DISTRIBUTION OF SINGLE NUCLEOTIDE POLYMORPHISM MARKERS TOWARDS TAGGING SOURCES OF RESISTANCE TO CASSAVA BROWN STREAK DISEASE IN CASSAVA

KAROLINE LEONARD SICHALWE

A DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CROP SCIENCE OF SOKOINE UNIVERSITY OF AGRICULTURE.

MOROGORO, TANZANIA.

2015

EXTENDED ABSTRACT

Cassava roots represent the future of food and income generation for over 800 million people in the world however, its production is threatened by virus disease; Cassava brown streak disease (CBSD). Biotechnology approaches are fast and powerful methodologies in cassava improvement and breeding. Construction of high-density and high quality genetic map in cassava would be of great benefit in combating CBSD. Conventional study was conducted in Naliendele, Kibaha and Dodoma for Tanzania and Molecular work performed at the International Livestock Research Institute, Kenya and University of Berkeley in United States of America. This study involved assessing the integrity of F_1 population from a cross between AR40-6 x Albert cassava cultivars using simple sequence repeat (SSR) polymorphisms. An F_1 population of 156 individuals were developed. Population evaluation resulted into 72% individuals as true F_1 hybrids, 18.7% were non hybrids and 8.2% were selfed individual plants. Evaluation of F_1 population validated SSR markers to be useful and efficient tools in identification of true F_1 hybrids in controlled crosses. On the other hand, the true F_1hybrids obtained were used to construct high dense SNP based linkage map using high throughput genotyping by sequencing (GBS) approach. The GBS is simple, low cost and de novo sequencing that makes an attractive option for large number of markers and individuals. Linkage analysis resulted into comprehensive genetic map with 19 linkage groups with a total of 4784 SNP markers: 2159 of these were mapped to the female genetic map, 2169 to the male map, and 3449 SNP markers to the integrated genetic map. Comprehensive genetic map encompassed 4250cM with mean distance of 1.26cM between the markers. This high density SNP based genetic linkage map of cassava can be used as base in locating genes controlling resistance to cassava brown streak disease and other genomic studies such as QTL detection, sequence assembly and genome comparison of the crop.

DECLARATION

I, Karoline L. Sichalwe, do hereby declare to the Senate of Sokoine University of Agriculture that this dissertation is my own original work done within the period of registration and that it has neither been submitted nor being concurrently submitted in any other institution.

_____ _____

Karoline Leonard Sichalwe Date

(MSc. Crop Science Candidate)

The above declaration is confirmed by:

_____ _____

Dr. Paul M. Kusolwa Date

(Supervisor)

COPYRIGHT

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of Almighty God. Thank you for keeping me strong and healthy throughout my academic endeavor. Thank you for giving me other people to whom I owe my deepest gratitude for their support.

I am grateful to the National Root and Tuber Research Program specifically Dr. Geoffrey Mkamilo, the National Root and Tuber coordinator for selecting me as one of project beneficiaries (student). My gratitude goes to Bill and Melinda Gates Foundation for funding the project 'Biotechnology application to combat cassava brown streak disease' from which my school fees and research studies was enabled.

I would like to sincerely thank my supervisor Dr. Paul Kusolwa for his guidance, patience and understanding during my studies. Despite his many other academic and professional commitments, he was able to provide his time for scientific advices, knowledge, many insightful discussions and suggestions.

I would like to thank Dr. Morag Ferguson of IITA for useful guidance, constructive critism, supervision and providing me an excellent atmosphere for doing my research. Thank you for sourcing my molecular laboratory work at BecA and your support in genetic linkage analysis.

I would like to acknowledge Ms. Esther Masumba and Fortunus Kapinga for assisting in planting and taking care of my mapping population. Thanks to Dr. Mkamilo and your team Julius and Gama for helping in genetic crosses at ARI Naliendele. Thanks to staffs at

ARI Kibaha in particular Mr. S. Tollano for taking care of my seedlings in the screen House. Thanks to Dr. Mrosso of ARI Makutupora for providing me with field site for establishing F_1 population.

I am also grateful to Arthur for his tireless assistance in laboratory work during genotyping of mapping population. Thank you for letting me experience laboratory hurdles in my life time. Thanks to IITA staff, Susan Karonga for administrative facilitation and Lorna Wanjiku for making me feel at home during my stay in Nairobi. I am also thankful to BecA laboratory staffs for their support. My research work would not have been possible without your help guys, May almighty God bless you all.

I would like to appreciate my parents who have been a source of inspiration to me in my life, to my lovely sisters for your genuine love and support throughout my studies. Thanks to my family members for encouraging me with their best wishes. Special thanks to my soul mate, my husband for your patience, practical and emotional support. I thank you for sticking by my side when I was depressed and lost faith. May almighty God bless you abundantly.

Lastly I would like to express my sincerely appreciations to my friend Salum Kasele for inspiring my effort despite enornemous work pressure we faced together and my friend Catherine for encouraging me pursue research in molecular breeding.

DEDICATION

This dissertation is dedicated to my wonderful husband, Anthony G. Tarimo for his constant support, encouragement, love and patience during my study period. The pursuit for my goals would not be possible without you. This work is also dedicated to my parents mother Alituhiga-Marry Kihaka and father Leonard Sichalwe who taught me to work hard for the things that I aspire to achieve, and my mother in law Esther Tarimo for her unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF PLATES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATION AND SYMBOLS

AFLP	Amplified Fragment Length Polymorphism
ARI	Agricultural Research Institute
BecA	Biosciences Eastern and Central Africa
Bp	Base pairs
CBB	Cassava bacterial blight
CBSD	Cassava brown streak disease
CBSV	Cassava Brown Streak Virus
CIAT	International Centre for Tropical Agriculture/Centro International de Agricultural Tropical
cM	centiMorgan
CMD	Cassava mosaic disease
CP	Cross-pollinated (out-breeder full-sib family)
DNA	Deoxyribonucleic acid
EST	Expressed sequence tags
FAO	Food and Agricultural Organization
GBS	Genotyping by sequencing
IFAD	International Fund for Agricultural Development
IITA	International Institute for Tropical Agriculture
Kg	Kilogram
LG	Linkage group
M	metre
MAS	Marker-assisted selection
Masl	Metre above sea level
Mb	Mega base pairs

Mg	milligram
$MgCl_2$	Magnesium chloride
min	Minute
ml	Mililitre
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
RAPD	Rapid amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
rpm	Revolution per minute
SDS	Sodium dodecyl sulphate
Sec	Seconds
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
t	Tonne
TBE	Tris Borate EDTA
TE	Tris-HCl EDTA
UCBSV	Uganda brown streak virus
ul	Microlitre
UV	Ultraviolet
V	Volt
w/v	Weight per volume

CHAPTER ONE

1.0 INTRODUCTION

Cassava (*Manihotesculenta* Crantz) is grown throughout tropical Africa, Asia and the Americas. It is one of the six staple crops in the world and forms the main source of calories for over 800 million people (Nassar and Ortiz, 2010). Cassava starch is a source of industrial raw materials and for bio-ethanol in Asia and constitutes a cash income for rural poor. Cassava production is affected mainly by abiotic factors, market conditions, socio-economic factors and biotic factors. Among the biotic factors, cassava mealy bug, cassava mosaic disease (CMD), cassava bacterial blight (CBB) and cassava brown streak disease (CBSD) cause major losses in yield and productivity resulting in the loss of billions of dollars (Waddington *et al*., 2010; Campo *et al*., 2011).

In East Africa, CBSD causes yield loss of up to 80% which translates to an annual monetary losses of up to 50 million dollars (FAO and IFAD 2005, IITA 2007). The disease is transmitted by a vector of whiteflies known as *Bemisia tabaci* Gennadius, family, Aleyrodidae (Maruthi *et al*., 2005). Discrete difference in geographical, biological and genetic features of this species makes them difficult to control and to find resistant varieties (Liu *et al*., 2007; Campo *et al*., 2011). Cassava is a highlyout-crossing monoecious species (mediated by protogyny) and suffers from inbreeding depression, making it difficult to develop appropriate stocks for classical genetics studies. The increased demand of cassava for industrial use needs high quality and reliable quantity of cassava. It is important to continue genetic improvement of cassava with higher productivity. Genetic markers are effective molecular tools for genomics study and development of new varieties in plant and animal breeding programs respectively (Collard *et al*., 2005). Here molecular markers are used for verification of the first filial population

and construction of genetic linkage maps which will be used for the identification of quantitative trait loci (QTL) specifically for resistance to CBSD.

Cassava (*Manihotesculenta* Crantz) is a root crop and a member of the family Euphorbiaceae. The crop was domesticated from the wild relative *Manihotesculenta* subspecies *flabellifolia* (Allem, 1994, 2002; Olsen, 2004), perhaps in several locations across the savannas of Brazil (Allem, 2002). The exact location(s) of cassava domestication is still being debated (Lebot, 2009). It was disseminated by the Portuguese during 16th and 17th centuries to tropical Africa, Asia and the Caribbean. The crop is a monoecious, perennial shrub possessing chromosomes (2n=36). Cassava plays an essential role in food security as it is main subsistence food crop in locations where food supply is constantly threatened by environmental constraints such as pests and diseases, drought, poor soils and increasing population growth. The crop is also considered as a famine relief crop when cereals fail due to prolonged drought and in areas where the environment has been affected by global climatic changes.

