoza and Stewart, 2004). Various techniques for microspore culture have been reported for Brassica species (Palmer et al. 1996). However, the success rate of Brassica microspore culture is very low and genotype-dependent (Palmer et al. 1996). Furthermore, it is generally well known that radish is one of the most difficult crops to regenerate microspore in tissue culture and this ability to regenerate plants through tissue culture using microspores is heritable (Curtis 2003; Zhang and Takahata 2001), indicating that the trait may be, to some extent, genetically regulated.

The success of producing DHs mainly depends on genotype, microspore stage, and cultural conditions of microspore in Brassica species (Cardoza and Stewart Jr 2004). So far, the technologies for making DHs have focused on culture conditions such as temperature and as well as on chemicals like colchicine for better efficiency (Charne et al. 1988; Chen and Beversdorf 1992; Lionneton et al. 2001). Homozygous DHs provide advantageous methods
for genetic studies and for breeding crops that cannot be self-pollinated, such as radish (Forster et al. 2007). Nonetheless, the double haploid approach has some drawbacks such as a high extent of segregation-distortion and the expression of lethal phenotypes, especially in cross pollinated species (Devaux et al. 1995).

In Brassicaceae, the genome sequences of Arabidopsis thaliana and Brassica rapa have been completed, and the function of many genes have been characterized (Yu et al. 2016). Following QTL analysis, synteny maps between these species offered a powerful tool for the identification of candidate genes. With the rapid development of next generation sequencing (NGS) technologies, whole-genome sequence analyses for various crops can be accomplished in a short time (Metzker 2010). The draft sequences of a Japanese radish cultivar, Aokubi, was published, comprising a total of 76,592 scaffolds covering 402 Mb, with 116 Mb of these assigned to chromosomes by incomplete genome assembly (Kitashiba et al. 2014). Using NGS, genotype-by-sequencing (GBS) technology that can simultaneously perform molecular marker discovery, and genotyping has been developed and applied to the sequencing of multiplexed samples (He et al. 2014; Kim et al. 2016; Poland and Rife 2012). GBS is a novel application of NGS protocols for discovering and genotyping single nucleotide polymorphisms (SNPs) in crop genomes and populations by constructing reduced representation libraries. SNPs found by GBS could be applied to a single-marker analysis (SMA) for target traits in non-structured populations thanks to its power to generate thousands of credible SNPs in a single run. The combination of the draft sequences of a Japanese radish cultivar genome and the SNPs associated with the target traits would increase the power to identify molecular markers and candidate genes based on previously reported information.

Here, we are interested in using GBS and SMA to find putative SNPs responsible for
the regeneration of radish plants from microspore culture. The objectives of this study were to identify genotypes with high regeneration rates from microspore culture and to develop molecular markers to enhance the efficiency to develop DHs for the establishment of inbred lines in radish. Hence, some accessions used for the current study will be used as parental lines for constructing F\textsubscript{2} mapping populations for the detailed investigation of regeneration rates.

**Materials and methods**

**Plant material**

Total 96 radish accessions were selected for microspore culture at the National Institute of Horticultural and Herbal Science, Jeonju, Korea between 2015 (Supplementary Table 1). The various regeneration rates of these resources are shown in Supplementary Table 1. Individual radish cultivars or lines were planted and vernalized in the previous year before collecting buds for microspore culture. Vernalization was done in the cold chamber at 5 °C (±1°C) for 10 weeks. And then transferred to a greenhouse maintained at 25 °C to induce the bolting for the flower bud collected.

