International Journal of Agriculture, Environment and Bioresearch

Vol. 3, No. 02; 2018

ISSN: 2456-8643

DISCOVERY OF NOVEL SINGLE NUCLEOTIDE POLYMORPHIC (SNP) MARKERS FOR GENETIC MAPPING OF CASHEW(Anacardium occidentale. L)

Geradina P. Mzena[1, 2, 3*] Paul M. Kusolwa[3], Gration M. Rwegasira[3] and Nasser K. Yao[4]

[1] Mikocheni Agricultural Research Institute, P. O. Box 6226, Dar es Salaam, Tanzania.

[2] Naliendele Agricultural Research Institute, P. O. Box 509, Mtwara, Tanzania.

[3]Department of Crop Science and Horticulture, Sokoine University of Agriculture, P. O. Box 3005, Morogoro,

[4]International Livestock Research Institute, P.O. Box 30709, Nairobi 00100, Kenya

ABSTRACT

Cashew (Anacardium occidentale. L) crop, is the major cash crop in Tanzania. Cashew breeding strategies rely on conventional breeding, which is time consuming and labor intensive. This study aimed at developing single nucleotide polymorphic (SNPs) markers and genetic linkage map for identifying quantitative traits loci (QTL) associated with yield, nut quality and plant size. In this study, Genotyping by Sequencing (GBS) technique, tagged as DArTSeq, was used to identify SNPs from reduced complexity cashew genome of F1 and F2 as well as ATA19/250 (Male parent) and Cook05 (female parent). The reduced libraries were sequenced using Hiseq 2500 illumina sequencer, besides, SNPs call was done by using DArTsoft v.7.4.7 (DArT P/L, Australia).The study identified 6,364 high quality SNPs, of which 57.6% were transition and 42.4% were trans version. The average call rate, reproducibility and polymorphic information content (PIC) of the SNPs was 99%, 98% and 0.31 respectively. Among the 6,364 markers,1009 were codominant markers of which 761 (75%) were assigned to 21 linkage groups of total length 2,330.6 cM with an average distance of 3.3cM between markers. This study revealed the first SNPs linkage map which can be used in identifying QTL for efficient breeding program through marker assisted selection

Keywords: Anacardium occidentale, genetic linkage map, single nucleotide polymorphic markers, quantitativetraitloci, marker assisted selection.

INTRODUCTION

Cashew (*Anacardium occidentale. L)* belongs to the family Anacardiaceae which is composed of 74 genera and 600 species [1]. Cashew is diploid with chromosome number 2n=42 [2]. It is a tropical perennial crop, widely adapted and drought tolerant [3]. In Brazil, India, Vietnam, and Tanzania and in some other Asian and African countries, cashew nuts play a key role as food security and income generating crop [4]. In Tanzania ,it is the fourth most important cash crop after coffee, tobacco and cotton [5] – [6]. Despite its importance, the production of cashew nut in Tanzania is lower (160 000 mt) as compared to other countries like India (725 000 mt), Nigeria (847 794 mt) and Vietnam (1,250,000 mt) [7]. The low production in Tanzania is duetoanumber

Vol. 3, No. 02; 2018

ISSN: 2456-8643

of production constraints such as low yields of local varieties, prevalence f economically important diseases and pests, shortage of good quality planting materials as well as limited application of good agricultural practices [8].

Considerable efforts have been made to develop improved varieties using conventional breeding approaches. However, the success is low due to the perennial nature and the long gestation period of at least 9 – 12 years to complete one breeding cycle [9]. In addition, the time taken to release varieties through application of conventional breeding is a major obstacle in the development of new varieties. The procedure is laborious, involving several crosses, several generations, and careful phenotypic selection making it further difficult to achieve the breeding objectives [10]. Based on that, researchers have been developing DNA based molecular markers to complement on going conventional breeding approaches in cashew. In cashew, the most commonly used molecular markers include Random Amplified Polymorphic DNA (RAPD) and Simple Sequence Repeat (SSR) markers but these markers are few and currently only 51 sequences are available in the public database [11]. The small number of available markers has been a great limitation in developing linkage mapping and diversity analyses of cashew genetic resources. Recently, Genotyping by Sequencing (GBS) as per the Diversity Array Technology protocol (DArTSeq), which is based on next generation sequencing technology and related bioinformatics, has facilitated large-scale discovery of SNPs markers for marker assisted selection (MAS) [12]. The application of SNP markers for crop improvement through MAS has been demonstrated with great success in a number of crops such as wheat [13]. MAS is now commonly employed in perennial crop breeding programs to accelerate cultivar development [14] – [15].In particular, MAS has shown to provide advantages for selection during the juvenile phase [16] – [17]and for replacing expensive, time-consuming or technically difficult traits [18] – [19]. Marker assisted selection have also been applied to quantitatively inherited traits (QTL) effect, including fruit acidity in peach [20], fruit size in tomato [21],and cherry [22]and grain yield in rice[23]. One strategy for implementing MAS involves the development of a genetic linkage map for localization of QTLs [24]. This study aimed at identifying SNP markers and developing genetic linkage map which will be used to identify and localize QTLs associated with economically importanttraits of cashew suchas yield, nut quality and plant size.