Cassava is grown for its enlarged starch-filled roots. Cassava fresh roots can produce 86% - 90% of cassava flour depending on variety (Kumoro *et al*., 2010). It is low in protein with 1.6g/100g. Roots can be eaten raw (sweet varieties), without processing, just peeled and either boiled, baked, or fried. In the coastal areas of Tanzania where cassava is also mostly grown, fresh roots are used to prepare a common food known as Iftar, prepared by boiling cassava fresh roots with coconut cream, commonly eaten when fasting in the Holy months of Ramadhan. The roots are also processed by grating, fermenting and then milling to make cassava flour. This is used to make different kinds of food such as porridge, stiff porridge (ugali), cakes, breads and other snacks. Young tender leaves are used as a cooked vegetable and contain vitamins and high levels of crude protein 29.3-

32.4% dry weight (Awoyinka *et al*., 1995). Cassava leaves are also used to produce cassava hay, and chips or pellets for animal feed. Cassava can be used as raw materials in agro-industries in the manufacture of adhesives, corrugated boards, gums, wallpaper, foundry, well drilling, paper industry, textile industry, wood furniture, particle board, biofuels, alcohol products, dusting powders, drugs, plastics, packaging, stain remover, concrete stabilizer and moisture sequester (Pattron, 2008).

In developing countries where poverty is common and severe, cassava can be grown in marginal areas with very unfavourable and fragile environments. Being the second most important staple crop after maize in these regions, yet far more resilient to adverse conditions, cassava was identified as the crop with the greatest potential to combat poverty as well as food and nutritional insecurity in Africa (African Agriculture, 2007). The current global cassava production is expected to be 282 million tons a year, of which more than 50% to occur in Africa, 33% in Asia and 15% in South and Central America (FAO, 2012). While the World average annual cassava consumption was around 19.9 kg/capita in 2012, Africa's annual per capita consumption is still above 130.9kg/year (FAO, 2012). In Africa total production was 141 million tonnes in 2011 with Nigeria being the largest producer, harvesting 52 million tonnes (FAOSTAT, 2011). Tanzania is the sixth largest producer of cassava in the world and the second producer of cassava in East Africa (Table1). The annual cassava production in Tanzania was between 4.5 and 6.0 million tons between 2009 and 2011 with about 761 000 hectares of land under cassava cultivation (FAO, 2011).

Table 1: Cassava production in eight African countries (million tons)

Country	2008	2009	2010	2011
Nigeria	46571	36822	42533	52403
Democratic Republic of Congo	15527	15054	15049	15569
Ghana	11858	12231	13504	14241
Angola	10506	12828	13858	14334
Mozambique	4236	5670	5700	6272
United Republic of Tanzania	5111	5179	4548	4647
Uganda	5298	5179	5282	4753
Malawi	3670	3823	4001	4259

Source: FAO 2011

The main cassava growing areas in Tanzania are the coastal strip along the Indian Ocean, around Lake Victoria, Lake Tanganyika and along the shores of Lake Nyasa. About 48.8 % of total production is from the coastal strip along the Indian Ocean (Mkamilo and Jeremiah, 2005). In Tanzania cassava is mainly a subsistent crop where 84% of the total production is utilized for food, making it second key food crop after maize in the order of importance. In general, about 16% of cassava is used for animal feed, alcohol brewing, starch making and for export. Cassava products have the potential to deliver significant economic benefits to rural poor and make substantial improvements to people's livelihoods.

1.1 Breeding and Genetic Improvement of Cassava

Although cassava is an important food crop, it has long been ignored by agricultural sectors until recently. Cassava genetic improvement (breeding) is challenging due to a high degree of genetic heterozygosity, variable flowering, low pollen fertility, self incompatibility, long-generation time and low fruit setting (Turyagyenda *et al.,* 2013). Conventional breeding has made progress in improving productivity, tolerance to biotic and abiotic stress and post-harvest physiological disorders (PPD) but with limited success (Ceballos *et al*., 2004). Due to the factors listed above, it takes at least 7 years to develop

an improved variety. Cassava Brown Streak is a cassava disease caused by Cassava brown streak virus (CBSV) and Uganda cassava brown streak virus (UCBSV), both virus species belong to the genus *Ipomovirus*, family *Potyviridae*. The major symptoms of CBSD occur in the leaves, stems and roots. However, these symptoms are inconsistent. The plant can have very severe symptoms in the leaves but no symptom in roots. Lack of knowledge and variability of symptoms and virus species has made breeding for CBSD resistance in cassava complicated. Molecular marker-assisted breeding (MAB) has the potential to increase the efficiency of breeding by reducing the number of plants that need to be phenotyped, increasing the accuracy of selection and hence reducing the time for varietal development and finally release. Molecular markers for root quality traits like dry matter content, protein and delayed postharvest physiological deteoration (PPD) as well as disease resistance have been identified at Centro International de Agricultural Tropical (CIAT) (Akano *et al*., 2002; Fregene *et al*., 2006). Identification of molecular markers tightly linked to CBSD resistance genes will be useful in breeding for CBSD resistance in cassava.

Cassava Brown Streak Disease and Cassava Mosaic Disease are two common virus diseases affecting the crop in East, Central and Southern Africa (ECSA). The disease has been known for many years, but is becoming increasingly damaging and posses a great threat to the livelihoods of millions of people who depend on cassava for food and income. Fields which are infected with CBSD have average yields of 3.5 t/ha in comparison with 10t/ha in fields without CBSD (Herron, 2009). The estimated economic loss due to CBSD is about US$130/ha, based on a sample in eastern Tanzania (Herron, 2009). The estimated annual yield loss for East Africa is as high as $202 million (Manyong *et al*., 2008). National Root and Tuber Programme have been evaluating many germplasm for response to CBSD resistance in their agricultural research stations using conventional procedures

which take around nine years to develop a new variety resistant to the disease. Molecular-marker assisted breeding approach will markedly reduce the time taken to develop potential resistant varieties, and significantly reduce population sizes during the evaluation and selection in breeding process.

For durable resistance, it is important that molecular markers are identified from many sources of CBSD resistance or tolerance. These can then be utilized for sustainable resistance to CBSD, by avoiding a single strong selection pressure. Breeding of vegetatively propagated crops require modified crossing hence the use of multiple parent stock for generating breeding populations. These populations might provide novel genetic recombination of chromosomes and eventually better resistance variations in cassava germplasm.

Vegetatively propagated crops are sensitive to selfing and thus each cross behaves independently and may provide differences in combining ability for agronomic trait of economic importance. A cassava genotype AR 40-6 has been evaluated in eastern zone of Tanzania for more than 10 years and observed to be resistant to the CBSD disease. This genotype was developed at CIAT and has 12.5%genomic contribution from wild species *M. esculenta* subsp. *flabellifolia*. Albert is cassava landrace originated from Mtwara, in Southern Tanzania. The genotype is very susceptible to CBSD. Genetic linkage maps are important for the purpose of gene discovery or allocating markers associated with quantitative trait loci (QTL) conferring CBSD resistance. The aim of this study was to assign and identify Single Nucleotide Polymorphic (SNPs) markers in a modified F_1 cross population of cassava cultivars AR40-6 x Albert for CBSD resistance.

1.2 Objectives

1.2.1 Overall objective

To assign and identify Single Nucleotide Polymorphic (SNPs) markers for CBSD resistance residing in a heterogeneous population of AR40-6 x Albert

1.2.2 Specific objectives

(i) To assess the integrity of simple sequence repeat (SSR) polymorphisms in an F1 population developed from a cross between AR40-6 x Albert.

(ii) To map SNP markers in heterogeneous F1 population derived from a cross between AR40-6 x Albert cassava cultivars

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Cassava Origin and Distribution

Manihotesculenta Crantz. ssp. *esculanta* is an important tropical root crop that provides food security and income generation for many farmers worldwide. Molecular studies carried out by using DNA markers SSR, RAPD, AFLP and isozyme marker G3phd, unravel *M. flabellifolia* to be the progenitor of cultivated cassava (Cabral *et al.,* 2000; Schaal and Olsen 2000; Olsen and Schaal 2001, Roa 2000; Colombo *et al.,* 1998). The new insights for initial domestication of cassava are likely to be in the savannas, the Brazilian Cerrado to the south of Amazon rainforest (Isendahl, 2011). Portuguese navigators in the 16th the century introduced cassava from Brazil to the west coast of Africa (Jones, 1969) and later to East Africa through Madagascar and Zanzibar (Hillocks *et al.,* 2002). The crop was not widely grown in East Africa until the late 18th or early 19th century when its value as food security crop was realized (Hillocks, 2002). Through rivers, land trade routes, farmer to farmer diffusion and mass migration, cassava spread to the interior of East African countries including Tanzania (Masumba, 2006). Today the crop is widely grown in the southern, eastern and Lake Victoria zones of Tanzania as a source of food and income.