**Microspore culture and regeneration rate**

Microspore culture was conducted according to Na et al. (2009) with some modifications. Flower buds were collected when a couple of flowers were blossoming. Among those buds, 2-3 mm of each closed bud segment were selected and wrapped in sterilized gauze. Each wrapping included 30 buds to be dipped in 1% of sodium hypochlorite (NaOCl) solution for 15 minutes. Then, the wrapping was washed three
times in sterilized water for 3 minutes. The washed buds were ground with mortar and pestle in 2-3 mL basal NLN medium (Lichter 1982; Nitsch and Nitsch 1967) supplemented with 13 % sucrose and filtrated through 45 µm metal mesh. The ground buds were then suspended in 30 ml of NLN medium. Suspension was centrifuged at 1,000 rpm for 3 minutes to separate microspores. Then, the supernatant was removed and NLN medium was added to repeat the previous step twice. The separated microspores were suspended in 75 ml of NLN medium at a density of 40,000 microspores to 1 ml. Then, microspore suspension was dispensed in four petri dishes (60 × 15 mm), followed by sealing with parafilm. One bud, which was suspended in 2.5 ml of NLN medium, was subjected to a high temperature treatment to induce embryogenesis. It was subjected to 30 °C regime for 48 hours. It was then moved to 25 °C for at least 2 weeks in darkness. After this treatment, dishes were cultured under the light, on a shaker at 70 ~ 80 rpm, at 25 °C and induced embryos were counted with bare eyes two to three days later. These counted embryos were defined as the embryos from the microspores which were collected in the early step. Regeneration rate was calculated as follows:

\[
\text{Regeneration rate} = \left( \frac{\text{the total number of induced embryos from microspore culture}}{\text{the total number of used buds in each line} \times 100,000} \right) \times 100
\]

**GBS library construction**

A GBS library was constructed based on Poland and Rife (2012) with slight modifications. The GBS libraries were constructed in 96-plex using the P384A adaptor set (Poland et al 2012). Genomic DNA was co-digested with two restriction enzymes, *Nm*<sub>SI-HF</sub> (New England Biolabs, Ipswich, MA, United States) and *MseI* (Enzynomics, Daejeon, Korea), and barcoded adapters were ligated to individual samples. Ninety six samples were pooled by plate into a single 15 ml falcon tube and cleaned using a phenol-chloroform-
isoamyl-alcohol (PCI) extraction method and re-precipitated with 2-propanol. The cleaned sample was size-selected with AMPure XP beads (Beckman Coulter, High Wycombe, UK) to clean out small fragments such as primers. The size-selected sample was then amplified by polymerase chain reaction (PCR). The PCR product was size-selected once again to become a final product in the GBS library. This library was checked for quality control using Library Quant Illumina Kit (KAPA Biosystems, MA, U.S.A.) according to manufacturer’s specification.

**Genotyping 96 entries**

Raw FASTQ files containing multiplexed reads from a barcoded library were obtained from Illumina HiSeq2000. The FASTLinkQ files were deposited in NCBI’s SRA database (SRR5804111). The radish reference genome was downloaded from NODAI Radish genome database (http://www.nodai-genome-d.org/) (Mitsui et al. 2015) and indexed with Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009). For sequence analysis and genotyping, the GBS SNP-Calling Reference Optional Pipeline (GBS-SNP-CROP, https://github.com/halelab/GBS-SNP-CROP) (Melo et al. 2016) was used. First, adaptors and low-quality sequences were trimmed using Trimmomatic v.0.33 (Bolger et al. 2014). The trimmed reads were then aligned to the reference genome with the BWA-mem algorithm (Li and Durbin 2009). Aligned sequences were indexed and sorted by SAMtools (Li et al. 2009). For generating the final SNP genotyping matrix, only high-confidence SNPs which passed stringent genotyping filtering criteria were collected. Recommended filtering options include a minimum depth for calling homozygotes (mnHoDepth)/ heterozygotes (mnHetDepth), a minimum acceptable proportion of genotyped individuals to retain a SNP (mnCall), and a maximum average depth of acceptable SNP (mxAvgDepth). The actual parameters used in this study are mnHoDepth = 5, mnHetDepth =3, mnCall =0.75 and mxAvgDepth=200.
Estimating linkage disequilibrium (LD)

For 18,071 SNPs, pairwise linkage disequilibrium ($r^2$) within sliding window size 50 was calculated using Tassel version 5.0 (Bradbury et al., 2007). The decay of Linkage disequilibrium ($r^2$) over distance in base pairs (bp) was estimated Hill and Weir expectation of $r^2$ between adjacent sites (Hill and Weir, 1988). With a low level of mutation and finite sample size $n$, the expectation becomes the following (Remington et al., 2001):

$$E(r^2) = \left[\frac{10 + C}{(2 + C)(11 + C)}\right]\left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)}\right]$$

The expected value of $r^2$ under drift-recombination equilibrium is $E(r^2) = 1/(1 + C)$, where $n$ is the effective population size, $C$ is the recombination parameter between sites, and $C = 4Nc$ (Sved, 1971). To fit equation to the data, nonlinear least squares (nls) model, in R package were used (Marroni et al., 2011).