2.0MATERIALS AND METHODS

2.1 Planting materials

The study was carried out at the Experimental Station of Naliendele Agriculture Research, in Mtwara region of Tanzania. Naliendele Agriculture Research Institute is located at 10.3539° S, 40.1682° E and 380m above sea level.

One hundred and sixty five (165) F_2 individuals, 3 F_1 progenies as well as Cook05 (Female parent and dwarf cashew with small nuts <6g) and ATA19/250 (Male parent and common giant with big nuts >6g), both grown at the Naliendele Agricultural Research Institute, Mtwara, Tanzania were used in this study.

Vol. 3, No. 02; 2018

ISSN: 2456-8643

2.2DNA Extraction

DNA was extracted based on the cetyl trimethyl ammonium bromide (CTAB) method of Doyle and Doyle, [25] with little modification. Approximately 0.25g of leaf samples were grounded into fine powder using mortar and pestle. The powdered tissue was transferred into 1.5 ml eppendorf tube and mixed with 700µl of preheated (65^{o}C) 2% CTAB extraction buffer. The extraction buffer consisted of 0.2M Tris (hydroxyl methyl) amino methane hydrochloride (Tris-HCl), 1.4M sodium chloride (NaCl) at pH of 8.0, 2% (w/v) polyvinyl pyrrolidone (PVP), 0.05M ethylene-diamine tetra acetate (EDTA) and 0.2% (v/v) β-mercarptoethanol. The mixture was shaken vigorously for 10 min until all tissue were dispersed in the buffer and transferred to 65^{o}C heating block for 15 min and thereafter cool at room temperature for 2 min. 250 µl of ice-cold 5M potassium acetate was added and mixed by gently inverting the tubes 5-6 times before incubation on ice for 20 min. The mixture was then centrifuged at 14 000 rpm for 10 min to separate supernatant. The supernatant was transferred to a new 1.5 ml eppendorf tube and 700 µl of ice-cold isopropanol was added and mixed by gently inverting the tube 8-10 times. The samples were incubated overnight and centrifuged at 14 000 rpm for 10 min.

The supernatant was poured off and the last drops of isopropanol were removed by placing the eppendorf tube face down on paper towels. The pellet was air dried on a paper towel for 1h. Finally, 100 µl of 10 mM Tris-HCl, 1mM EDTA containing 10 mg/ml RNase was added and stored overnight at 4^{o}C to dissolve the pellet.

2.3 Determination of DNA quantity and quality

Quantity and quality of isolated DNA samples were determined using the NanoDrop spectrophotometer (ND-1000 V3.5, NanoDrop Technologies, Inc.) and agarose gel electrophoresis, respectively. The DNA quantity of each sample was determined through direct reading from the NanoDrop. The A260/A280 nm ratio was used to provide an estimate of DNA purity. The quality of each DNA was determined by a clear band pattern after running the sample on a 0.8% (w/v) agarose gel in 1x Tris-acetate EDTA (TAE) (40 mM Tris-HCl and 1 mM EDTA at pH 8.3 adjusted with acetic acid) buffer containing ethidium bromide (10 mg/ml) at 120 V for 30 min. The simultaneous loading of standard λ-DNA at concentrations of 50, 100, 150 and 200 ng/µl also allowed the determination of DNA concentration. DNA was visualized under ultraviolet (UV) illumination, using a transilluminator spectrophotometer (Super imagine UK). After determination of DNA concentration and quality, each sample was diluted to a working concentration of 10 ng/µl using double distilled water.

