**SNP diversity within and among *Brassica rapa* accessions
reveals no geographic differentiation**

Journal:	<i>Genome</i>
Manuscript ID	gen-2015-0118.R1
Manuscript Type:	Article
Date Submitted by the Author:	20-Oct-2015
Complete List of Authors:	Tanhuanpää, Pirjo; Natural Resources Institute Finland (Luke), Green Technology Erkkilä, Maria; Natural Resources Institute Finland (Luke), Internal Expert Services Tenhola-Roininen, Teija; Natural Resources Institute Finland (Luke), Green Technology Tanskanen, Jaakko; Natural Resources Institute Finland (Luke), Green Technology Manninen, Outi; Boreal Plant Breeding Ltd.
Keyword:	<i>Brassica rapa</i>, SNP, genetic diversity, GBS, STRUCTURE

SCHOLARONE™
Manuscripts

Original article

SNP diversity within and among *Brassica rapa* accessions reveals no geographic differentiation

P. Tanhuanpää^{1*}, M. Erkkilä², T. Tenhola-Roininen¹, J. Tanskanen¹ & O. Manninen³

¹Green Technology, Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland

²Internal Expert Services, Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland

³Boreal Plant Breeding Ltd., FI-31600 Jokioinen, Finland

With 3 figures (including one as an ESM) and 3 tables (including one as an ESM)

Draft

Correspondence to:

Pirjo Tanhuanpää

Green Technology

Natural Resources Institute Finland (Luke)

FI-31600 Jokioinen

Finland

tel: +358 295326574

e-mail: pirjo.tanhuanpaa@luke.fi

Abstract

Genetic diversity was studied in a collection of 61 *Brassica rapa* accessions, which were mostly oiltype turnip rapes but also 2 oiltype ssp. *dichotoma* and 5 ssp. *trilocularis* accessions, as well as 3 leaf type (ssp. *japonica*, *pekinensis*, and *chinensis*) and 5 turnip cultivars (ssp. *rapa*) were included. Two-hundred and nine SNP markers, which had been discovered by amplicon resequencing, were used to genotype 893 plants from the *B.rapa* collection using Illumina BeadXpress. There was great variation in the diversity indices between accessions. With STRUCTURE analysis, the plant collection could be divided into 3 groups which seemed to correspond to morphotype and flowering habit but not to geography. According to AMOVA analysis, 65% of the variation was due to variation within accessions, 25% among accessions, and 10% among groups. A smaller subset of the plant collection, 12 accessions was also studied with 5727 GBS-SNPs. Diversity indices obtained with GBS-SNPs correlated well with those obtained with Illumina BeadXpress SNPs. The developed SNP markers have already been used and will be used in future plant breeding programs as well as in mapping and diversity studies.

Keywords: *Brassica rapa*, SNP, genetic diversity, GBS, STRUCTURE

Introduction

Brassica rapa L. ($2n=20$, AA, haploid genome size 485 Mb) is an obligate out-crossing crucifer grown worldwide. The species is morphologically very variable, and contains a large and diverse group of vegetable, oil, fodder, and condiment crops. Whether there are one or two centres of origin of *B.rapa* has been debated but according to a recent study based on Simple Sequence Repeat (SSR) analysis (Guo et al. 2014), there is one centre of origin in the classical Old World but several centres of diversity in Asia. The oleiferous form of *B.rapa*, namely ssp. *oleifera*, turnip rape, is the third important *Brassica* oilseed crop, and is grown in China, Canada, India, and northern Europe (Ramchiary and Lim 2011). In Finland, spring turnip rape is the most important oilseed crop. Other oiltype subspecies include *B.rapa* ssp. *trilocularis* (yellow sarson), *B.rapa* ssp. *sarson* (brown sarson), and *B.rapa* ssp. *dichotoma* (toria), which are grown in Indian continent (Kimber and McGregor 1995).

The assessment of diversity in a species is important in plant breeding programs, and for effective conservation, management and utilization of genetic resources of the species. The most efficient markers for studying diversity are molecular markers, and several marker systems exist. Single nucleotide polymorphism (SNP) markers describe a single base pair change or indel (insertion or deletion), and occur frequently throughout the genome in large numbers. Publicly available SNP panels exist only for a limited number of species, and a great effort is needed when designing SNP markers for a species of your choice. SNPs can be found by comparing sequences from several genotypes obtained either from sequence databases or from sequencing projects. Fortunately, powerful next-generation sequencing (NGS) techniques enable large-scale resequencing studies for discovering novel SNPs. Sequence assembly is greatly simplified if a reference genome exists, and for *B.rapa* (*Brassica* A genome), a reference genome was released in 2011 (Wang et al. 2011). Recently, genotyping by sequencing (GBS) has become an option for genetic diversity studies in plant populations. With the ability of multiplexing as many as 384 barcoded samples per sequencing lane in NGS, unbiased SNP polymorphism detection has become economical. However, analyzing such data is more demanding due to the dimensions of data as well as the big amount of missing values.

Genetic diversity in *B.rapa* has been studied with different marker types such as allozymes, isozymes, Restriction Fragment Length Polymorphisms (RFLPs), Amplified Fragment Length Polymorphisms (AFLPs), and SSRs (McGrath and Quiros 1992; Zhao et al. 2005; Takuno et al. 2007; Zhao et al. 2009; Pino Del Carpio et al. 2011; Annisa et al. 2013; Guo et al. 2014). SNP markers have only been used to study genetic variation in turnips (Zhang et al. 2014) but not in other morphotypes.

The specific aims of the present study were 1) to gather a genetically diverse collection of *B.rapa*, mainly oiltypes, from different parts of the world, 2) to develop a set of SNP markers identified by amplicon resequencing and with some SNPs designed to genes

affecting important breeding traits (fatty acid and glucosinolate metabolism, and disease resistance), 3) to analyze the SNP diversity in the *B.rapa* collection, and 4) to compare our SNP genotyping data with GBS data.