Construction of phylogenetic tree

A maximum-likelihood phylogenetic tree based on the SNP data was constructed using SNPhylo (Lee et al. 2014). The SNPhylo filters SNP datasets with 5% minor allele frequency (MAF) and 20% missing data. An additional function includes the removal of redundant SNPs based on linkage disequilibrium (LD) information. Optionally, the SNPhylo performs multiple sequence alignment using MUSCLE (Edgar 2004). Then, a maximum-likelihood tree was constructed by DNAML embedded in the PHYLIP (Felsenstein 1989). Finally, a newick file and a phylogenetic tree image were automatically generated using default parameters.

Population analysis with STRUCTURE version 2.3.4

STRUCTURE (Pritchard et al. 2000) software version 2.3.4 was used to detect the
subpopulation structure. This program implements a model-based clustering method for inferring population structure using multi-locus genotypic data. For selecting best K, STRUCTURE HARVESTER was used (Earl 2012). It is a program for visualizing STRUCTURE output and implementing the Evanno method (Earl 2012). A STRUCTURE plot was then visualized with STRUCTURE PLOT (Ramasamy et al. 2014).

**Single marker analysis**

A marker-trait analysis (MTA) was performed by a single-marker analysis approach (Lynch and Walsh 1998; Soller et al. 1976). This method consisted of fitting a regression model and running an F test analysis of variance (ANOVA) at each marker locus with the statistical model given by:

$$
\log(y_i + 1) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i
$$

where $y_i$ is the regeneration rate of pollen culture; $\beta_0$ is the intercept of the model; $\beta_1$ is the additive effect; $x_{1i}$ is an explanatory variable that assumes -1, 0, and 1 if the marker is coded as A, H, and B, respectively; $\beta_2$ is the dominance effect; $x_{2i}$ is other explanatory variable that assumes 1 if the marker were coded as H and -1/2 otherwise; and $e_i$ is the random error.

**Locating SNPs within genes**

The location of each SNPs within the scaffold of the radish genome (http://www.nodai-genome-d.org/) was compared to the locations of those genes found by the annotation from BLAST results. Only those SNPs within the genes in the scaffolds were considered as candidate genes related to the regeneration trait. SNPRelate, parallel computing toolset(Zheng et al. 2012), was used for relatedness and principal component analysis (PCA) of SNP data using Identity-By-Descent measures.
Results and discussion

Genotyping

The sequencing of the GBS library consisting of 96 radish accessions resulted in a total 305,013,185 raw read pairs. Average length of read was 101 bp. The percentage of usable paired reads was 95.9%. A total of 39,957 SNPs was obtained with an average depth of 37.15. After filtering SNPs with MAF > 0.05 and allowing fewer than 10 missing genotypes per SNP, a total of 18,071 SNPs were used for downstream analyses.

LD decay

Average distance between markers was 152 kb (calculated by the total intervals of markers in physical distance divided by the number of SNPs) and average LD ($r^2$) was 0.708 in the samples of current study. The rate of LD decay varies depending on species because it is dependent on multiple factors such as the population size, the number of founding chromosomes in the population, and the number of generations for which the population has existed (Bush and Moore, 2012). Among those, population mating patterns and admixture can strongly influence LD. In general, LD decays more rapidly in outcrossing species like radish as compared to selfing species because recombination is less effective in selfing species, where individuals are more likely to be homozygous, than in outcrossing species (Long et al., 1998). In the current study, LD decayed rapidly within 15 kb (Fig. 1). Our results showing moderate LD decays indicate the resolution is not highly acceptable, considering the mode of reproduction is outcrossing (Rafalski, 2002). This may be caused by the limited genetic background of each accession. Thus, those SNPs detected may not be suitable to capture all the variation in our samples but the coverage of SNPs across the whole genome was still reasonable. This led us to use SMA rather than genome wide association study (GWAS).