Cassava has been ignored for years and only left as a food security reserve crop in time of poor harvest and drought, despite its high potential. As a result of this neglect, the productivity of cassava has remained low and has sometimes decreased over years. In Tanzania, average production of fresh cassava root is 10 t/ha which is below the continent's average yield of 12.5 t/ha and below Africa average yields of 15t/ha (FAO, 2011). This yield gap is caused by abiotic and biotic factors. Abiotic factors include;

weather vagaries, infertile soils, post-harvest root deteoration and cultural practices such as continued use of varieties with low genetic potential, lack of adequate planting materials and inadequate farming practices (Bull *et al.,* 2011; Campo *et al.,* 2011; and Kapinga *et al.,* 2005). Biotic factors which affect crop production in cassava include cassava green mites, cassava mealy bug, cassava bacterial blight (CBB), cassava mosaic disease (CMD) and cassava brown streak disease (CBSD). Cassava brown streak disease (CBSD) causes substantial losses in production through a loss in quality and quantity of the crop due to root necrosis in storage tissue.

2.2 Cassava Brown Streak Disease

Cassava brown streak disease is an important viral disease in the coastal lowlands of Eastern Africa (below 500m above sea level). The disease is also endemic in the coastal strip of Lake Malawi and Mozambique. Lately, CBSD has been reported at mid-altitude levels (1200-1500 m asl) in the Democratic Republic Congo (Mahungu *et al.,* 2003), Uganda, Mozambique, Zambia and Malawi (Alicai *et al.,* 2007; Campo *et al*., 2011) and the Lake zone areas of Tanzania (Jeremiah and Legg, 2008; Legg *et al.,* 2011) areas which were not reported to be at risk earlier. According to a UN agency, CBSD is on the verge of becoming an epidemic in countries of the Great Lakes if urgent measures are not taken. Surveillance conducted by Rwanda's National Agricultural Research Institute in 2010 showed a 15.7% rate of infection on local varieties and 36.9 % in improved varieties. Estimated production data from FAO (2005), revealed that about 350 000 MT of fresh cassava is lost annually to CBSD infection in coastal regions of Tanzania. Kanju (2003 and 2007) gave the estimated annual loss of US$ 16.5 million that occurred due to CBSD in the Tanzanian coastal lowland (price of fresh cassava is estimated at US$ 15 per tonne). The disease had caused a loss of 80% in East Africa and annual loss of $50 million for farmers in Tanzania (FAO and IFAD 2005; Campo *et al.,* 2011; IITA 2007). Cassava

Brown Streak Disease poses significant threat to food security on its own but become exaggerated when other agricultural infections such as root rot and bacterial diseases are involved. So far no cassava varieties currently being distributed to farmers seem to be resistance to CBSD.

There are two distinct virus species that were identified in East Africa to cause CBSD; *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak virus* (UCBSV). Both species belong to the family, *Potyviridae*: genus, *Ipomovirus* (Mbazibwa *et al.,* 2009). Another species has been identified in Mozambique as cassava brown streak Mozambique virus (Liu *et al.,* 2011, Winter *et al.,* 2010). There is evidence that the whitefly species *Bemisia tabaci* is the vector that transmits the cassava brown streak viruses at low rates (Maruthi *et al.,* 2005). Another whitefly species *Bemisia afer* is also mentioned as a possible vector of the disease (Mware *et al.,* 2009). In Tanzania *B. tabaci* is more abundant than *B. afer* and they occupy different parts of the cassava plant. *B. tabaci* adults are more abundant on the upper green leaves, while *B. afer* is more abundant on the lower semi-senescent leaves (Maruthi *et al*., 2004). Little is known on the variability of different isolates of Cassava Brown Streak Virus (CBSV). Additionally, sharing and distribution of infected planting materials is responsible for rapid spread of the disease especially in areas where the disease is re-emerging such as Uganda.

2.3 Characteristics and Symptoms of CBSD

A cassava brown streak virus affects the entire cassava plant; leaves, stems and the roots. The leaf symptom characteristics include irregular yellow blotches to foliar chlorosis which are associated with secondary veins (Plate 1). Foliar symptoms vary according to specific variety, age of the plant, growing conditions (temperature, rainfall, and altitude) and the virus isolate that cause symptoms. Resistant or tolerant varieties show foliar

symptoms with no or delayed root symptoms. At high temperature new leaves do not show symptoms. In older plants, CBSD symptoms are hard to recognize as leaf senesce is similar to the symptoms (Mohammed *et al.,* 2012). Lesions and dark brown streaks occur on stems (Plate 2). The diseased plants may show shoot tip death and progress to die-back. Symptoms on roots include yellow to brown necrosis on the starch bearing tissue (Plate 3) which is sometimes accompanied by root constriction (Alicai *et al.,* 2007, Mtunda 2002, Hillocks, 1997). Root constrictions and malformations are a common feature of severe CBSD. Roots which have symptoms become susceptible to secondary soil-borne pathogens such as soft rot.

Plate 1: Cassava Brown Streak Disease symptom on leaves of Albert cultivar

Plate 2: Cassava brown streak disease symptoms on stem

Plate 3: Symptoms of CBSD in roots showing brown necrosis ten months after planting: Left is the longitudinal cross-section and right cross-section cutting of the roots infected with CBSD

2.4 Molecular Markers and Cassava Genetic Improvement

A molecular marker is a characterisable protein, nucleic acid segment or section of chromosome that can be used to infer identity or the presence of a feature of interest that cannot be measured directly. Molecular markers are currently the most efficient tools for detection of genetic variations among populations and individual organisms. Applications include Marker Assisted Selection (MAS), identification and fingerprinting of genotypes and detection of quantitative trait loci (QTL). The use of marker-assisted selection (MAS) increases the precision of selection in plant populations. It also reduces the population size in the seedling stage, which allows breeders to work on the large number of genotypes simultaneously hence reducing time for varietal development (Ferguson *et al.,* 2011, Semagn *et al.,* 2006).

Three types of DNA markers exists; those based on the hybridization of short DNA probes to genomic DNA which is referred to as restriction fragment length polymorphism (RFLP) (Botstein *et al.,* 1980; Semagn *et al.,* 2006), secondly DNA markers that involve the amplification of short sections of DNA through the use of primers and the Polymerase chain reaction (PCR) examples being the Amplified fragment length polymorphism (AFLP) (Vos *et al.,* 1995), Random Amplified Polymorphism (RAPD) (Williams *et al.,* 1990) and Simple Sequence Repeat, (SSR) (Akkaya *et al.,* 1992) and lastly DNA markers based on sequence information which include (Single Nucleotide Polymorphism (SNP) (Jordan and Humphries, 1994), Expressed Sequence Tags (ESTs) and Diversity Array Technologies, DArTs (Jaccoud *et al.,* 2001). RFLP was used to assess the genetic variability in a traditional cassava *(M. esculenta* Crantz) and study of genetic diversity within the genus Manihot (Elias *et al.,* 2000; Marmey *et al.,* 1994). This technique is time consuming, labour intensive and requires large amounts of good quality DNA hence limits application of this marker to crop improvement (Yamamoto *et al*., 1994; Liu, 1998).

RAPD has been used to study phylogenetic relationship (Mignouna and Dixon, 1997; Schaal *et al*., 1997) and genetic linkage mapping (Michelmore *et al*., 1991; Fregene *et al.*, 1997).

RAPD markers have been widely used to determine genetic diversity (Marmey *et al*., 1994; Laminski *et al*., 1997; Colombo *et al*., 1998; Asante and Offei, 2003; Zacarias *et al*., 2004) and Quantitative trait loci (QTL). A QTL is a chromosomal region containing gene(s) responsible for variation of complex, or quantitative genetic traits in plants. The main disadvantage of this technique is lack of reproducibility and its dominance (each allele cannot be observed hence heterozygosity cannot be determined). Benesi *et al.* (2006) used AFLP to determine genetic relationship within Malawian cassava germplasm. Jorge (2000) also used AFLP markers in genetic mapping of Cassava Bacterial Blight (CBB) disease resistance in cassava.

2.4.1 Simple sequence repeat (SSR) markers

Simple Sequence Repeats are stretches of 2 to 6bp nucleotide units repeated in tandem. They are highly informative polymerase chain reaction (PCR)-based markers that detect length polymorphisms at loci with simple sequence repeats. Such length-polymorphisms can easily detect on high resolution gels by using a unique pair of flanking repeat through polymerase chain reaction. These markers are largely co-dominant, multi-allelic and evenly distributed all over the genome. SSR markers provide a powerful tool for genetic linkage map construction that can be applied for identification of quantitative trait loci (QTL). In cassava several groups have developed a few thousand SSR markers from expressed sequence tags (ESTs) and enriched genomic DNA libraries (Chavarriaga-Aguirre *et al.,* 1998; Mba *et al.,* 2001; Raji *et al.,* 2009; Sraphet *et al.,* 2011). The SSR markers have been used in different experiments to verify the integrity of the population as

used in peanuts (Gomez *et al.,* 2008), *Musa* species (Mbanjo *et al.,* 2012) and *Vigna radiata* (Khajudparn *et al.,* 2012). Microsatellites have also been used to assess the genetic diversity and degree of relationship between cassava and its wild relatives (Chavarriaga *et al.,* 1998). Despite the advantages above, SSR markers has high development cost and effort required obtaining working primers when large number of markers (more than 1000) is used unlike SNPs which occur in unlimited numbers and do not need primers.