Materials and methods

Plant material

Sixty-one *B.rapa* accessions, mostly oiltype, were selected from genebanks and commercial sources (Table 1). The collection contained local accessions, cultivars, and breeding lines. The turnip rape (*B.rapa* ssp. *oleifera*) accessions included cultivars from Canada (4), Sweden (4) and Finland (1), breeding lines from Finland (6), accessions from South and Middle America (4), North America (1), Australia (1), Africa (1), Europe (8), and Asia (14). Five of the accessions were wild type (var. *silvestris*, 3 from Europe and 2 from Asia). Because there are several centres of diversity of *B.rapa* in Asia, several Asiatic accessions were included. In addition to turnip rape, 2 oiltype ssp. *dichotoma* and 5 ssp. *trilocularis* accessions were included, as well as 3 leaf type (ssp. *japonica*, *pekinensis*, and *chinensis*) and 5 turnip cultivars (ssp. *rapa*).

Four plants from those accessions that were presumably self-compatible (ssp. *dichotoma* and *trilocularis*, Gómez-Campo and Prakash 1999), and 16-20 plants from the other accessions were sown in greenhouse. DNAs were extracted using the method of Tinker et al. (1993) with modifications described in Tanhuanpää and Manninen (2012). DNA concentrations were measured with NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Delaware, USA). All seeds did not germinate and in addition, some individuals had to be excluded from the final data because of more than 10% missing SNP results (Table 1 contains the final amounts of individuals) to end up with a total of 893 individuals. For resequencing, a panel of 8 individuals was selected, one individual from each of the following accessions: Cordelia (spring turnip rape from Finland), SW Petita (spring turnip rape from Sweden), Salut (winter turnip rape from Sweden), SW Rebus (spring turnip rape from Sweden), Milan White (turnip), PI 459018 (*B.rapa* ssp. *trilocularis*), CR 3396 (turnip rape from Tibet), and K-145 (turnip rape from China). The selected accessions represented different types of *B.rapa* and oiltypes that interested the breeder.

Amplicons

Amplicon sequencing was used for SNP discovery. Genes for amplicons were selected by taking advantage of the recently sequenced *B.rapa* ssp. *pekinensis* accession Chiifu-401-42 (Wang et al. 2011). A list of all the annotated genes was made, and 149 genes were selected based on their function: those involved in fatty acid or glucosinolate metabolism, or defence. In

addition, 235 genes situating evenly in the genome (every 100th gene) were selected to end up with a total of 384 genes.

Primers for amplifying amplicons from the selected genes were designed using the program eprimer3 (Rice et al. 2000). From the 5 alternative primer pairs for each amplicon, we selected primers that attached to exons so that there was at least one intron between them, the GC percentage was as high as possible, and the amplicon size was around 400 bp.

PCR program for amplicon amplification consisted of 35 cycles of 10 s at 98°C, 30 s at 58°C (for some amplicons 54°C or 62°C), and 30 s at 72°C in a PTC-220 DNA Engine Dyad™ Peltier Thermal Cycler (MJ Research, Waltham, Massachusetts, USA). The program started with an initial denaturation step of 30 s at 98°C and was followed by a final extension step of 10 min at 72°C. PCR amplification reactions in 10 µl contained 0.2 U of Phusion® Hot Start II DNA polymerase (Fisher Scientific, Vantaa, Finland), the buffer HF with 1.5 mM MgCl₂ supplied by the enzyme manufacturer, 200 µM each dNTP, 10 ng of DNA, and 0.25 µM each primer. PCR products were run on a 1% agarose gel (Bioline, London, UK) to verify the amplification of only one band.

First, the functionality of the primers was verified by amplifying all amplicons in one of the 8 genotypes to be resequenced (one individual from the following accessions: Cordelia, SW Petita, Salut, SW Rebus, Milan White, PI 459018, CR 3396, and K-145). Primer pairs that did not produce any band or produced more than one band, were optimized using a lower or a higher annealing temperature, respectively. Three-hundred and fifteen (= 82%) amplicons were amplified correctly either in the first trial or after optimization, and thereafter, they were amplified in all the 8 genotypes. PCR products were normalized to a concentration of 1.25 ng/µl of DNA with the aid of a SequalPrep™ normalization plate kit (Invitrogen, Carlsbad, CA, USA). Some random checks on an agarose gel were done after normalization to verify the functioning of PCR.

Resequencing

All amplicons from one genotype were combined and concentrated with a GFX PCR DNA and gel band purification kit (GE Healthcare, Buckinghamshire, UK) to a theoretical concentration of 500 ng DNA in 16 µl. Sequencing was done with a Roche 454 GS Junior (Roche Applied Science, Penzberg, Germany) using the ligated adaptors amplicon sequencing design. With the aid of Multiplex identifiers (MIDs), all the amplicons could be finally combined and sequenced in only one run.

SNP design and genotyping

SNPs or indels were visually looked for with the aid of the program Tablet 1.12.02.06 (Milne et al. 2010). It was not possible to design SNP markers for all amplicons because sometimes the

amplicons were too short, contained too few reads, did not contain any polymorphism or contained too many polymorphisms in the SNP design region. In total, 332 SNP markers were designed. To increase the amount, two public databases were used. From the ACPFG database (<http://autosnpdb.appliedbioinformatics.com.au/>) we selected 28 genes either having an interesting function or sequence homology to a candidate gene known previously. Thirty-two sequences with clear SNP polymorphism between individual *B.rapa* accessions were used for SNP marker design. In addition, 20 SNPs were selected from the Quiros lab *Brassica* SNP discovery page <http://informatics.plantsciences.ucdavis.edu/haggard/brassica.html>. These were in sequences that found only one good hit in Chiifu sequence when using BLAST on the BrassEnsembl net page (<http://www.brassica.info/Multi/blastview>).