Subpopulation structure
To infer evolutionary relationships among 96 radish samples, a maximum likelihood phylogenetic tree was constructed using SNPhylo (Lee et al., 2014). A total of 3,327 SNPs, obtained after removing redundant SNPs from initial 18,071 SNPs based on LD block information, were used to construct the tree (Fig. 2). The branch styles and colors of the tree were modified using MEGA7 software (Kumar et al. 2016).

To investigate population structure, STRUCTURE (Pritchard et al. 2000) was performed using a subset of a 3,327 filtered SNP dataset of all 96 samples. Linkage disequilibrium was assumed to be absent in the filtered SNP dataset after SNPRelated (Zheng et al. 2012) filtering in the SNPhylo (Lee et al. 2014) pipeline. The estimation of clusters (K) was performed in 20 replications using K values from 2 to 10. According to the best K value (Evanno’s methods) (Evanno et al. 2005), 8 subpopulations were defined in the 96 radish accessions used in the current study.

The result from STRUCTURE showed mixture origins among accessions which we used in this study, implying that there are not clear differences among them (Fig. 3). Cultivation of the radish has a long history (Smart and Simmonds 1995). Another theory about the origin of the radish is that it was domesticated from the Mediterranean area and spread to East Asia and Europe (Wang et al. 2008). However, none of these ideas gives definitive evidence on its origins because they are mostly based on morphological observations and have no convincing basis in genetic, cytological, or molecular data. Overall, the origin of radish has not been clarified not only due to the long history but also due to diversity of cultivated radishes (Lewis-Jones et al. 1982). However, a study showed that cultivated radishes have multiple origins (Yamagishi and Terachi 2003). During their domestication, the root of radishes was the focus for selection and they have a wide variation in shape although there also are variations in color and size of roots. Ellstrand and
Marshall (1985) reported that domesticated radish retained a population structure similar to that of wild populations based on allozyme variation within and among varieties of *R. sativa*. This is supported by Pistrič (1987) who reported that all of the taxa of wild and cultivated species might be included in *R. sativus*. Consequently, there were no significant differences in the average genetic diversity among cultivars collected from Europe, North Africa, the Middle East, and South and East Asia across all loci (Wang et al. 2008), indicating that no clear bottleneck effect occurred during segregation of these groups (Wang et al. 2008). In addition to the fact that enough seeds might have been transferred from one area to another, there maybe have not been so much breeding effect which could had reduced genetic diversity. This could be not only because enough seeds might have been transferred from one area to another during the distribution process to conserve biodiversity (Wang et al. 2008), but also because radishes is an outcrossing and open-pollination system (Brown 1978). Therefore, the regeneration-rate-oriented sampling in this study rather than a core-collection-sampling should not be an issue.

**Regeneration rate and SMA**

Regeneration rates from microspores of each individual line had a wide variation (Supplementary Table 1). As stated above, entries in the current study include enough diversity to observe a wide range of regeneration rates. Notably, the lines with regeneration rates below 10% were 63, while those above 50% were only 6 out 96 entries. These regeneration rate characteristics (low rate and genotype-dependence) are consistent with the report by Palmer et al. (1996). This indirectly reflects that the regeneration rate was not the selection target in radish cultivars or germplasms.

Out of 52 SNPs associated with regeneration rates based on SMA at *p*-value less than 0.001 (Table 1), SMA using the regeneration rate trait data found several SNPs, which are
within the genes imbedded in the scaffold, involved in the most important activities that take place during embryogenesis such as programmed cell death (SNP8594), copper ion regulation (SNP6450, SNP6451, and SNP6454), chromatin-remodeling (SNP9872), and cell cycle arrest (SNP10352 and SNP15840). Among those, two SNPs (SNP6450 and SNP6454) were very significant at \( p\)-value less than 0.0001 and on SNP (SNP6451) was very significant at \( p\)-value less than 0.00001, which were related with copper ion regulation. In addition, SNP8508 and SNP8512 (\( p\)-value 0.00000001 and 0.0000000001, respectively) would be worth to be investigated further although no genes were associated with. This will be discussed further below. Among 52 SNPs, 16 were physically close to the genes reported as candidate genes for regeneration rates in other species. Those may be linked to genes for unique pathways unlike other species in the very specific experimental condition in this study. The SMA generally provides a rough estimate of genotype-phenotype associations. Thus, it will be necessary to further investigate those 16 SNPs by analyzing quantitative trait loci (QTL) using biparental populations by crossing accessions with extremely opposite regeneration rates.