2.4 Preparation of Genomic Library Representation (Targets)

Following the Diversity Array Technology (DArT) protocol, genomic representations were generated by digesting 100ng of DNA samples using a combination of rare cutter (PstI) and a frequent cutter (MSEI). A PstI adapter (5′-CAC GAT GGA TCC AGT GCA-3′ annealed with 5′-CTG GAT CCA TCG TGC A-3′) was ligated with T4 DNA ligase.A 1-µl aliquot of the ligation product was used as a template in 50-µl amplification reactions with DArT-*Pst*I primer (5′-GAT GGA TCC AGT GCA G-3′). For amplification, the following PCR program was used: 30 cycles

Vol. 3, No. 02; 2018

ISSN: 2456-8643

of 94°C for 20 sec, 58°C for 40 sec, 72°C for 1 min, and followed by a final extension step of 7 min at 72°C. The quality of PCR products was determined using 1.2% agarose gel.

2.5 Cluster generation and DNA Sequencing

The well amplified pooled targets were sent to c-Bot for cluster generation. The clusters were formed through hybridization reaction mix of 994µl of hybrid buffer mixed with 6µl of denatured DNA. From the mix, 75µl were loaded into the template strips and run in the cBot to generate clusters for sequencing using Illumina Hiseq2500 sequencer. The sequencing (single read) was run for 77 cycles in two lanes with a sequencing depth of 30X. The sequencer operates by immobilizing the DNA template onto a flow-cell and then each single DNA molecule is amplified using a 'bridging PCR' amplification reaction. The DNA sequence is finally recorded utilizing four different fluorescent labels captured by an optical camera.

2.6 SNP Calling

The images from Hiseq2500 Illumina sequencer were analyzed using DArTsoft v.7.4.7 (DArT P/L, Canberra, Australia) to identify polymorphic candidate SNP and silicoDArT markers. The latter were scored as binary data (1/0), indicating presence or absence of a marker in genomic representation of each sample as described by Wenzl*et al*. [26]. Clones with an average genotype call rate of 98% and a scoring reproducibility of 99% were selected as markers. The informativeness of the DArTSeq markers was determined by calculating the polymorphism information content (PIC), whereby the maximum PIC was 0.5.

2.7 Genetic Linkage analysis

Since the mapping population was derived from a self-pollination of the F_1 progeny, we first selected SNPs that were heterozygous in the F_1 to ensure that they segregated in F_2 progeny. The informative markers were expected to segregate in a 1:2:1 ratio in the F_2 mapping population, and

those that deviated significantly from the expected Mendelian segregation ratio ($\chi 2$ test p-value < 0.01) were excluded from further analysis. Linkage analysis was performed using JoinMap® v 3.0 software package [27] using the parameters set for F_2 method and "Create Population Node" function. Grouping and order of markers was done using independence logarithm of odds (LOD) threshold value of 6. We used linkage with recombination ratio (REC) threshold of 0.5, a map LOD value of 0.05 and a goodness of fit jump threshold value of 5.0 for calculation of linear order of the markers within the linkage group. Recombination frequencies were converted to map distances in centiMorgan (cM) using the Kosambi mapping function [28] and linkage map was drawn by using MapChart V. 2.30 software of Wageningen University, Netherlands [29].

3.0 RESULTS

A total of 96 380 248reads with a total of 7 421 279 118 base pairs (bp) were generated from reduced library of cashew genomic representations of the two parents (ATA19/250 and Cook05),

Vol. 3, No. 02; 2018

ISSN: 2456-8643

3F_1 and 165 F_2 individuals of the mapping population (Table 1).After read trimming, 70,789,261 quality reads (73.44%) and 5 450 773 155 base pairs were obtained, which is equivalent to a coverage of 8.0% of the cashew genome size of 419 Mb.