Three hundred ninety-three SNP-containing sequences (a SNP flanked by ~60 bp on each side) were submitted to Illumina for evaluation of marker designability, and 384 best ones were submitted for Oligo Pool Assay synthesis. SNP genotyping was carried out using Illumina BeadXpress (San Diego, CA, USA). With GoldenGate assay and Veracode technology it was possible to genotype 96 samples with 384 SNP markers on a single plate. For genotyping the data, the GenomeStudio genotyping software with the GenCall score threshold of 0.30 was used.

GBS

Twelve *B.rapa* accessions (Table 1) from the plant collection, represented by 4-8 individuals per accession were chosen for GBS experiment (84 individuals in total). GBS libraries were prepared and analyzed at the Institute for Genomic Diversity (IGD, Cornell, USA), according to Elshire et al. (2011), using the enzyme *PstI* for digestion and creating a library with 96 unique barcodes. The GBS analysis pipeline [Tassel Version 3.0.157], an extension to the Java program TASSEL (Bradbury et al. 2007), was used to call SNPs from the sequenced GBS library with the following options: <10% missingness and a minor allele frequency (MAF) >2%. These thresholds diminished the total number of SNPs from 9161 to 5727 (63%). Tags were aligned to the *B.rapa* reference genome downloaded from ftp://ftp.jgi-psf.org/pub/compugen/phytozome/v9.0/Brapa/assembly/Brapa_197.fa.gz on 31st July 2013 and indexed for use with bwa version 0.7.3a-r367.

Data analyses

The program GenAlex 6.5 (Peakall and Smouse 2006, 2012) was used for counting genetic diversity indices for accessions: expected heterozygosity (H_e), observed heterozygosity (H_o), Shannon's diversity index I (Shannon 1948), and the percentage of polymorphic loci. The diversity indices for SNPs: polymorphic information content (PIC), gene diversity, heterozygosity, and major allele frequency, were counted with the program Powermarker v3.0

(Liu and Muse 2005). Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was conducted to partition genetic variance within and among accessions, and among groups, and was performed with GenAlex 6.5 (Peakall and Smouse 2006, 2012). Significance of the results was tested by permuting the SNP marker data 999 times. Four accessions that according to SNP analysis seemed to be self-compatible were omitted from AMOVA analysis. A neighbor-joining (NJ, Saitou and Nei 1987) dendrogram containing all the individuals in the study was built with MEGA ver 6 (Tamura et al. 2007) using the Euclidean distances counted with GenAlex 6.5. In addition, NJ trees were built for the 84 individuals that were involved in the GBS experiment, using Euclidean distances based on either GBS-SNPs or Illumina BeadXpress SNPs, and distance matrices were compared with a Mantel test (Mantel 1967) in the software GenAlex.

The population structure of *Brassica* accessions was investigated by STRUCTURE ver. 2.3.2.1 software using admixture model (Pritchard et al. 2000; Falush et al. 2003; Hubisz et al. 2009). The number of clusters or populations (K) was estimated by using five runs of 20,000 iterations following 10,000 burn-in iterations for values of K between 1 and 10. The best number of clusters (3) was identified by using the delta K method which was based on the estimated probability of data $\ln P(D)$ (Evanno et al. 2005). The STRUCTURE program counts membership fractions for the K clusters for each individual. For each accession, the average membership fraction was counted, and accessions were assigned to groups according to this mean membership value, into that group the value of which was the highest.

Results

SNP discovery

From the 384 SNP markers (including 28 indels) run with Illumina BeadXpress, 209 (54.4%) functioned well, were polymorphic (ESM Table S1), and were used to genotype plants from our *B.rapa* collection. Six of the SNPs were actually indel differences. From the nucleotide substitutions, 54% were transitions. The number of SNP markers in different chromosomes (A01-A10) was as follows (number in parenthesis refers to the amount of different loci): 28 (24), 14 (12), 26 (21), 19 (15), 21 (15), 20 (19), 26 (24), 10 (10), 30 (24), 15 (13).

In the final set of 209 SNP markers, the PIC value varied from 0.039 to 0.482 (mean 0.284 ± 0.104), heterozygosity from 0 to 0.572 (0.207 ± 0.108), gene diversity from 0.039 to 0.570 (0.345 ± 0.141), and major allele frequency from 0.496 to 0.980 (0.748 ± 0.139) (Table S1).

Genetic diversity

There was great variation in diversity indices of the 61 accessions (Table 1). Four of the accessions of ssp. *dichotoma* and *trilocularis* (CR 1643, PI 459015, PI 459018, and PI 649205) showed very low values for diversity indices, e.g. percentage of polymorphic loci from 0 to

5.3%, and I from 0 to 0.027, where as in 3 accessions (CR 2301, PI 173847 and PI 649203), the percentage of polymorphic loci varied from 31.6% to 69.9%, and I from 0.191 to 0.390. The naming of ssp. *dichotoma* and *trilocularis* is very vague, and they have been mentioned to be self-compatible (Gómez-Campo and Prakash 1999). However, our results indicated that there are both breeding types among these subspecies, accessions CR 1643, PI 459015, PI 459018, and PI 649205 being presumably self-compatible. In the 57 self-incompatible accessions from our plant collection, the percentage of polymorphic loci varied from 31.6% (PI 173847, ssp. *trilocularis* accession from India) to 91% (K-229, a Korean accession), with mean $66.6 \pm 20.9\%$, H_e from 0.133 (PI 173847) to 0.318 (K-31 from China) with mean 0.226 ± 0.069 , H_o from 0.061 (Bor01, a self-pollinated breeding line from Finland) to 0.326 (PI633159 from Korea) with mean 0.203 ± 0.087 , and I from 0.191 (PI 173847) to 0.475 (K-31) with mean 0.339 ± 0.104 . In most accessions, H_o was approximately the same as H_e , indicating that the accessions were under Hardy-Weinberg equilibrium. Exceptions included breeding material that had been selfed a couple of times (Table 1).