The following are the SNPs detected within genes embedded in the scaffold of the radish genome (http://www.nodai-genome-d.org/) in the current study. The SNP8594 was found within the gene encoding \textit{ring-h2} finger protein \textit{atl1-like} in the scaffold of the radish genome. \textit{ATL1} is a positive regulator of programmed cell death (Serrano et al. 2014). Programmed cell death plays an important role at the late stages of microspore-derived embryo development (Maraschin et al. 2005). Chromatin condensation and DNA degradation are accompanied by an increase in the activity of other enzymes. The orchestration of such a cell death program is linked to the elimination of the small generative domain, and further embryogenesis is carried out by the large vegetative domain.
The SNP6450, SNP6451, and SNP6454 were found within the gene encoding \textit{cox19} family protein (having Coiled Coil-Helix-Coiled Coil-Helix motifs). It has been proposed that the conserved cysteine residues of the CHCH motif in \textit{Cox19} mediate as ligands for copper ions (Nobrega et al. 2002; Smits et al. 2007). In general, copper is included in plant tissue culture media at low concentrations but has a crucial role as a cofactor of several key enzymes. There is a report that copper at 10 µM, which is higher than the concentration of copper in general medium, significantly enhanced shoot regeneration of wheat, triticale, and tobacco (Purnhauser and Gyulai 1993). The level of copper appears to be key to promoting plant regeneration but the effect of copper on plant regeneration depends on the type of in vitro culture and the cultivars tested (Szafranska et al. 2011). However, as in wheat, triticale, and tobacco, copper could be an important element for higher regeneration rate in radish judging from the association of those detected SNPs with \textit{cox19} family protein.

The SNP4477 was found within the \textit{leucine aminopeptidase 1 (lap-1)} based on the BLAST results. \textit{Leucine aminopeptidase (LAP)} is an aminoaeryl-peptide hydrolase that catalyzes the removal of amino acids from the N-terminus of a peptide or protein Matsui et al. 2006). \textit{LAPs} are known to play critical roles in many cellular ‘housekeeping’ functions, including the maturation of proteins, digestive and intracellular protein metabolism, and hormone regulation (Taylor 1993). However, very high embryo recovery rate was observed in the cultivar bearing \textit{lap-1} according to Orton and Browers (1985), implying that it may be associated with embryogenesis.

The SNP4965, SNP4669, SNP4670, SNP4671, and SNP13203 were found within the gene encoding \textit{ribonuclease h-like} protein. RNases H1/H2 participate in DNA replication (Arudchandran et al. 2000). RNases H also have been suggested to be involved in removing the RNA of RNA-DNA hybrids, particularly the primers of Okazaki fragments in lagging-
strand DNA synthesis (Kogoma and Foster 1998). This protein is known to be related to mitochondrial DNA replication during embryogenesis in mouse (Cerritelli and Crouch 2009). Furthermore, vegetative cells of pollen always have to re-enter the cell cycle before embryogenesis can occur and embryogenic cultures can be started with microspores from late G1 to G2-phase (Hause and Hause 1996). Hause and Hause (1996) showed that the dynamics of the replication with respect to microspore-derived embryogenesis was revealed based on nuclear DNA synthesis using their qualitative and quantitative analyses of nuclear DNA synthesis. The SNP9872 was found within the gene encoding a helicase-like protein. Members of large helicase-like grouping can be subdivided according to particular motif variations into two large subdivisions - superfamily I (SF-I) and superfamily II (SF-II) (Petukhova et al. 1999; Tan et al. 1999; Van Komen et al. 2000). Several members of the SF-II superfamily of helicase-like proteins have been shown to alter the twisting of DNA fragments. Most recently, this has expanded to include certain ATP-dependent chromatin-remodeling activities (Flaus and Owen-Hughes 2001), which is one of the important phase for regeneration.