2.4.2 Single nucleotide polymorphism markers

Single nucleotide polymorphisms, or SNPs, are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered. They occur in coding and non-coding regions of the genome. Single-nucleotide polymorphisms (SNPs) are the most frequent polymorphisms within and between individuals and populations and have been preferred recently as the genetic markers of choice for the study of complex genetic traits. These markers allow large quantities of DNA to be analyzed at a high rate of speed (high-throughput genotyping). The identification of high density of SNPs in cassava would facilitate progress in cassava genomics and breeding (Ferguson *et al.,* 2011). Recent studies of sequence variations in cassava identified 136 SNPs from EST sequence and 50 SNPs from bacterial artificial chromosome (BAC) end sequence with an average frequency of one SNP every 62 bp (Lopez *et al.,* 2005), Sakurai (2007) reported identification of 2356 SNPs when presented transcript of genes involved in stress response from full-length cassava cDNA library. Twenty-six SNPs were identified by Kawuki (2009) when studying sequence diversity in gene fragments of cassava. A total of 1190 SNP markers were identified and validated technically and biologically using 53 cassava varieties from Africa and America (Ferguson *et al.,* 2012). Single-nucleotide polymorphisms markers have been currently used to develop high-density genetic map by using more advanced genotyping technology (Genotyping by sequencing: GBS).

Genotyping by sequencing is a reduced-representation approach which uses enzyme-based complexity reduction to target a small portion of genome, coupled with DNA barcoded adapters to produce multiplex libraries of samples ready for next generation sequencing. The method is suitable for population studies, characterization of germplasm, breeding for crop genetic improvement and trait mapping in diverse organisms. GBS was first described for maize and barley (Elshire *et al.,* 2011) and for wheat and oats (Poland *et al.,* 2012). Elshire used GBS as an approach for high diversity species where roughly 200 000 maize sequence tags and 25 000 barley sequence tags were mapped using recombinant inbred populations. Large number of markers is obtained through GBS approach at reasonable price.

2.5 Genetic Linkage Maps

Molecular markers have immense by contributed to cassava breeding and genetics, in the assessment of genetic diversity, taxonomical studies, understanding the phylogenetic relationships in the genus, confirmation of ploidy and in the development of genetic maps (Lokko *et al*., 2004).

Genetics maps for yield and Cassava Mosaic Disease has been developed using molecular markers (Fregene *et al.,* 1997; Fregene 2001; Lokko *et al.,* 2004). A molecular genetic map of cassava is a basis for identifying molecular markers linked to other traits of interest such as starch, beta carotene, CMD and CBSD which can then be applied in cassava improvement. The first genetic linkage map of cassava was published from an intra-specific cross between TMS30572, an improved line from IITA, Ibadan, Nigeria and CM2177-2, an elite line from CIAT, Cali Colombia (Fregene *et al*., 1997). The map included 132 RFLP, 30 RAPD, 3 microsatellite, and 3 is oenzyme loci which defined 20 linkage groups spanning 931.6 cM, with an average marker density of 1 marker every 8

cM. Another linkage map was developed using SSR markers consisted of 100 markers assigned to 22 linkage groups (LG1 – LG22), each with 2 -8 markers, and a linkage group length varying from 9.7 cM (LG19) to 129.9 cM (LG3) (Fregene *et al*., 2006) The first SNP based map was published by Rabbi *et al.* (2012). Linkage maps were constructed using two approaches: one step approach whereby genetic map was constructed using the cross-pollinator (CP) option of Joinmap and two-step approach whereby individual parental maps were first developed and then merged into one integrated map. In one-step linkage map of 1837 cM containing 568 markers (435 SNPs and 133 SSRs) distributed across 19 linkage groups was constructed using cross-pollinator (CP) option of JoinMap. In two-step method, a map of 1541 cM and 483 markers (348 SNPs and 128 SSRs) distributed across 18 linkage group with an average map distance of 90 cM per group was developed. Other crops have developed genetic maps by using different types of markers. Examples being A SSR based genetic map for cultivated peanuts (Hong *et al.,* 2010), maternal and parental linkage map using 1125 AFLP markers in sweet potato (Kriegner *et al.,* 2000) and high- density genetic map for grape using next generation restriction-site associate DNA sequencing (Wang *et al.,* 2012).

High-throughput SNP markers were used to develop genetic linkage map for CBSD resistance using AR40-6 F1's population. Number of linkage groups was equivalent to haploid number of chromosomes in cassava that is eighteen linkage groups. However, practice more than or less than haploid number of chromosome is a common phenomena. Single Nucleotide Polymorphism genotyping by using advanced GBS approach offers high cost effective and efficiency method for high-throughput genotyping. This approach is useful for marker assisted genomics and assessment of genetic resources in wide range of species.

2.6 References

African Agriculture (2007). Nigerian cassava production booms but processing capacity lags. [http://africanagricultureblog.com] site visited on 12/12/2009.

Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. and Colvin, J. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant Disease* 91: 24 – 29.

Allem, A. C. (1994). The origin of *Manihotesculenta* Crantz (*Euphorbiaceae*). *Genetic Resources and Crop Evolution* 41: 133 – 150.

Allem, A. C. (2002). The origins and Taxonomy of Cassava. In: *Cassava biology, production and utilization.* (Edited by Hillocks, R. J., Thresh, J. M. and Bellotti, A. C). CABI Publishing, Wallingford, UK. pp.1 - 16.

Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E. and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theory Applied Genetics* 105: 521–525.

Akkaya, M. S., Bhagwat, A. A. and Cregan, P. B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 132: 1131-1139.

Asante, I. K. and Offei, S. K. (2003). RAPD-based genetic diversity study of fifty cassava (*Manihotesculenta* Crantz) genotypes. *Euphytica* 131: 113-119.

Awoyinka, A. F., Abegunda, V. O. and Adewusi, S. R. A. (1995). Nutrient content of young cassava leaves and assessment of their acceptance as a green vegetable in Nigeria. *Plant Foods for Human Nutrition*, 47: 21 – 28.

Benesi, I. R. M., Labuschagne, M. T., Dixon, A. G. O., Viljoen, C. D. and Mahungu, N. (2006). Genetic distance analysis of elite cassava (*Manihotesculenta* Crantz) genotypes in Malawi using morphological and AFLP marker technique. *South African Journal of Plant and Soil* 23: 58 – 61.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314 – 331.

Bull, S. E., Ndunguru, J., Gruissem, W., Beeching, J. R. and Vanderschuren, H. (2011). Cassava: constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Reproduction* 30: 779–787.

Cabral, G. B., Carvalho, L. J. C. B. and Schaal, B. A. (2000). Relationship analysis of closely related species to cassava (*Manihot esculenta* Crantz) based on microsatellite-primed PCR. In: *Proceedings of the Fourth International Scientific Meeting of the Cassava Biotechnology Network*. (Edited by Carvalho, L. J. C. B., Thro, A. M. and Vilarinhos, A. D.) Embrapa Recursos Geneticos Biotechnologia /CBN, Brazil pp. 36 - 50.

Campo, B. V. H., Hyman, G. and Bellotti, A. (2011). Threats to cassava productions: known and potential geographical distribution of four key biotic constraints, *Journal Article* 3: 329 – 345.

Ceballos, H., Iglesias, C. A., Perez, J. C. and Dixon, G. O. (2004). Cassava breeding opportunities and challenges. *Plant Molecular Biology* 56: 503 – 516.

Chavarriaga-Aguirre, P., Maya, M. M., Bonierbale, M. W., Kresovich, S., Fregene, M. A., Tohme, J. and Kochert, G. (1998). Microsatellites in cassava (*Manihotesculenta* Crantz): discovery, inheritance and variability. *Theory Applied Genetics* 97: 493 – 501.

Colombo, C., Second, G., Valle, T. L. and Charrier, A. (1998). Genetic diversity characterization of cassava cultivars (*Manihotesculenta* Crantz). RAPD markers. *Genetic Molecular Biology* 21: 105 – 113.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica.* 142: 169–196.

Elias, M., Panaud, O. and Robert, T. (2000). Assessment of genetic variability in a traditional cassava (*Manihotesculenta* Crantz) farming system, using AFLP markers. *Heredity* 85: 219-230.

Elshire, R., Glaubitz, J., Sun, Q., Poland, J. and Kawamoto, K. (2011). A robust simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6 (5): 19 - 21.

FAO (2012). Food Outlook, Global Market Analysis ISSN 0251-1959. 22pp.