Genetic distances between individuals were visualized with a NJ tree (ESM Fig. S1). Individuals belonging to the same accession mainly grouped together. However, especially spring oilseed accessions did not always form clear groups but individuals were somewhat mixed. There were also some individuals with erroneous positioning to a wrong accession type (two examples have been pointed out with arrows in Fig. S1), due to perhaps a sample mixup. The presumably self-compatible accessions of ssp. *dichotoma* and *trilocularis* (CR 1643, PI 459015, PI 459018, and PI 649205) formed a clearly separate branch, as also accession K-145 together with ssp. *trilocularis* accession PI 173847. The turnips located near to each other in one branch. Each of the 3 leaf type accessions formed a branch of its own but they were located near to each other. Winter type accessions Largo, Salut, and K-193 belonged to the same branch, where as 2 other winter type accessions K-238 and K-239 belonged to a different branch. The spring oilseed breeding material seemed to form three different groups. Each var. *silvestris* accession formed a clear group, and they were located either near the leaf types or near the turnips.

STRUCTURE deduces population structure using the marker data to identify K clusters. The delta K values for 2 or 3 clusters were almost the same. Based on the K value 3 the material was divided into 3 groups (Fig. 1). The assignment of accessions into groups was usually very straightforward but for some accessions, the membership values for 2 different groups (mainly groups 1 and 2) were close to each other and these accessions could be considered admixed. The 3 groups were: 1) leaf type, ssp. *dichotoma* and *trilocularis* accessions, and K-145 (from China), 2) turnips, winter type accessions, var. *silvestris* accessions, PI 633159 (from Korea), and CR 3396 (from Tibet), and 3) ssp. *oleifera* spring types. In the 3 groups (group 1: 73 individuals when the 4 presumably self-compatible

accessions were omitted, group 2: 262 individuals, and group 3: 543 individuals), the diversity indices were as follows: the amount of polymorphic loci $54.5 \pm 14.5\%$, $63.8 \pm 9.5\%$, and $78.4 \pm 7.5\%$, I 0.293 ± 0.072 , 0.330 ± 0.049 , and 0.393 ± 0.039 , H_e 0.197 ± 0.048 , 0.221 ± 0.032 , and 0.261 ± 0.026 , and H_o 0.205 ± 0.044 , 0.223 ± 0.054 , and 0.216 ± 0.085 , respectively. According to AMOVA analysis of self-incompatible accessions, 65% of the variation was due to variation within accessions, 25% among accessions, and 10% among STRUCTURE groups (Table 2).

GBS-SNPs versus Illumina BeadXpress SNPs

From the 12 accessions selected for GBS analysis, a total of 5727 SNPs were called. NJ trees built either using distance matrices obtained with GBS-SNPs or 209 Illumina BeadXpress SNPs grouped the individuals from the 12 accessions roughly similarly (Fig.2). In both trees, individuals from accessions Debra, Siloganova, Largo, PI 649203 (ssp. *trilocularis*), and Bor01 were grouped correctly to their corresponding accessions without any exceptions. Instead, individuals from e.g. cultivars SW Petita and Cordelia were somewhat mixed with each other and with the Bor breeding lines in both trees. Individuals from accessions PI 173847 (ssp. *trilocularis*) and K-145 grouped together in the same branch. The positioning of accessions in the trees was also quite similar. However, some differences existed, e.g. cv. Largo was more closely grouped with the turnips Debra and Siloganova in the tree built with Illumina BeadXpress SNPs. In addition, regarding the 3 spring oilseed breeding lines (Bor01, Bor03 and Bor05), Bor01 and Bor03 grouped close to each other in the tree based on GBS-SNPs, whereas Bor03 and Bor 05 in the tree based on Illumina BeadXpress SNPs.

Diversity indices in the 12 accessions studied with GBS-SNPs were smaller when compared to those with Illumina BeadXpress SNPs. The percentage of polymorphic loci studied with GBS-SNPs varied (results not shown) from 31% to 46% (mean $39.7 \pm 1.4\%$), I from 0.176 to 0.251 (mean 0.210 ± 0.020), H_e from 0.117 to 0.173 (mean 0.140 ± 0.013), and H_o from 0.038 to 0.213 (mean 0.104 ± 0.048). The corresponding numbers with Illumina BeadXpress SNPs were as follows: the percentage of polymorphic loci from 32% to 69% (mean $58.6 \pm 3.3\%$), I from 0.191 to 0.367 (mean 0.316 ± 0.050), H_e from 0.133 to 0.246 (mean 0.212 ± 0.032), and H_o from 0.060 to 0.267 (mean 0.161 ± 0.061). The values correlated well with each other ($r=0.744$, 0.770 , and 0.974 for I , H_e , and H_o , respectively) when two accessions (K-145 and accession PI 173847) were omitted from the results due to many missing observations. The distance matrices also correlated well with each other ($r=0.897$, $P<0.0001$).

Discussion

SNPs are the most abundant type of DNA sequence variation and SNP markers have been developed for several species. For *Brassica*, the first public Infinium SNP array (60K, A and C genomes) was developed in 2011 based on the work of the international *Brassica*

SNP consortium, which was established together with Illumina Inc. (San Diego, CA, USA). Our project began before the array was available, and we wanted to develop a collection of *B.rapa* SNP markers. Our aim was made possible with the release of the first *B.rapa* sequence from a leaf type cultivar 'Chiifu' (Wang et al. 2011). Thereafter, genomes from two other morphotypes, turnip and oiltype, have been sequenced (Lin et al. 2014). Nowadays, several studies have been published which describe the use of Infinium SNP array (Liu et al. 2013; Raman et al. 2014), or the development and use of novel SNP markers - mainly for linkage mapping - in *Brassica A* genome (Li et al. 2009; Park et al. 2010; Delourme et al. 2013; Paritosh et al. 2013; Devisetty et al. 2014; Chung et al. 2014; Zhang et al. 2014; Song et al. 2015). However, genetic diversity in oilseed *B.rapa* has not been studied with SNPs.