The SNP10352 was found within the DNA replication licensing factor minichromosome maintenance 5 (mcm5-like, MCM5) is one of seven MCM family members, which are highly conserved in eukaryotes (Mlechkovich and Frenkel 2007). Uncovering the role of proteins in the MCM family may hold the key to understanding how DNA is replicated once per cell cycle (Kearsey et al. 1996). In yeast, mcm1 exhibits a temperature-dependent cell-cycle arrest, which is also the important phase of regeneration. The mcm5-a-like protein may have a similar function to mcm1 considering that the MCM family shares similar functions; this is worthwhile to investigate further.

The SNP15840 was found within the gene encoding pelota-like protein. The Pelota
protein homolog is a protein in humans that is encoded by the *PELO* gene (Eberhart and Wasserman 1995; Shamsadin et al. 2000). This gene encodes a protein which contains a conserved nuclear localization signal. The encoded protein may have a role in spermatogenesis, cell cycle control, and in meiotic cell division (Eberhart and Wasserman 1995). Eberhart and Wasserman (1995) found that spermatogenesis in pelota and twine homozygotes in Drosophila is normal until the late meiotic prophase, at which point the cell cycle arrests due to a lack of spindle formation (Eberhart and Wasserman 1995). Cell cycle arrest was observed in tobacco and rapeseed microspores (Binarova et al. 1993; Žárský et al. 1992) and this will be discussed further later.

Numerous studies lead to the prediction of the mechanism of microspore embryogenesis at the molecular level using functional genomics such as microarray-based transcriptomics for gene discovery in microspore embryogenesis, differential screens, subtractive hybridization to cDNA microarrays, suppression subtractive hybridization and metabolomics (Hosp et al. 2007). Nevertheless, the overall picture of the complex regulation of the developmental switch from the gametophytic to sporophytic pathway during microspore embryogenesis has not yet emerged even with the large amount of data obtained from the different model species. However, there are some clues, such as cell cycle arrest and chromosome re-modelling, suggesting that some common aspects govern the re-programming of microspores toward becoming embryos (Hosp et al. 2007). Cell cycle arrest may lead to an early stage of an autophagic response that enables the fittest microspores to de-differentiate and develop as embryos according to evidence in yeast (Wysocki and Kron 2004). Chromosome re-modelling might be a further requirement in an early stage of adaptation based on reports of a number of genes involved in chromosome re-organization in diverse species upon differential screens (Hosp et al. 2007). Reconstruction of the
cytoskeleton might be involved in the morphological changes from young pollen to embryogenic cells along with chromosomal alterations. Changes in the transcriptional and translational machinery are likely to occur. Afterwards, it is assumed that proteins are to be recycled and selectively destroyed in favor of synthesis of new proteins that are able to better serve the novel conditions of embryogenic induction (Aubert et al. 1996). During the transition from multi-cellular structures to embryos, sufficient evidence for the importance of programmed cell death at late stages of microspore embryo development was found (Maraschin et al. 2005).

A master gene capable of initiating embryogenic development from microspores has not been identified so far. It would be unlikely that a single gene could induce a large number of profound cellular adaptations and changes. Rather, initiating embryogenic development from microspores is more likely to have a concerted series of events taking place out of multiple genes during haploid embryo development from microspore. Each event may be triggered by switches that initiate stress-induced metabolic alterations leading to higher order changes in chromatin structure and gene expression, and end with the activation of cell-cycle and regulatory genes (Hosp et al. 2007). Notably, none of the same genes including those genes associated with the markers found in the current study has been identified in other species. This could be due to the complexity of the process of microspore embryogenesis. Indeed, embryo-abundant gene expression is spatially and temporally specific (Reynolds and Kitto 1992). It is also possibly because of the different stages of microspore embryogenesis and the limited number of genes analyzed. Different species might have different pathways to regulate genes for embryogenesis. In addition, the differences in the culturing conditions and the variations from different buds, even within a single plant, created diverse stress guiding different pathways for signaling to express different genes. These speculations could be
supported by circumstantial evidence like the following studies. Lane (1978) reported bud location effects on regeneration frequency whereas Chuong et al. (1988) showed there is no bud position effect. Even the effect of sampled-bud location is not consistent based on the independent research using different species. Furthermore, the position and orientation of the explant affect the frequency of shoot regeneration and the number of adventitious buds produced from the regenerated shoots (Zhao et al. 2007). In *R. sativus*, temperature treatment and bud size were found to have an effect on plant regeneration (Takahata et al. 1996). In addition, there might be genotype by environmental effect on the results for the microspore regeneration rate. All these results indicate the complexity of different gene expression regulations among different species within plants and among subtle differences of the incubation condition even within the same plant. This might have led this inconsistent unity of those genes in the independent studies for embryogenesis. Hence, it would be worth to establish a sampling method and standard conditions for culture to uncover the mechanisms underpinning embryogenesis.