FAOSTAT (2011). Food and agriculture organizations statistics database. FAO, Rome [http://faostat3.fao.org/home/index.html] site visited on 5/15/2013.

FAO and IFAD (2005). Proceedings of the validation forum on the Global Cassava Development Strategy Roma. FAO 17pp.

Ferguson, M. E., Hearne, S. J., Close, T. J., Wanamaker, S., Moskal, W. A., Town, C. D., De Young, J., Marri, P. R., Rabbi, I. Y. and De Villiers, E. P. (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theory Applied Genetics.* 124: 685 - 695.

Ferguson, M., Rabbi, I., Kim, D. J., Gedil, M., Lopez-Lavalle, L. A. B. and Okogbenin, E. (2011). Molecular markers and their application to cassava breeding: Past, present and future. *Tropical Plant Biology*. 5: 95 – 109.

Fregene, M., Morante, N., Sanchez, T., Marin, J., Ospina, C., Barrera, E., Gutierrez, J., Guerrero, J., Bellotti, A., Santos, L., Alzate, A., Moreno, S. and Ceballos, H. (2006). Molecular markers for the introgression of useful traits from wild *Manihot* relatives of cassava: Marker-assisted selection of disease and root quality traits. *Journal of Root Crops* 32: 1 – 31.

Fregene, M., Okogbenin, E., Mba, C., Angel, F., Maria, C., Suarez, G. J., Chavarriaga, P., Roca, W. and Bonierbale, M. J. (2001). Genome mapping in cassava improvement: Challenges, achievements and opportunities. *Euphytica* 120: 159 – 165.

Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriagga, P., Roca, W., Tohme, J. and Bonierbale, M. (1997). A molecular genetic map of cassava (*Manihotesculenta Crantz*).*Theoretical and Applied Genetics* 95: 431 – 441.

Gomez, S. M., Denwar, N. N., Ramasubramanian, T., Simpson, C. E., Brow, G., Burkes, J. J., Puppala, N. and Burow, M. D. (2008). Identification of peanut hybrids using microsatellite markers and horizontal polyacrylamide gel electrophoresi *Peanut Science* 35: 123 – 129.

Herron, C. (2009). CBSD: Enemy number 1, IITA, research for Development Review, Issue 2. [http://r4dreview.org/2009/03/cbsd-enemy-number-1/] site visited on 9/8/2013.

Hillocks, R. J., Thresh, J. M., Tomas, J., Botao, M., Macia, R. and Zavier, R. (2002). Cassava brown streak disease in northern Mozambique. *International Journal of Pest Management* 48: 178 – 181.

Hillocks, R. J., Thresh, J. M. and Bellotti, A. (2002). *Cassava Biology, Production and Utilization*. CABI Publishing, Wallingford, UK. 332pp.

Hillocks, R. J. (1997). Cassava virus diseases and their control with special reference to Southern Tanzania. Integrated Pest Management Reviews 2: 125-138. [http://www.livestrong.com/article/463568-nutrition-in-cassava-flour] site visited on 4/16/2013.

Hong, Y., Chen, X., Liang, X., Liu, H., Zhou, G., Li, S., Wen, S., Holbrook, C. C. and Guo, B. (2010). A SSR-based composite genetic linkage map for the cultivated peanut (*Arachishypogaea* L.) genome BMC Plant Biology DOI: 10.1186/1471-2229-10-17 [http://www.biomedcentral.com] site visited on 10/4/2014.

IITA, (2007). Solution Found for Cassava Root-rot Devastation in Africa. IITA News, Dar es Salaam, Tanzania. 23pp.

Isendahl, C. (2011). The domestication and early spread of Manioc *(Manihotesculenta* Crantz: A brief synthesis *Latin America antiquity* 22(4): 452 – 468.

Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29(4): 1 - 7.

Jeremiah, S. C. and Legg, J. P. (2008). Cassava brown streak virus disease: Farmers' 419 perspectives on a new outbreak of this disease from the Lake zone of Tanzania in [www.youtube.com/watch?v=CJdws9CnUw] site visited on 15/8/2012.

Jones, W. O. (1969). Manioc in Africa. Stanford University Press. Palo Alto, California. 315 pp.

Jordan, S. A. and Humphries, P. (1994). Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35. *Human Molecular Genetics* 3: 1915 - 1925.

Jorge, V., Fregene, M. A., Duque, M. C., Bonierbale, M. W., Tohme, J. and Verdier, V. (2000). Genetic mapping of resistance to bacterial blight disease in cassava (*Manihotesculenta* Crantz). *Theoretical and Applied Genetics* 101: 865 – 872.

Kanju, E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. and Dixon, A. (2007). Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmers and consumer preferences. In: *Tropical root and tuber crops: Opportunities for poverty alleviation and sustainable livelihoods in developing countries*. (Edited by Kapinga, R., Kingamkono, R., Msabaha, M., Ndunguru, J., Lenmaga, B. and Tusiime, G.) Proceedings of the thirteenth triennial symposium of the International society for tropical root crops (ISTRC), 10-14 November 2003, Arusha Tanzania. pp. 95 - 101.

Kanju, E. N., Mahungu, A. G., Dixon, O. and Whyte, J. (2003). Is resistance/tolerance to cassava brown streak disease associated with the zigzag stem trait? *Roots* 8: 15 - 19.

Kapinga, R., Mafuru, J., Jeremiah, S., Rwiza, E., Kamala, R., Mashamba, F. and Mlingi, N. (2005). *A Review of Cassava in Africa with Country Case Studies on Nigeria, Ghana, the United Republic of Tanzania, Uganda and Benin.* Status of Cassava in Tanzania. FAO and IFAD, Rome, Italy. 66pp.

Kawuki, R., Ferguson, M., Labuschagne, M., Herselman, L. and Kim, D. J. (2009). Identification, characterization and application of single nucleotide polymorphism for diversity assessment in cassava (*Manihotesculenta* Crantz) *Molecular Breeding* 23: 669 – 684.

Khajudparn, P., Prajongjai, T., Poolsawat, O. and Tantasawat, O. P. (2012). Application of ISSR markers for verification of F_1 hybrids in mungbean (*Vignaradiata*).*Genetics Molecular Research* 11(3): 3329 – 3338.

Kriegner, A., Cervantes, J. C., Burg, K., Mwanga, R. O. and Zhang, D. P. (2000). *A Genetic Linkage Map of Sweetpotato* (*Ipomoea batatas* (L.) Lam.) Based on AFLP Markers, CIP Programme. 313pp.

Kumoro, A. C., Retnowati, D. S. and Budiyati, C. S. (2010). Microwave assisted synthesis and characterization of acetate derivative cassava starch. *American Journal of Food Technology* 5: 100 - 110.

Laminski, S., Robinson, E. R. and Gray, V. M. (1997). Application of molecular markers to describe South Africa cassava cultivars. African Journal of Root Tuber Crops: 132 - 134.

Lebot, V. (2009). *Tropical Root and Tuber Crops:* Cassava, Sweet potato, Yams and Aroids, crop production science in horticulture series. CAB International. 85pp.

Legg, J. P., Jeremiah, S. C., Obiero, H. M., Maruthi, M. N., Ndyetabula, I., Okao-Okuja, G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., Mkamilo, G., Alicai, T. and Lava Kumar, P. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research.* 159(2): 161-170.

Liu, J., Zheng, Q., Gadidasu, K. K. and Zhang, P. (2011). Cassava genetic transformation and its application in breeding. *Journal Integrated Plant Biology* 53(7): 552 – 569.

Liu, S. S., De Barro, P. J., Xu, J., Luan, J. B., Zang, L. S. and Ruan, Y. M. (2007). Asymmetric mating interactions drive widespread invasion and displacement in a whitefly. *Science* 318(5857): 1769 – 1772.

Liu, B. H. (1998). Statistical genomics: linkage, mapping, and QTL analysis. CRC Press, LLC. 12pp.

Lokko, Y., Danquah, E. Y., Offei, S. K., Dixon, A. G. O. and Gedil, M. A. (2004). Molecular markers associated with a new source of resistance to the cassava mosaic disease. *African Journal of Biotechnology* 4(9): 873 – 881.

López, C., Piegu, B., Cooke, R., Delseny, M., Tohme, J. and Verdier, V. (2005). Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihotesculenta* Crantz). *Theoretical and Applied Genetics* 110: 425 – 431.

Mahungu, N. M., Bidiaka, M., Tata, H., Lu-Kombo, S. and N'luta, S. (2003). Cassava brown streak-like symptoms in Democratic Republic of Congo. *Roots* 8: 8 – 9.

Manyong, V. M., Kanju, E. E., Mkamilo, G., Saleh, H. and Rweyendela, V. J. (2008). *Baseline Study on Livelihood Status and Technology Adoption levels in Cassava Brown Streak Disease Infected Areas of Eastern Tanzania and Zanzibar*. Technical Report. IITA, Dar es Salaam, Tanzania. 37pp.

Marmey, P., Beeching, J. R., Hammon, S. and Charrier, A. (1994). Evaluation of cassava (*Manihotesculenta* Crantz) germplasm using RAPD markers. *Euphytica* 74: 203 – 209.