Our plant material was analyzed with 209 SNP markers and was divided into 3 diverse STRUCTURE groups, which were quite definite. The groups seemed to correspond at least partially to morphotype and flowering habit but not to geographic origin. Group 1 contained all leaf types and ssp. *dichotoma* and *trilocularis* accessions, group 2 contained all winter types (including turnips), and group 3 contained most spring oiltypes from all over the world. Five wild accessions (var. *silvestris*, CR 2210, CR 2211, CR 2212, CR 2213, and CR 2616) belonged to group 2. These accessions have been classified as spring types according to genebank but CR 2210, CR 2211, and CR 2616 have been noticed to be semi-winter types (CR 2211 is also a root type) and belong to the same SSR group (Annisa et al. 2013), and Guo et al. (2014) demonstrated that CR 2212 is a winter type thus justifying the assignment of these 4 wild types to group 2 with other biannuals. Unfortunately, we did not grow plants long enough to verify their flowering habit. Also in Guo et al. (2014) all the 7 var. *silvestris* accessions included in the study clustered to the same group. Diversity in our 3 STRUCTURE groups was different, group 1 being the least variable and group 3 the most variable. However, this might merely reflect the fact that our genotype collection was deliberately biased towards spring oiltypes, and therefore group 3 was biggest and contained accessions widely from all over the world.

There were only a few exceptions to the STRUCTURE grouping corresponding to morphotype and flowering habit. This might indicate incomplete information or misclassification in the genebank databases or sample mixup in greenhouse or during DNA extractions or analyses. As a matter of fact, accession K-145, which belongs to group 1 (*dichotoma-trilocularis*-leaf types), and not to group 3 as expected, appears to be ssp. *trilocularis* because in the NJ tree its individuals very clearly belonged to the same branch with those of ssp. *trilocularis* accession PI 173847. This was not due to the quite a high amount of missing data in these accessions (see Results).

A general finding is that *B.rapa* accessions are separated into groups corresponding to their geographical origin (Song et al. 1988, 1990; Zhao et al. 2005; Takuno et

al. 2007; Annisa et al. 2013; Guo et al. 2014, Zhang et al. 2014). However, we did not perceive any effect of origin on grouping, e.g. the biggest group 3 contained spring oiltypes from several countries all over the world. In groups 1 and 2, there were not so many accessions and e.g. the 5 turnips were all European, and from the 3 leaf types 2 were Chinese and one was Japanese. Therefore, it is difficult to differentiate the effect of origin from that of the morphotype in these groups. At least partly convergent results compared to ours include clustering of winter and spring oiltypes separately in *B.rapa* (Zhao et al. 2010; Annisa 2013) and in *B.napus* (Hasan et al. 2006; Bus et al. 2011). Further, Pino del Carpio (2011) could cluster *B.rapa* accessions into 4 groups showing correspondence to morphotype and geographic origin (Asia and Europe). Comparison of the results from various diversity studies is complicated because different marker types and plant collections of different makeup have been used. Marker types differ from each other regarding e.g. the efficiency to measure distances between individuals (dominant vs. codominant, random vs. mapped markers). Moreover, one crucial factor is the amount of individuals analyzed per accession (from one to several) and if they are studied separately or bulked.

Most of the variation in our *B.rapa* collection was within accessions (65%), which is typical for a self-incompatible species. Variation among accessions was 25% and among STRUCTURE groups 10%. Previously, it has not usually been possible to investigate variation within *B.rapa* accessions because only one individual per accession or bulked individuals have been used. However, in one study with 8 oilseed *B.rapa* cultivars, genetic variation was reported to be 60%, 22% and 18% within cultivars, across cultivars and across 2 geographic regions, respectively (Zhao et al. 2009). When accessions have been grouped according to STRUCTURE or hierarchical cluster analysis, the among group variation has been from 11% to 16% (Pino del Carpio 2011; Annisa et al. 2013; Guo et al. 2014).

It has been debated that SNP markers developed by resequencing of only a few genotypes may suffer from ascertainment bias. Consequently, estimates of population genetic parameters, allele frequency distribution and linkage disequilibrium can be biased (Heslot et al. 2013). In our study, SNPs were generated based on sequencing of 8 *B.rapa* genotypes. We therefore also studied diversity with GBS-SNPs developed in a set of 84 individuals from our plant material. GBS is a method where discovery and genotyping of SNPs is done in one step which leads to marker data with less bias. The diversity indices and distance matrices with GBS-SNPs and Illumina BeadXpress SNPs correlated well with each other, and the NJ trees looked quite similar. The higher diversity values for Illumina BeadXpress SNPs refer to the fact that during the SNP marker design process, we had selected the most polymorphic SNPs whereas all GBS-SNPs (excluding monomorphic ones) were used in the analyses. There seems to be no bias in our SNP marker set, which is due to the fact that we had succeeded to select genotypes representing all the 3 STRUCTURE groups for resequencing, and the genotypes thus

covered the diversity of *B.rapa* quite well. Consequently, this smaller number of SNP markers was evidently big enough for diversity analyses.

The outcome from the present study was a genetically diverse collection of mainly oilseed *B.rapa* accessions and a set of 209 polymorphic SNP markers, which was used to investigate genetic diversity, genetic distances between individuals, and structure in the *B.rapa* collection. As far as we know, genetic relationships in oilseed *B.rapa* have not previously been investigated with SNP markers, and new interesting information was revealed. According to our results, the accessions formed 3 groups which seemed to correspond to morphotype and flowering habit but not to geography as generally perceived in previous studies. Assessment of genetic diversity is the basis of all plant breeding and also crucial when studying genetic resources of the species. The SNP markers developed can be used in future mapping and diversity studies as well as for plant breeding programs.