Table 1: Summary of the DArTseq Genotyping by Sequencing Reads Before and After Trimming

	Before trimming	After trimming
Total reads	96 380 248	70 789 261
Average reads per sample	566 942	416 407
Total bp	7 421 279 118	5 450 773 155
Average genome size per sample (bp)	43 654 583	32 063 339
% Coverage of genome[1)]	10.4	8.0

bp = base pair based on cashew genome size of 419 Mb as estimated by Aliyu, [30]

A total of 9965 SNPs were obtained using DArTsoft v.7.4.7 (DArT P/L, Canberra, Australia). After filtering with compound filter, 6364 (63.9%) putative SNPs revealed an average call rate of 99% and average reproducibility of 98%. The SNPs were classified into transitions (Ti) and transversions (Tv) based on nucleotide substitution. About 57.6% of the base changes were transitions and 42.4% were transversions (Table2).

Table 2: Summary of SNPs identified in cashew

	Type of marker	No of markers	Percentage	
	Transition			
	G>A	966	15.2	
	A>G	890	14	
	T>C	855	13.4	
	C>T	953	15	
	Transversion			
	A>C	310	4.9	
	G>C	323	5.1	
	G>T	340	5.3	
	T>A	345	5.4	
	A>T	381	6	
	C>A	396	6.2	
	C>G	306	4.8	
	T>G	299	4.7	
	Grand Count	6,364	100	

The informativeness of the markers based on the polymorphic information content (PIC) ranged from 0.1 – 0.5 (Table 3). The average PIC value of the 6 364 selected markers was 0.37.

Vol. 3, No. 02; 2018

ISSN: 2456-8643

Among the 6 364 high quality SNPs, 1 345 (21.1%) were heterozygous in the F_1 parents (that were selfed to produce the F_2 mapping population). The remaining 5 019 (78.9%) makers were homozygous and were not considered for linkage analysis. Out of the 1 345 segregating SNP markers, 1009 (75.0%) showed a Mendelian segregation ratio of 1:2:1 ($P<0.05$) in the F_2 mapping population of which only 761 SNPs were mapped onto 21 linkage groups (Fig.1**).**

Table 3: Polymorphic Information Content (PIC), number and percentage of polymorphic SNPs markers generated from cashew genotypes

PIC value range	Number of SNP markers	%
0.46 - 0.50	1 610	25.3
0.41 - 0.45	962	15.1
0.31 - 0.40	1 556	24.5
0.21- 0.30	1 230	19.3
0.10 - 0.20	1 006	15.8
Total	6 364	100

The genetic map had a total length of 2,330.6 cM, with an average marker interval of 3.3 cM (Table 4); and the length of the individual linkage groups (LGs) ranged from 25.8cM (LG19) to 206.6cM (LG4) (Table 4). The smallest linkage group was LG21, which consisted of 9 markers spanning a length of 38.3 cM, while the largest linkage group was LG3 with 83 markers and a length of 182.1 cM (Table 4) Fig. 1.

Table 4: Number of linkage groups derived from 165 F2 cashew genotypes and their corresponding number of markers, size (length) and average distance between markers (in centiMorgan)

Linkage Group	No. of Markers	Length (cM)	Average distance (cM) between markers
1	69	199.4	2.9
2	19	61.3	3.2
3	83	182.1	2.2
4	68	206.6	3.0
5	77	191.4	2.5
6	51	132.9	2.6
7	47	130.8	2.8
8	56	195.5	3.5
9	37	135.8	3.7
10	37	104.1	2.8
11	31	146.5	4.7
12	34	113.1	3.3
13	26	59.4	2.3
14	21	104.5	5.0

Vol. 3, No. 02; 2018

ISSN: 2456-8643

15	21	56.7	2.7
16	21	72.6	3.5
17	16	47.8	3.0
18	15	61.4	4.1
19	12	25.8	2.2
20	11	64.6	5.9
21	9	38.3	4.3
Total/Average		**2 330.6**	**3.3**

Figure1 Linkage map of cashew derived from 165 F2 cashew genotypes obtained from selfing F1 developed through controlled crossing of two distinct cashew clones, namely a dwarf cashew clone (Cook 05) and a common (giant) cashew clone (ATA19/250). Distances are given in centiMorgans (cM) on the left with markers placed on the right.