The regeneration rates in 96 samples of the current study were low and genotype-dependent as stated. Here, we found several SNPs which are potentially associated with genes for the important stages during the regeneration from microspores, including programmed cell death, copper ion regulation, chromatin-remodeling, and cell cycle arrest. Our findings could be used to develop molecular markers for regeneration rate. In addition, those unknown SNPs might lead to new findings for clarifying genes underlying the traits and for developing molecular markers associated with regeneration rates beneficial to generating homozygous inbred lines. To achieve this, we are planning to make F2s using two parents among the 96 entries presented in this study, one with a low regeneration rate and the other with a high one. This will enable us to find useful markers to screen for candidate lines at
early growing stages for making DHs in the breeding of *R. sativa*.

**Acknowledgment**

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Golden Seed Project, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (213006-05-1-WTH11).

**References**


Aubert S et al. (1996) Ultrastructural and biochemical characterization of autophagy in higher plant cells subjected to carbon deprivation: control by the supply of mitochondria with respiratory substrates. *J Cell Biol* 133:1251-1263. doi: 10.1083/jcb.133.6.1251


male and female recombination-derived, doubled haploid populations. MGG, 249:600-608.


ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5. doi: 10.3389/fpls.2014.00484


https://mc.manuscriptcentral.com/cjps-pubs


https://mc.manuscriptcentral.com/cjps-pubs


https://mc.manuscriptcentral.com/cjps-pubs


Figure captions

Fig. 1. Plot of linkage disequilibrium (LD, r^2) against distance between SNPs in Radish genome. Red line show the expected LD decay, estimated by nonlinear regression of r^2.

Fig. 2. Phylogeny tree of 96 radish entries using SNPs from genotype-by-sequencing. Each color (orange, yellow, green, blue, purple, violet, magenta, red, and peach) corresponds to the same color of STRUCTURE result in Figure 3.

Fig. 3. STRUCTURE result of 96 radish entries using SNPs from genotype-by-sequencing. Each colour (orange, yellow, green, blue, purple, violet, magenta, red, and peach) corresponds to the same color of phylogeny tree in Figure 2.
Table 1 SNPs for microspore regeneration rate and their associated candidate genes using single marker analysis in 96 radish entries