Acknowledgements

We want to thank Marja-Riitta Arajärvi, Leena Lohermaa, Sirpa Moisander, Kirsti Mäkelä, Leena Ramstedt, and Anneli Virta for their excellent technical assistance. We also appreciate the assistance of Ilma Tapio, Terhi Iso-Touru and Daniel Fischer. The help of Katja Hämäläinen, the breeder, and the staff at Boreal Plant Breeding Ltd. is also greatly acknowledged. We are grateful for the genebanks CGN, IPK, USDA and VIR, Boreal Plant Breeding Ltd., and Kevin Falk (Canadian accessions) for providing the seeds for our study. The Ministry of Agriculture and Forestry in Finland, The Finnish Funding Agency for Technology and Innovation, Neste Oil Ltd., Raisioagro Ltd., and Boreal Plant Breeding Ltd. are acknowledged for their financial support.

References

- Annisa, Chen, S., and Cowling, W.A. 2013. Global genetic diversity in oilseed *Brassica rapa*. *Crop Pasture Sci.* 64: 993-1007.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23: 2633-2635.
- Bus, A., Körber, N., Snowdon, R.J., and Stich, B. 2011. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. *Theor. Appl. Genet.* 123: 1413-1423.

- Chung, H., Jeong, Y.-M., Mun, J.-H., Lee, S.-S., Chung, W.-H., and Yu, H.-J. 2014. Construction of a genetic map based on high-throughput SNP genotyping and genetic mapping of a TuMV resistance locus in *Brassica rapa*. *Mol. Genet. Genomics*, 289: 149-160.
- Delourme, R., Falentin, C., Fomeju, B.F., Boillot, M., Lassalle, G., André, I., Duarte, J., Gauthier, V., Lucante, N., Marty, A., Pauchon, M., Pichon, J.-P., Ribière, N., Trotoux, G., Blanchard, P., Rivière, N., Martinant, J.-P., and Pauquet, J. 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics*, 14: 120.
- Devisetty, U.K., Covington, M.F., Tat, A.V., Lekkala, S., and Maloof, J.N. 2014. Polymorphism identification and improved genome annotation of *Brassica rapa* through deep RNA sequencing. *G3*, 4: 2065-2078.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6: e19379.
- Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites. *Genetics*, 131: 479-491.
- Falush, D., Stephens, M., and Pritchard, J. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164: 1567-1587.
- Gómez-Campo, C., and Prakash, S. 1999. Origin and domestication. In *The biology of Brassica coenospecies*, vol 4. Edited by C. Gómez-Campo. Elsevier, Amsterdam. pp. 33-58.
- Guo, Y., Chen, S., Li, Z., and Cowling, W.A. 2014. Center of origin and centers of diversity in an ancient crop, *Brassica rapa* (turnip rape). *J. Hered.* 105(4): 555-565.

- Hasan, M., Friedt, W., Pons-Kühnemann, J., Freitag, N.M., Link, K., and Snowdon, R.J. 2008. Association of gene-linked SSR markers to seed glucosinolate content in oilseed rapa (*Brassica napus* ssp. *napus*). Theor. Appl. Genet. 116: 1035-1049.
- Heslot, N., Rutkoski, J., Poland, J., Jannink, J.-L., and Sorrells, M.E. 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One, 8: e74612.
- Hubisz, M., Falush, D., Stephens, M., and Pritchard, J. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9(5): 1322–1332.
- Kimber, D.S., and McGregor, D.I. 1995. The species and their origin, cultivation and world production. In Brassica oilseeds, production and utilization. Edited by D.S. Kimber and D.I. McGregor. CAB international, UK, pp. 1-7.
- Li, F., Kitashiba, H., Inaba, K., and Nishio, T. 2009. A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. DNA Res. 16: 311-323.
- Lin, K., Zhang, N., Severing, E.I., Nijveen, H., Cheng, F., Visser, R.G.F., Wang, X., de Ridder, D., and Bonnema, G. 2014. Beyond genomic variation – comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics, 15: 240.
- Liu, K., and Muse, S.V. 2005. Powermarker: integrated analysis environment for genetic marker data. Bioinformatics, 21 (9): 2128-29.
- Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., Snowdon, R.J., and Li, J. 2013. A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. PLoS One, 8: e83052.
- Mantel, N.A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27: 209-220.
- McGrath, J.M., and Quiros, C.F. 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor. Appl. Genet. 83: 783-790.

- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. 2010. Tablet – next generation sequence assembly visualization. *Bioinformatics*, 26 (3): 401-402.
- Paritosh, K., Yadava, S.K., Gupta, V., Panjabi-Massand, P., Sodhi, Y.S., Pradhan, A.K., and Pental, D. 2013. RNA-seq based SNPs in some agronomically important oleiferous lines of *Brassica rapa* and their use for genome-wide linkage mapping and specific-region fine mapping. *BMC Genomics*, 14: 463.
- Park, S., Yu, H.-J., Mun, J.-H., and Lee, S.C. 2010. Genome-wide discovery of DNA polymorphism in *Brassica rapa*. *Mol. Genet. Genomics*, 283: 135-145.
- Peakall, R., and Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, 6: 288-295.
- Peakall, R., and Smouse, P.E. 2012. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, 28: 2537-2539.
- Pino del Carpio, D., Basnet, R.M., De Vos, R.C.H., Maliepaard, C., Visser, R., and Bonnema, G. 2011. The patterns of population differentiation in a *Brassica rapa* core collection. *Theor. Appl. Genet.* 122: 1105-1118.
- Pritchard, J., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- Ramchiary, N., and Lim, Y.P. 2011. Genetics of *Brassica rapa* L. In *Genetics and genomics of the Brassicaceae*. Edited by R. Schmidt and I. Bancroft I. Springer, New York pp 215-260.
- Raman, H., Dalton-Morgan, J., Diffey, S., Raman, R., Alamery, S., Edwards, D., and Batley, J. 2014. SNP markers-based map construction and genome-wide linkage analysis in *Brassica napus*. *Plant Biotechnol. J.* 12: 851-860.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16 (6): 276-277.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Shannon, C.E. 1948. A mathematical theory of communication. *AT&T Tech. J.* 27(3): 379-423 and 623-656.