4.0 DISCUSSION

This study aimed at developing novel single nucleotide polymorphism (SNPs) markers by using Next generation sequencing methods (Hiseq 2500). NGS has been facilitating sequencing of complex genome of which the reference genome sequence are not yet available including cashew, tobacco. In this study, the obtained 9,965 SNPs were filtered based on call rate (percentage of target that could be scored as '0' or '1') and reproducibility value (the proportion of technical replicate assay pairs for which the marker score is consistent) as per the metadata description developed by DArT, Canberra, Australia. After filtration, the selected 6364 SNPs had average call rate, reproducibility and PIC value of 99%, 98% and 0.37 respectively. This confirmed that the 6364 SNPs are high quality markers and suggested to be used in cashew genomic analysis. This results are comparable to those obtained for pigeon pea [31], barley [32], cassava [33] and wheat [34]. Among the 6,364 obtained polymorphic SNPs, 57.6% were transitions SNPs (C/T and A/G), and 42.4% were transversions SNP (A/T, A/C, T/G and C/G). This indicates that transitions SNPs are higher than transversions SNPs which is due to the fact that transitions SNPs are more easily substituted than transversion SNPs since purine-purine (transitions) substitution is easy and stable than purine-pyrimidine (transversions) substitution. This result is similar with other previous research as reported by Prasad [35].

The highly polymorphic SNP markers identified in this study were used to develop a genetic map that can be used for implementing MAS in cashew breeding. The genetic linkage map was developed using 761 codominant SNP markers that segregated at a ratio of 1:2:1 in the F_2 population. Codominant markers allow accurate estimation of recombination frequencies for accurate alignments of linkage map. According to Ritter et al., [36]the use of mixed dominant and co-dominant markers has an inherent disadvantage of allowing contradictions and unsafe alignments due to the different degrees of accuracy in estimating recombination frequencies.

Vol. 3, No. 02; 2018

ISSN: 2456-8643

In this study, 25.0% of the codominant markers showed distorted segregation from the 1:2:1 segregation ratio as revealed by the Chi-square test ($P<0.01$). Segregation distortions are deviations from normal Mendelian genetic ratios caused by various reasons that include technical problems in genotyping and scoring errors, genetic drift, locus duplication and chromosome rearrangement. Segregation distortions have been reported in several studies and a high frequency of markers showing distorted segregation is common in outcrossing species [37]. Besides, deviations from Mendelian segregation ratios may be due to various processes amongst which may be the presence of gametophytic selection for sub-lethal genes, genes controlling the viability of pollen, zygote or seedlings, putatively located on one or more of the linkage groups Similar results have been reported in various studies such as Peach [38], Quercus [39], Eucalyptus, [40]. The distorted markers were not considered in the analysis to avoid reduction in map length because according to Liu and Muse [41] markers exhibiting significant deviation from the expected Mendelian ratio lead to a significant under- or overestimation of recombination fractions. This consequently biases the calculation of genetic distances and order of markers which can underestimate distances between closely linked markers.

The distribution of marker along the linkage groups (LG) was not uniform, as evident by the mixture of tightly linked loci (LG 3) and regions with low density (LG 21) as observed in the constructed map. This suggests that either recombination events or mapped loci were not evenly distributed throughout the genome. The low density of markers in some of the linkage groups might also correspond to regions that are highly homozygous and/or subject to higher recombination frequency events [42]. The use of threshold LOD score of 6.0 produced a map with 21 genetic linkage groups, which correspond to the haploid number of chromosomes of the cashew genome (n=21). According to Pootakhamet al., [43] a linkage group consists of loci from the same chromosomes, and leads to a linkage group number similar to that of haploid number of chromosomes of a particular plant.

The total linkage map distance of 2,330.6 cM observed in this study was more than double the size of the map (1,050.7 cM) published earlier by Calvacanti and Wilkinson, [6] who used 11 SSR and 184 RFLP markers. This difference in size could be due to the different marker systems used in the two studies and also due to differences in the number of recombination events in the two maps as well as variations in the number and locations of mapped loci. In addition, the linkage map revealed in this study is denser as demonstrated by the relatively small average marker interval of 3.3 cM and high number of markers per linkage group (9 - 83) revealed in this study. This study has identified the first high density SNPs based genetic linkage map of cashew and is an important step for the detection of QTL associated with agronomically important traits. The linkage map and the highly polymorphic SNP markers identified in this study will also serve as a useful and practical tool not only for cashew breeders and the Cashew Research Programme in Tanzania but also for the cashew community worldwide.