Maruthi, M. N., Hillocks, R. J., Mtunda, K., Raya, M. D., Muhanna, M., Kiozia, H., Rekha, A. R., Colvin, J. and Thresh, J. M. (2005). Transmission of cassava brown streak virus by Bemisiatabaci. *Journal of Phytopathology* 153: 307 – 312.

Maruthi, M. N., Hillocks, R. J., Rekha, A. R. and Colvin, J. (2004). *Transmission of Cassava Brown Streak Virus by Whiteflies*. In: Sixth international scientific meeting of the cassava biotechnology network adding value to a small farmer crop. CIAT, Cali, Colombia. 80pp.

Masumba, E. A. (2006). Genetic diversity and field performance of cassava (*Manihotesculenta* Crantz) landraces commonly grown in eastern southern and Lake Zones. Dissertation for Award of MSc Degree at Sokoine University of Agriculture, Morogoro, Tanzania, 89pp.

Mba, R. E. C., Stephenson, P., Edwards, K., Melzer, S., Mkumbira, J., Gullberg, U., Apel, K., Gale, M., Tohme, J. and Fregene, M. (2001). Simple sequence repeat (SSR) markers survey of the cassava (*Manihotesculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theory Applied Genetics* 102: 21–31.

Mbanjo, E. G. N., Tchoumbougnang, F., Mouelle, A. F., Oben, J. E., Nyine, M., Dochez, C., Ferguson, M. E. and Lorenzen, J. (2012). Development of expressed sequence tags-simple sequence repeats (EST-SSRs) for Musa and their applicability in authentication of a Musa breeding population. *African Journal Biotechnology* 11: 13546 - 13559.

Mbanzibwa, D. R., Tian, Y. P., Mukasa, S. B. and Valkonen, J. P. (2009). Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology* 83: 6934 – 6940.

Michelmore, R. W., Paran, I. and Kesseli, R. V. (1991). Identification of markers linked to disease-resistant genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl Acad. *Science*. 88: 9828-9832.

Mignouna, H. D. and Dixon, A. G. O. (1997). Genetic relationships among cassava clones with varying levels of resistance to African mosaic disease using RAPD markers. *African Journal Roots Tuber Crops* 2: 28–32.

Mkamilo, G. S. and Jeremiah, S. C. (2005). Current status of cassava improvement programme in Tanzania. *African Crop Science* 7: 1311 – 1314.

Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. and Maruthi, N. M. (2012). The Symptom and Genetic Diversity of Cassava Brown Streak Viruses Infecting Cassava in East Africa. Hindawi Publishing Corporation. 10pp.

Mtunda, K. and Muhanna, H. (2002). *Report on a Survey of Cassava Disease in Tanzania*. Ministry of Agriculture and Food Security, Dar es Salaam, Tanzania. 17pp.

Mware, B. O., Narla, R., Amata, R., Olubayo, F., Songa, J. and Ateka, E. M. (2009). Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *Journal of General and Molecular Virology* 1(4): 040-045.

Nassar, N. and Ortiz, R. (2010). *Breeding Cassava to Feed the Poor.* Scientific American, USA. 83pp.

Olsen, K. M. and Schaal, B. A. (2001). Microsatellite variation in cassava and its wild relatives: further evidence for a southern Amazonian origin of domestication. *American Journal Botanic* 88(1): 131 – 142.

Olsen, K. M. (2004). SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology* 56: 517 – 526.

Pattron, D. D. (2008). Non food uses cassava.[httpc//goarticles.com/article/Non-food-uses-of-cassava] site visited on 27/6/2012.

Poland, J. A., Brown, P. J., Sorrells, M. E. and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7(2): 1 - 8.

Rabbi, I. Y., Kulembeka, H. P., Masumba, E., Marri, P. R. and Ferguson, M. (2012). An EST-derived SNP and SSR genetic map of cassava (*Manihotesculenta* Crantz), *Theory Applied Genetics* 125(2): 329 – 342.

Raji, A. A., Anderson, J. V., Kolade, O. A., Ugwu, C. D., Dixon, A. G. and Ingelbrecht, I. L. (2009). Gene-based microsatellites for cassava (*Manihotesculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant Biology* 9: 118 – 120.

Roa, A. C., Chavarriaga-Aguirre, P., Duque, M. C., Maya, M. M., Bonierbale, M. W., Iglesias, C. and Tohme, J. (2000). Cros - species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites. *American Journal Botanic* 87 (11): 1647 – 1655.

Schaal, B. A. and Olsen, K. M. (2000). Gene genealogies and population variation in plants. *Academy of Sciences* 97: 7024 – 7029.

Schaal, B., Carvalho, L. C. C. B., Prinzie, T., Olsen, K., Hernandez, M., Cabral, G. and Moeller, D. (1997). Phylogenetic relationships and genetic diversity in *Manihot* species. In: *Proceedings of the Third International Scientific Meeting, Cassava Biotechnology Network* (Thro, A.M. and Akoroda M. O. eds.). Kampala, Uganda, 26-30 August, 1996. *Africa Journal Root Tuber Crops*. *2*: 147-149.

Sakurai, T., Plata, G., Rodriguez-Zapata, F., Seki, M., Salcado, A., Ishiwata, A., Tohme, J., Sakaki, Y., Shinozaki, K. and Ishitani, M. (2007). Sequence analysis of 20 000 full length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology* 7: 66 – 69.

Semagn, K., Bjørnstad, A. and Ndjiondjo, M. N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology* 25: 2540 – 2569.

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., Shirasawa, K., Isobe, S., Lightfoot, D. A., Tangphatsornruang, S. and Triwitayakorn, K. (2011). SSR and EST-SSR based genetic linkage map of cassava (*Manihotesculenta* Crantz). *Theory Applied Genetics* 122: 1161– 1170.

Turyagyenda, L. F., Kizito, E. B., Ferguson, M., Baguma, Y., Agaba, M., Harvey, J. J. W. and Osiru, D. S. O. (2013). Physiological and molecular characterization of drought responses and identification of candidate tolerance genes in cassava. AoB Plants 5: 1 – 17.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijers, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407 – 4414.

Waddington, S. R., Li, X. Y., Dixon, J., Hyman, G. and de Vicente, M. C. (2010). Getting the focus right: production constraints for six major food crops in Asian and African farming systems. *Food Security* 2(1): 27 – 48.

Wang, N., Fang, L., Xin, H., Wang, L. and Li, S. (2012). Construction of a high-density genetic map for grape using next generation restriction-site associate DNA sequencing. *BMC Genomics* 12: 143 – 148.

Winter, S., Koebler, M., Stein, B., Pietruszka, A., Paape, M. and Butgereitt, A. (2010). The analysis of cassava brown streak viruses reveals the presence of a distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology* 91(477): 365 – 376.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingley, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531 – 6535.

Yamamoto, T., Nishikawa, A. and Oeda, K. (1994). DNA polymorphism in *Oryza sativa* L. amplified by arbitrary primer PCR. *Euphytica* 78: 143 – 148.

Zacarias, A. M., Botha, A. M., Labuschagne, M. T. and Benesi, I. R. M. (2004). Characterization and genetic distance analysis of cassava (*Manihotesculenta* Crantz) germplasm from Mozambique using RAPD fingerprinting. *Euphyti ca* 138: 49 – 53.

CHAPTER THREE

Integrity of SSR Markers in a modified F1 population from a cross between AR40-6 x Albert cassava cultivars

3.0 Abstract

Genetic improvement of cassava requires the identification of proper identity of true F_1 hybrids from controlled crosses before further evaluation and selection, a concern that has been an important challenge in many cassava breeding programs. In this study, objective was to authenticate between expected parental recombination and non-parental recombination. An F_1 population was developed from a cross between genotype AR40-6 x Albert. Out of 350 seeds collected and sowed, only 219 seedlings grew to its full potential. Leaf samples were collected and DNA extracted from 219 F_1seedlingsusing modified Dellaporta method. Thirty eight SSR markers were used to genotype F_1 population and ten markers was found to be polymorphic to female and male parent. Comparison of resulting banding patterns to those of the parents was done and 72% F_1individuals were identified as true hybrids of their parents due to presence of marker alleles from both female and male parents. The results also revealed that 18.7% of the individuals were non hybrids and 8.2% were selfed individual plants. Application of SSR markers have been useful and efficient tools in identification of true F_1 hybrids in controlled crosses thereby enhancing the effectiveness of our breeding programs. The F_1population obtained in this study was further used for construction of genetic linkage map by using SNP markers.

3.1 Introduction

Cassava (*Manihotesculenta)* is a source of food security and income generation for millions of people in developing countries of Africa, South and Central America and Southeast Asia. The crop can grow in poor soil, can withstand drought and is an important

famine reserve crop in areas with prolonged periods of drought and unreliable rainfall. Cassava roots are rich in carbohydrates, mainly starch, and are a major source of energy (Charles *et al.,* 2005). Cassava leaves contain more protein and are used for human consumption and animal feed. According to the Food and Agriculture Organizations' Global Cassava Development Strategy, cassava is the third most important source of calories in the tropics, after rice and maize (FAO 2004). The crop is vital for both food security and income generation and grown by poor farmers, many of them women, often on marginal and fragile land.