- Song, K.M., Osborn, T.C., and Williams, P.H. 1988. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B.rapa* (syn. *campestris*) and *B.oleracea*. Theor. Appl. Genet. 76: 593-600.
- Song, K.M., Osborn, T.C., and Williams, P.H. 1990. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 3. Genome relationships in *Brassica* and related genera and the origin of *B.oleracea* and *B.rapa* (syn. *campestris*). Theor. Appl. Genet. 79: 497-506.
- Song, X., Ge, T., Li, Y., and Hou, X. 2015. Genome-wide identification of SSR and SNP markers from the non-heading Chinese cabbage for comparative analysis. BMC Genomics, 16: 328.
- Takuno, S., Kawahara, T., and Ohnishi, O. 2007. Phylogenetic relationships among cultivated types of *Brassica rapa* L. em. Metzg. as revealed by AFLP analysis. Genet. Resour. Crop Ev. 54: 279-285.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.
- Tanhuanpää, P., and Manninen, O. 2012. High SSR diversity but little differentiation between accessions of Nordic timothy (*Phleum pratense* L.). Hereditas, 149 (4): 114-127.
- Tinker, N.A., Fortin, M.G., and Mather, D.E. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor. Appl. Genet. 85: 976-984.
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weisshaar, B., Liu, B., Li, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Yu, J., Meng, J., Min, J., Poulain, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Sato, S., Sun, S.,

- Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Z., Li, Z., Xiong, Z., and Zhang, Z. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43: 1035–1040.
- Zhang, N., Zhao, J., Lens, F., de Visser, J., Temesgen, M., Fang, W., Xiao, D., Bucher, J., Basnet, R.K., Lin, K., Cheng, F., Wang, X., and Bonnema, G. 2014. Morphology, carbohydrate composition and vernalization response in a genetically diverse collection of Asian and European turnips (*Brassica rapa* subsp.*rapa*). *PloS One*, 9 (12): e114241.
- Zhao, J., Wang, X., Deng, B., Lou, P., Wu, J., Sun, R., Xu, Z., Vromans, J., Koornneef, M., and Bonnema, G. 2005. Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. *Theor. Appl. Genet.* 110: 1301-1314.
- Zhao, J., Artemyeva, A., Pino del Carpio, D., Basnet, R.M., Zhang, N., Gao, J., Li, F., Bucher, J., Wang, X., Visser, R.G.F., and Bonnema, G. 2010. Design of a *Brassica rapa* core collection for association mapping studies. *Genome*, 53: 884-898.
- Zhao, Y.-G., Ofori, A., and Lu, C.-M. 2009. Genetic diversity of European and Chinese oilseed *Brassica rapa* cultivars from different breeding periods. *Agr. Sci. China*, 8(8): 931-938.

Figure captions:

Fig. 1 Grouping of *B.rapa* accessions with the program STRUCTURE with $K=3$. Accessions are represented by a bar composed of segments with different colours according to the accession's average membership fractions of 3 clusters. Symbols indicate different *Brassica* types.

Fig. 2 Dendrograms showing relationships between individuals from 12 *B.rapa* accessions, based on a neighbor-joining algorithm using Euclidean distance matrices constructed with either 209 Illumina BeadXpress SNPs or 5727 GBS-SNPs.

ESM Fig.S1 A neighbor-joining tree of 893 *B.rapa* individuals from 61 accessions based on Euclidean distances with 209 SNP markers. Individuals have been marked with different symbols based on different *Brassica* types: leaf types, turnips, winter oilseeds (cultivars marked separately), spring oilseeds (breeding material and accession K-145 marked separately), ssp. *trilocularis*, ssp. *dichotoma*, and wild types (var. *silvestris*). As an example of individuals with erroneous localization, two spring oilseed individuals (from accessions K-228 and K-303) have been pointed out with arrows.

Draft

Table 1. List of 61 *Brassica rapa* accessions with their observed diversity indices: % polymorphic loci, Shannon's diversity index (H'), observed heterozygosity (H_o), and d_{ST} .

Eight accessions marked in bold were also involved in GBS analysis.