Vol. 3, No. 02; 2018

ISSN: 2456-8643

ACKNOWLEDGEMENTS

The authors would like to acknowledge contributions made by Dr. Andrzej Kilian, Dr. Jason Carling and Colleen Hopper at DArT, Canberra, Australia for the development of SNPs markers. We would also like to thank Dr. Emmarold Mneney for establishing the F_2 mapping population. We wish to sincerely thank the Tanzania Commission for Science and Technology (COSTEC), the Cashew Research Program (CRP) and the Integrated Genotyping Service and Support (IGSS), Nairobi, Kenya for funding this work.

REFERENCE

Bezerra, M. A., de Lacerda, C. F., Filho, E. G., Carlos, E. B. and Prisco, J. T. (2008). Physiology of cashew plants grown under adverse conditions. *Brazilian Journal of Plant Physiology* 19: 449 – 461.

Aliyu, O. M. and Awopetu, J. A. (2007b). Assessment of genetic diversity in three populations of cashew (*Anacardium occidentale* L.) using protein-isozymes-electrophoretic analysis.*Genetic Resource Crop Evaluation*54: 1489 – 1497.

Aliyu, O. M. and Awopetu, J. A. (2007a). Multivariate analyses of cashew (*Anacardium occidentale* L.) germplasm in Nigeria. *Silvae Genetic* 56: 170 – 179.

Masawe, P. A. L. (2009). Modern agro-practices in Cashew. *Journal of Science, Technology and Management* 2: 102 – 114.

Masawe, P. A. L. (2012). Sustainable cashew production – challenges and Opportunities. In: *Proceeding of International Conference*. (Edited by Jnanadevan, R. *et al.*,)Panjim, Goa, India. pp. 36 – 42.

Calvacanti, J. V. and Wilkinson, M. J. (2007). The first genetic maps of cashew (*Anacadium occidentale*). *Euphytica* 157 131 – 143.

FAO (2011). Cashew production in Africa. [http://faostat3.fao.org/download/Q/QC/E] site visited on 20/5/2013.

Masawe, P. A. L. (2006). *Tanzanian Cashew Cultivars, Selected Clones*. (1st Edition), Cashew Research Programme, Naliendele Agriculture Research Institute, Nthara, Tanzania, 40pp.

Mohan, J. S. and Priyadarshan, P. M. (2009). *Breeding Plantation Tree Crops. Temperate Species.* Springer Science Business, New York, USA. 290pp.

Mneney, E. E. and Mantell, S. H. (2002). Clonal propagation of cashew (*Anacardium occidentale*. L.) by tissue culture. *Journal of Horticulture Science Biotechnology* 77(6): 649 – 657.

Croxford, A. E. (2003). A molecular study of the breeding system of cashew (*Anacadium occidentale*) in Tanzania. Thesis for Award of PhD Degree at University of Reading, UK,

Wittenberg, A. H. J., van der Lee, T., Cayla, C., Kilian, A., Visser, R. G. F. and Schouten, H. J. (2005). Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Molecular Genetic Genomics*274: 30 – 39.

Gupta, P. K., Mir, R. R., Mohan, A. and Kumar, J. (2008). Wheat genomics: present status and future prospects. *International Journal of Plant Genomics* 2: 1 – 36.

Myles, S. (2013). *Improving fruit and wine: what does genomics have to offer?Trends Genetic* 29(4):190 – 196.

Ru, S., Main, D., Evans, K. and Peace, C. (2015). *Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding.Tree Genet Genomes* 11: 1–12.

Bus, V. M., Esmenjaud, D., Buck, E. and Laurens, F. (2009). Application of genetic markers in rosaceous crops. *Genetics and Genomics of Rosaceae*6: 563–599.

Karaagac, E., Vargas, A., de Andrés, M., Carreño, I., Ibáñez, J. and Carreño, J. (2012). *Marker-assisted selection for seedlessness in table grape breeding*. *Tree Genet Genomes*8: 1003–1015.

Collard, B. C. Y. and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Biological Sciences* 363(1491): 557–572.

Varshney, R. K., Beier, U., Khlestkina, E., Kota, R., Korzun, V. and Röder, M. (2007). Single nucleotide polymorphisms in rye: discovery, frequency and applications for genome mapping and diversity studies. *Theory Applied Genetic*114: 1105 – 1116.