<table>
<thead>
<tr>
<th>Scaffold number</th>
<th>SNP</th>
<th>P-value</th>
<th>Sequence Name</th>
<th>Sequence Description</th>
<th>Blast Top Hit Gene Name</th>
<th>Blast Top Hit E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs_scaf1279</td>
<td>SNP2031</td>
<td>3</td>
<td></td>
<td></td>
<td>ARALYDRAFT_48127</td>
<td>0</td>
</tr>
<tr>
<td>Rs_scaf1283</td>
<td>SNP2082</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf14</td>
<td>SNP2725</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf14282</td>
<td>SNP2974</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf14282</td>
<td>SNP2975</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf1498</td>
<td>SNP3348</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf16382</td>
<td>SNP4095</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf1707</td>
<td>SNP4477</td>
<td>3</td>
<td>RSG46506.11</td>
<td>leucine aminopeptidase 1</td>
<td>ARALYDRAFT_48127</td>
<td>0</td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4957</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4958</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4965</td>
<td>4</td>
<td>RSG14716.11</td>
<td>ribonuclease h-like protein</td>
<td>AT4G29090</td>
<td>7.99E-31</td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4969</td>
<td>4</td>
<td>RSG14716.11</td>
<td>ribonuclease h-like protein</td>
<td>AT4G29090</td>
<td>7.99E-31</td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4970</td>
<td>4</td>
<td>RSG14716.11</td>
<td>ribonuclease h-like protein</td>
<td>AT4G29090</td>
<td>7.99E-31</td>
</tr>
<tr>
<td>Rs_scaf182</td>
<td>SNP4971</td>
<td>4</td>
<td>RSG14716.11</td>
<td>ribonuclease h-like protein</td>
<td>AT4G29090</td>
<td>7.99E-31</td>
</tr>
<tr>
<td>Rs_scaf184</td>
<td>SNP5045</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf21962</td>
<td>SNP6450</td>
<td>4</td>
<td>RSG58882.11</td>
<td>cox19 family protein (chch motif)</td>
<td>AT4G29090</td>
<td>3.16E-06</td>
</tr>
<tr>
<td>Rs_scaf21962</td>
<td>SNP6451</td>
<td>5</td>
<td>RSG58882.11</td>
<td>cox19 family protein (chch motif)</td>
<td>AT4G29090</td>
<td>3.16E-06</td>
</tr>
<tr>
<td>Rs_scaf21962</td>
<td>SNP6454</td>
<td>5</td>
<td>RSG58882.11</td>
<td>cox19 family protein (chch motif)</td>
<td>AT4G29090</td>
<td>3.16E-06</td>
</tr>
<tr>
<td>Rs_scaf249</td>
<td>SNP7569</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf249</td>
<td>SNP7570</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf2540</td>
<td>SNP7773</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf2547</td>
<td>SNP7791</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf25616</td>
<td>SNP7853</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf2799</td>
<td>SNP8508</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>SNP</td>
<td>Gnome Position</td>
<td>Gene Function</td>
<td>Accession Number</td>
<td>P-Value</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf2799</td>
<td>SNP8512</td>
<td>9</td>
<td>ring-h2 finger protein atl1-like</td>
<td>AT5G43420</td>
<td>1.89E-132</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf281</td>
<td>SNP8594</td>
<td>3</td>
<td>receptor-like protein kinase hsl1-like</td>
<td>ARALYDRAFT_48944</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf297</td>
<td>SNP9014</td>
<td>3</td>
<td>receptor-like protein kinase hsl1-like</td>
<td>ARALYDRAFT_48944</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf3</td>
<td>SNP9108</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf317</td>
<td>SNP9660</td>
<td>3</td>
<td>helicase-like protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf344</td>
<td>SNP10352</td>
<td>3</td>
<td>dna replication licensing factor mcm5-a-like</td>
<td>ARALYDRAFT_48024</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf4221</td>
<td>SNP11967</td>
<td>8</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf511</td>
<td>SNP13203</td>
<td>3</td>
<td>ribonuclease h-like protein</td>
<td>AT1G10000</td>
<td>6.10E-39</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf560</td>
<td>SNP13809</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf6236</td>
<td>SNP14579</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf703</td>
<td>SNP15451</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf703</td>
<td>SNP15452</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf7372</td>
<td>SNP15761</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf7372</td>
<td>SNP15763</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf749</td>
<td>SNP15840</td>
<td>3</td>
<td>protein pelota-like</td>
<td>AT3G58390</td>
<td>5.02E-148</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf8219</td>
<td>SNP16582</td>
<td>3</td>
<td>protein hsl1-like</td>
<td>AT1G52950</td>
<td>4.85E-46</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf8219</td>
<td>SNP16583</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf866</td>
<td>SNP17016</td>
<td>3</td>
<td>protein hsl1-like</td>
<td>AT1G52950</td>
<td>4.85E-46</td>
<td></td>
</tr>
<tr>
<td>Rs_scaf889</td>
<td>SNP17235</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf889</td>
<td>SNP17236</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rs_scaf889</td>
<td>SNP17237</td>
<td>3</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2.

233x238mm (96 x 96 DPI)
Fig. 3.

301x65mm (96 x 96 DPI)