Accession ¹	Accession name	Source ²	Type ³	Country of origin	Flowering habit ⁴	No. of ind.	% polymorphic loci
Mizuna Kyoto		seed company	leaf type (<i>ssp. japonica</i>)	Japan	S	16	46.4
Michihili		seed company	leaf type (<i>ssp. pekinensis</i>)	China	S	15	70.3
Cantong White		seed company	leaf type (<i>ssp. chinensis</i>)	China	S	16	54.6
CR 1643	Yellow Sarson	IPK	oiltype (<i>ssp. dichotoma</i>)	unknown	S	4	1.0
CR 2301		IPK	oiltype (<i>ssp. dichotoma</i>)	China	S	4	69.9
PI 173847	9751	USDA	oiltype (<i>ssp. trilocularis</i>)	India	S and/or W	4	31.6
PI 459015	NU-42929	USDA	oiltype (<i>ssp. trilocularis</i>)	India		3	0.0
PI 459018	NU 47865	USDA	oiltype (<i>ssp. trilocularis</i>)	India		4	4.3
PI 649203	Ames 26166	USDA	oiltype (<i>ssp. trilocularis</i>)	United States, Iowa	S	4	64.1
PI 649205	Ames 26168	USDA	oiltype (<i>ssp. trilocularis</i>)	United States, Iowa	S	4	5.3
K-145	Danus	VIR	oiltype	China	S	14	44.5
CGN7159	Gelria A	CGN	turnip	Netherlands	W	15	46.4
CGN7171	Debra	CGN	turnip	Netherlands	W	15	64.1
CGN11009	Siloganova	CGN	turnip	Netherlands	W	16	61.7
Milan White		seed company	turnip	Italy	W	16	52.2
Snowball		seed company	turnip	United Kingdom	W	16	69.9
CR 2210		IPK	oiltype, var. <i>silvestris</i>	Slovakia	S (SW ⁵)	16	52.2
CR 2211		IPK	oiltype, var. <i>silvestris</i>	Georgia	S (SW ⁵)	16	69.9
CR 2212		IPK	oiltype, var. <i>silvestris</i>	China	S (W ⁶)	16	60.8
CR 2213		IPK	oiltype, var. <i>silvestris</i>	Austria	S	16	64.1
CR 2616		IPK	oiltype, var. <i>silvestris</i>	Nepal	S (SW ⁵)	15	72.7
CR 3396		IPK	oiltype	Tibet	S	14	83.3
PI 633159	Su Weon	USDA	oiltype	Korea		14	50.2
K-193	Izumrudnaya	VIR	oiltype, var. <i>biennis</i>	Ukraine	W	16	60.8
K-238	Vesnyanka	VIR	oiltype, var. <i>biennis</i>	Ukraine	W	16	70.3
K-239	Gorlitca	VIR	oiltype, var. <i>biennis</i>	Ukraine	W	14	68.4
Largo_s		BOR*	oiltype, var. <i>biennis</i>	Sweden	W	15	66.5
Salut_s		BOR*	oiltype, var. <i>biennis</i>	Sweden	W	16	70.8
BRA 2218		IPK	oiltype	Canada	S	16	78.5
K-12		VIR	oiltype	Argentina	S	16	86.1
K-13		VIR	oiltype	Argentina	S	15	86.6
K-17	Local n 5034	VIR	oiltype	China	S	16	88.0
K-26	Polan uzai n 5083	VIR	oiltype	China	S	15	90.4
K-27	Chju-Shahun n 5091	VIR	oiltype	China	S	15	89.5
K-31	Local n 5163	VIR	oiltype	China	S	15	90.4
K-148	Kuimin	VIR	oiltype	China	S	16	86.1
K-218	Nabo silvestre	VIR	oiltype	Peru	S	16	78.0
K-223	San Lorenzo	VIR	oiltype	Mexico	S	16	73.7
K-226		VIR	oiltype	Finland	S	16	82.3
K-227		VIR	oiltype	Finland	S	16	80.4
K-229	Mun Yuz	VIR	oiltype	Korea	S	16	90.9
K-252	Mun Von Sojie	VIR	oiltype	Korea	S	16	72.3
K-253	NAT-CZJS 1366	VIR	oiltype	Tanzania	S	15	81.8
K-254	Tumbuck/Jumbuck	VIR	oiltype	Australia	S	16	68.4
K-303	Nels	VIR	oiltype	Nepal	S	16	86.6
K-308	Sals	VIR	oiltype	Nepal	S	15	68.4
ACS-C29		AAFC	oiltype	Canada	S	16	79.4
Early One		AAFC	oiltype	Canada	S	16	76.6
Mawerick_s		BOR*	oiltype	Canada	S	19	73.7
Synergy		AAFC	oiltype	Canada	S	16	71.8
Bor01_s		BOR	oiltype	Finland	S	19	62.7
Bor02_s		BOR	oiltype	Finland	S	17	72.3
Bor03_s		BOR	oiltype	Finland	S	19	69.9
Bor04_s		BOR	oiltype	Finland	S	14	72.7
Bor05_s		BOR	oiltype	Finland	S	15	75.6
Bor06_s		BOR	oiltype	Finland	S	16	79.4
Cordelia		BOR	oiltype	Finland	S	18	77.0
Cordelia_s		BOR	oiltype	Finland	S	17	78.0
SW Petita		seed company	oiltype	Sweden	S	21	76.6
SW Petita_s		BOR*	oiltype	Sweden	S	20	74.6
SW Rebus_s		BOR*	oiltype	Sweden	S	18	67.5

*original seed from commercial source

¹ s= selfed, lines that have been self-pollinated

² genebanks: CGN, IPK, USDA, VIR; plant breeding companies: AAFC, BOR

³ the oiltypes are *ssp. oleifera* if not mentioned otherwise

⁴ S=spring, SW=semi-winter, W=winter

⁵ SW according to Annisa et al. 2013

⁶ W according to Guo et al. 2014

[illegible]

Table 2. Analysis of molecular variance in 57 self-incompatible *B.rapa* accessions based on SNP markers.

Source	df	SS	MS	Variance components	% total
Among STRUCTURE groups	2	5814.45	2907.23	10.74	10
Among accessions	54	25083.75	464.51	25.77	25
Within accessions	821	55198.43	67.23	67.23	65
Total	877	86096.64		103.75	100

Stat	Value	P(rand >= data)
PhiRT	0.104	0.001
PhiPR	0.277	0.001
PhiPT	0.352	0.001

Probability, P(rand >= data), for PhiRT, PhiPR and PhiPT is based on 999 permutations across the full data set.

PhiRT = AG / (WA + AA + AG) = AG/ TOT

PhiPR = AA / (WA + AA)

PhiPT = (AA + AG) / (WA + AA + AG) = (AA + AG) / TOT

Key: AG = est. var. among groups, AA = est. var. among accessions, WA = est. var. within accessions.

Draft

GBS-SNPs Genome Illumina BeadXpress SNPs