Production of cassava is affected by abiotic and biotic factors. Abiotic factors include infertile soils, poor agricultural practices and use of infected planting materials. Biotic factors affecting cassava production include pests and disease such as cassava mealy-bug, cassava green mite, cassava bacterial blight (CBB), cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Bull *et al*., 2011; Campo *et al.,* 2011; Kapinga *et al.,* 2005). CBSD is currently a major devastating disease which causes loss of quantity and quality of cassava roots. The disease used to be restricted to low elevations 1000m below sea level but current research reports revealed that, the disease now occurs in highland areas 1000m above sea level (Alicai *et al.,* 2007). CBSD causes yield losses of up to 100% in the worst affected cassava fields. Symptoms of the disease include yellow cholosis in the mature parts of leaves, brown streaks on the green portion of stems and cracking on the brown portion of stems. In severe cases leaves dry, shoot tip die and progress to stem die-back (Ntawahurungu *et al.,* 2007). Storage roots of diseased plants show brown necrosis in the starchy part and sometimes root constrictions. Cassava brown streak disease is caused by *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak virus* (UCBSV), both species belonging to family, *Potyviridae*: genus, *Ipomovirus* (Mbazibwa *et al.,* 2009).

Cassava is a monoecious and predominantly out-crossing plant with the female flowers open before the male ones on the same branch. Self-pollination can occur when male and female flowers on different branches or on different plants of the same genotype open simultaneously (Chavarriaga-Aguirre and Halsey, 2005). Flowering time depends on the genotype and environmental conditions. Cassava flower best at moderate temperature (24^0C) and flowering is promoted by long days (photoperiods range between 11.2h – 14.8h) depending on genotypes (Alves, 2002). Scarcity of flowers may sometimes hinder crossing of elite clones. Controlled crossing involves bagging of female flowers before they open and pollinating using specific pollens. This minimizes effect of seed contamination hence minimizing variations in population of F_1 individuals. Variation in F_1 can also be caused by variations in selected parents due to genetical difference in the same parental genotype. For these reasons we cannot depend on the controlled crossing to completely eliminate cross contamination and thus produce pure hybrid population. Confirmation of genotype identity is important for breeders and geneticists in attaining breeding goals (Khasa *et al.,* 2003; Evans *et al.,* 2011). Morphological characteristics are not adequate for identifying the integrity of pure genotypes as they possess limited polymorphism, which varies with the environment (Asif *et al.,* 2006; Lin *et al.,* 2010; Al-Doss *et al.,* 2011). DNA markers are generally highly polymorphic and are not influenced by the environment and thus are useful in determining parentage-offspring relationships as well as validation of genotype identity (Fen *et al.,* 2008).

Application of molecular markers allows rapid identification of plant genotypes with high efficiency and low labour cost (Reddy *et al.,* 2002; Takrama *et al.,* 2005; Gomez *et al.,* 2008). They provide a robust, rapid and effective means to differentiate even closely related genotypes. Simple sequence repeats (SSRs) markers require small quantities of genomic DNA and are amenable to automation through the use of liquid handling devices

and capillary electrophoresis (Powell *et al.,* 1996; McCouch *et al.,* 1997). These markers are highly polymorphic, reproducible, multiallelic, easy assayed, locus specific, co-dominant, abundant and randomly distributed throughout the genome. These characteristics made them the most popular markers, before next-generation sequencing technologies made SNP genotyping more accessible (Mbanjo *et al.,* 2012; Rabello *et al.,* 2005). SSR markers has been used for evaluating genetic variability in a broad range of living organism (Li *et al.,* 2007; Kawka *et al.,* 2007; Chan *et al.,* 2008; Banhos *et al.,* 2008; Arif *et al.,* 2010).

The objective of this study was to assess the integrity of SSR markers in a modified F_1 population developed from a cross between AR40-6 x Albert and to detect possible non hybrids in the population. The identification of true to type genotypes is a first step toward developing a reliable genetic linkage map.

3.2 Material and Methods

3.2.1 Plant materials

Cassava genotype 'AR40-6' was used as a female parent and 'Albert' as a male parent (pollen donor) to generate an F_1 population via intra-specific crossing. The genotype 'AR40-6' was developed at Centro Internacional de Agricultura Tropical (CIAT) in Colombia and has 12.5% genomic contribution from wild species *M. esculenta* subsp. *flabellifolia* and 50% from Cassava Mosaic Disease (CMD) resistant *M. esculenta* variety (C-39) evaluated by International Institute for Tropical Agriculture (IITA). Genotype AR 40-6 has been evaluated in eastern zone of Tanzania for more than ten years and observed to be resistant to cassava brown streak disease (CBSD). Albert is a local landrace originating from Mtwara, Tanzania and is susceptible to CBSD. Both genotypes are resistant to Cassava Green Mite, Cassava Bacterial Blight and CMD. Planting materials

for genotype AR40-6 were collected from the research station while cuttings for Albert were collected from different farmer's field in the nearby Naliendele villages.

The crossing block was planted at Naliendele Agriculture Research Institute in southern Tanzania. Complete randomized block design was used to randomize two genotypes arranged in alternative rows of 10.0m long and 1.0m apart. Mature female flowers were closed in the morning by white cloth bags to prevent pollination from insects. Hand pollination was done in the afternoon by rubbing anthers with pollen onto stigmas of bagged flowers. Fertilized flowers were covered with a mesh cloth to prevent extraneous pollen contamination and to collect dehiscent mature seeds. Seeds were collected and sown as F1 progenies in a screen house at Sugarcane Research Institute (SRI) Kibaha. A mixture of forest soil and sand in a ratio of 1:3 respectively was placed in seed trays before direct sowing of seeds was done. After sowing, seeds were watered and sprayed with and insecticide. After three months the seedlings were transplanted into the field at Makutupora Research Station where there is low CBSD pressure. Spacing of 1.0m between rows and 1.0m within rows was used with 10 plants in each line. All cultural practices (including weed control, fertilizer application and irrigation) were carried out to ensure good crop growth performance. An F1 population in this study was referred as AA population followed by number of individual plant eg. AA100.

3.2.2 Genomic DNA extraction

Young leaves were collected and stored at -80^{0}C before DNA isolation. A total of 219 leaf samples were used to extract DNA. The protocol described by Dellaporta (1983) was used for DNA extraction. After extraction, the DNA was stored at -20^{0}C for long term use and 4^{0}C before genotyping.

3.2.3 Quantification and qualification deoxyribonucleic acid

The DNA quantity and quality were determined using NanoDropTMND-1000 Spectrophotometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis respectively. A volume of 1.5 µl of each DNA samples was used to measure absorbance of molecules samples at certain wavelength. Spectrophotometer was used to measure purity of DNA at the ratio of absorbance 260 nm and 280 nm and concentration at nanogram per microlitre (ng/µl). The quality of DNA was assessed by loading 2ul of each sample and running the samples through a 0.8% (w/v) agarose gel in 1XTBE buffer at 120V for 45minutes. TBE electrophoresis buffer was diluted from 10XTBE electrophoresis buffer which made up by mixing Tris Base 108g, Boric Acid 55g, 0.5M EDTA 20ml and distilled water. The gel purity was visualized under Ultraviolet (UV) illumination light using special installed software.

3.2.4 Polymerase chain reaction (PCR)

3.2.4.1 Primers and parental screening

SSRs marker used in this study (Table 2) were developed at CIAT (Chavariagga-Acquire *et al.,* 1998; Mba *et al.,* 2001). A total of 38 primer pairs were selected for screening genomic DNA of parents AR40-6 and Albert. Primers which were found to be polymorphic in the parents were used to genotype the F_1 hybrids.

3.2.4.2 Genotyping

A total of 219 F1 progenies of cassava from a cross between AR40-6 and Albert were genotyped using SSR markers. Genotyping involved SSR marker amplification by polymerase chain Reaction (PCR) and gel electrophoresis. Ten informative markers, which showed size polymorphism between the parents, were selected to genotype the F_1hybrids. Polymorphic markers either had a single allele but different sizes in two parents

or heterozygous allele in one parent but homozygous in the other parent. Amplification of genomic DNA was carried out using a Gene Amp PCR System 9700 Base Module (*Applied Biosystems Inc*) PCR machine. The total volume of PCR reaction was 10µl per reaction. These 10µl contains 10ng of genomic DNA, 1.0µl of 10X PCR reaction buffers, 2Mm $MgCl_2$, 0.2mM dNTPs, 0.38U of Taq polymerase and 0.08pmol of forward and reverse primer (Table 2). Polymerase Chain Reactions conditions were accomplished by initial denaturation for 3min at 95^0C then 30 cycles of denaturing at 95^0C at 30s, annealing at between 55^0C to 57^0C for 1min, initial extension of 72^0C for 1minute and final extension of 72^0C for 30min.