Boudehri, K., Bendahmane, A., Cardinet, G., Troadec, C., Moing, A., and Dirlewanger, E. (2009). *Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach*. *BioMed Central Plant Biology*9: 1 – 59.

Alpert, K. B. and Tanksley, S. D. (1996). *High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato*. *Proceeding National Academic Science* 93: 15503–15507.

De Franceschi, P., Stegmeir, T., Cabrera, A., van der Knaap, E., Rosyara, U. R. and Sebolt, A. M. (2013). *Cell number regulator genes in Prunus provide candidate genes for the control of fruit size in sweet and sour cherry*. *Molecular Breeding*32: 311–326.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J. and Harrington, S. (2013). *Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations.Theory Applied Genetic*126: 2699 – 2716.

Kulembeka, H. P. K. (2010). Genetic linkage mapping of field resistance to cassava brown streak disease in cassava (*manihotesculentacrantz*) landraces from Tanzania. Thesis for Award of PhD Degree at University of the Free State, South Africa,

Doyle, J. J. and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13 – 15.

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D. and Huttner, E. (2004). Diversity arrays technology (DArT) for whole genome profiling of barley. *Proceeding National Academic Science* 101(26): 9915 – 9920.

Van Ooijen, J. W. (2006). *Join Map 4. Software for the Calculation of Genetic Linkage Maps in Experimental Populations.*Kyazma BV, Wageningen, Netherlands. 59pp.

Kosambi, D. D. (1944). The estimation of map distance from recombination values. *Annual Eugene* 12: 172 – 175.

Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* 93 (1): 77 – 78.

Aliyu, O. M. (2014). Analysis of absolute nuclear DNA content reveals a small genome and intraspecific variation in cashew (*Anacardium occidentale L.*), Anacardiaceae. *Silvae Genetic* 63: 285 – 293.

Yang, S., Pang, W., Ash, G., Harper, J., Carling, J., Wenzl, P., Huttner, E., Zong, X. and Kilian, A. (2006). Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology.*Theory Applied Genetics* 113: 585 – 595.

Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E. (2006). High-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BioMed Central Genomics*7: 206 – 228.

Xia, L., Peng, K., Yang, S., Wenzl, P., de Vicente, M. C., Fregene, M. and Kilian, A. (2005) DArT for high-throughput genotyping of Cassava (*Manihotesculenta*) and its wild relatives. *Theory Applied Genetics*110: 1092 – 1098.

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L. and Yang, S. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theory Applied Genetics*113 (8): 1409 – 1420.

Prasad, S., Hendre, R. and Kamalakannan, M. V. (2012). High-throughput and parallel SNP discovery in selected candidate genes in Eucalyptus camaldulensis using Illumina NGS platform. *Plant Biotechnology Journal* 10: 646 – 656.

Ritter, E., Gebhardt, C. and Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* 125(3): 645 – 654.

Xian-Liang, S., Xue-Zhen, S. and Tian-Zhen, Z. (2006). Segregation distortion and its effect on genetic mapping in plants. *Chinese Journal of Agricultural Biotechnology* 3(3): 163 – 169.

Dettori, M. T., Quarta, R. and Verde, I. A. (2001). Peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. *Genome* 44(5): 783 – 790.

Barreneche, T., Bodenes, C., Lexerm, C., Trontin, J. F. and Fluch, S. A. (1998). Genetic linkage map of Quercusrobur L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, and isozyme and 5S r DNA markers. *Theory Applied Genetic* 97: 1090 – 1103.

Myburg, A. A., Griffin, A. R., Sederoff, R. R. and Whetten, R. W. (2003). Comparative genetic linkage maps of Eucalyptus grandis, Eucalyptus globulus and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theoretical and Applied Genetics* 107(6): 1028 – 1042.

Liu, K. and Muse, S. V. (2005). Power Marker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9): 2128 – 2129.

Castiglioni, P. P., Ajmone-Marsan, R., Van, W. and Motto, M. (1999). AFLP markers in a molecular linkage map of maize: Codominant scoring and linkage group distribution. *Theory Applied Genetic* 99(3): 425 – 431.

Pootakham, W., Jomchai, N., Ruang-Areerate, P., Shearman, J. R., Sonthirod, C. and Sangsrakru, D. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*105: 288 – 295.