The PCR products were electrophoresed on 2% agarose for 30min with 120V in 1 x TBE and later visualized under UV light. Gel Photographs were scanned through Gel Doc System (*Syngene*). Amplification products were then subjected to the sequencer for observation of allelic patterns. In the process, 2ul of two to three pooled PCR products mixed with 9µl formamide-standard mix (0.11µl GS500 LIZ and 8.89ul Hi-Di formamide, Applied Biosystems). The volume of 11µl mixture were denatured for 3min at 95^0C then sent to capillary electrophoresis on Applied Biosystem's 3730 DNA analyzer.

Table 2: Primer sequences of SSR markers used to genotype the F_1 population

Marker	Type of Repeat	Left primer (F 5'-3')	Right Primer (R 5'-3')	Anneal. Temp (^{0}C)	Product Size(bp)	Product Range (bp)	Label
SSRY5	(GA)38	TGATGAAATTCAAAGCACCA	CGCCTACCACTGCCATAAAC	55	173	100-180	6-FAM
NS911	NA[a]	TGTTGTTCAGACGATGTCCAA	TTGAAGCAGTTATGAACCGT	57	127	100-180	VIC
SSRY148	NA[a]	GGCTTCATCATGGAAAAACC	CAATGCTTTACGGAAGAGCC	57	114	100-180	VIC
SSRY171	(TA)5CATA(GATA)8GC(GA)23(GTGA)2	ACTGTGCCAAAATAGCCAAATAGT	TCATGAGTGTGGGATGTTTTTATG	57	291	250-350	NED
SSRY64	(CT)13CG(CT)6	CGACAAGTCGTATATGTAGTATTCACG	GCAGAGGTGGCTAACGAGAC	57	194	150-250	VIC
SSRY119	(GA)8(G)3(GA)3(N)4(GA)3(N)32(A)5(GT)2(N)6(GT)3	AACATAGGCATTAAAGTTTGGCA	GCAAATGTGTTTTCAATATAAGGC	57	155	100-200	6-FAM
SSRY233	(CT)9	CGAAACGATCGAAGTTCCAC	TGTGGCCATCACACTCATTT	57	207	150-250	PET
SSRY309	(CT)11CCT	TCCTCCTCCCTCTTCAGATTC	GCCACAGGCTAAGGAAACAA	57	218	200-300	6-FAM
SSRY331	(CT)11	TCAGATCCCTCGGTTCTCAG	TGCTTTGCTTCTCAAGTCCA	57	198	220-320	PET
NS 193	NA[a]	TTGGGGGCTTTAAGTTGTTG	AAAGCCCATCCCCTCTATGT	57	258	200-300	PET

NA[a] = Not Applicable

3.2.5 Analysis of SSR data

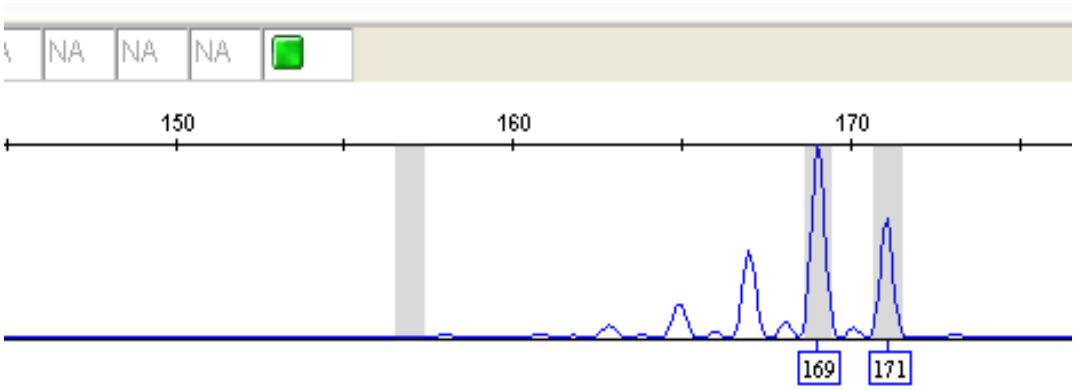
Allele sizes were determined using Genemapper version 3.7 Software (Applies Biosystems, foster City, CA, USA). After cleaning and analyzing the data extracted from genemapper, data were imported to Microsoft Excel for further individual observation. All SSR markers followed diallelic mode of inheritance. Expected recombinants were identified as well as unexpected recombinants (allele not present in either of the parents) eventually true to type progenies were detected and counted as well as self and off-types.

3.3 Results

3.3.1 Parents and population screening

Thirty eight SSR primer pairs were used to perform the amplification reactions, which generated distinct bands. All bands were scored for the presence, absence or ambiguous for each parent by visual inspection. Band size ranged from 100bp – 300bp. Based on 36 SSR primers used, ten primer pairs showed highly polymorphic bands. Each band produced by these primers was distinct and reproducible showing clear diallelic model of inheritance. The allele size selected was the one with highest picks (Fig. 1). Thirteen primers were monomorphic showing the single allele and fifteen primers did not generate recognizable PCR amplification. Therefore 28 primer pairs were not used during screening of population. All these primers were SSR markers that have been mapped in cassava genome hence, ten polymorphic markers viz. NS193, NS911, SSRY119, SSRY 223, SSRY309, SSRY331, SSRY64, SSRY148, SSRY171 and SSRY5 were used in PCR amplification for F_1 population between AR40-6 x ARAL (referred as ARAL population in this study) after parental screening. MarkersNS193, NS911, SSRY119, SSRY 223, SSRY309, SSRY331, SSRY64, SSRY148 and SSRY171amplified well at annealing temperature 57^0C except for SSRY5, which amplified at 55^0C.

AA112_SSRY119



AA113_SSRY119

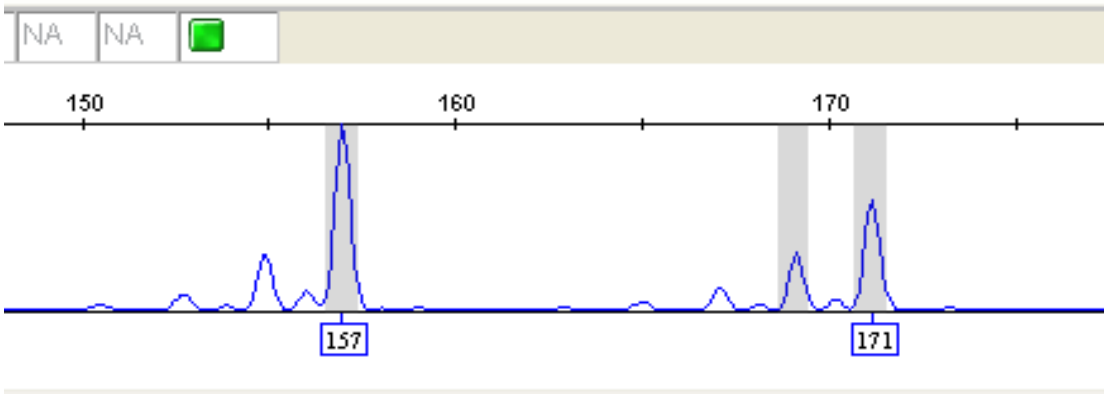


Figure 1: Marker SSRY 119 shows multiple allele sizes in highest peaks of individual AA112 and AA113

3.3.2 Population verification

Identification of true population cross was identified by observation of each marker patterns in the progenies. Marker patterns observed in progenies were consistent with what would be expected based on parental allele sizes (Table 3). The results reveal that 159 genotypes were true cross of AR40-6 x Albert (Table 4). The results also showed that there were 41 non hybrids (Table 5), which were caused by unexpected allele 200bp in SSRY64 and allele 290bp in SSRY331 (Fig. 2 and 3). One genotype (AA9) was removed because 90% of data were missing.

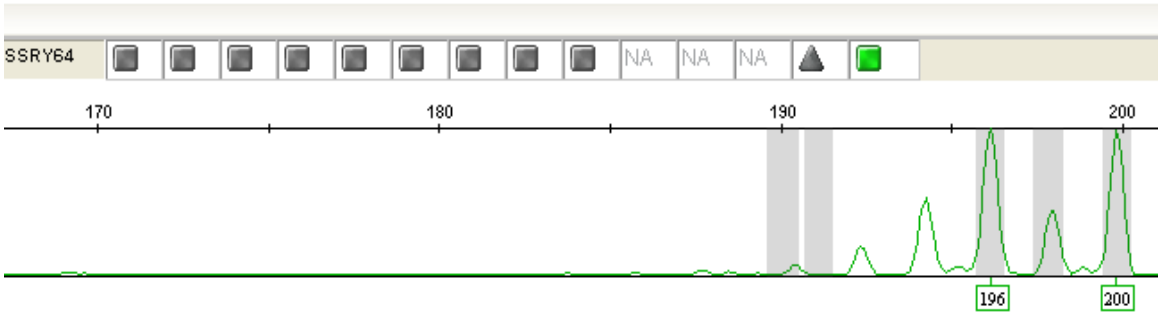


Figure 2: Marker SSRY 64 shows unexpected allele size 200bp

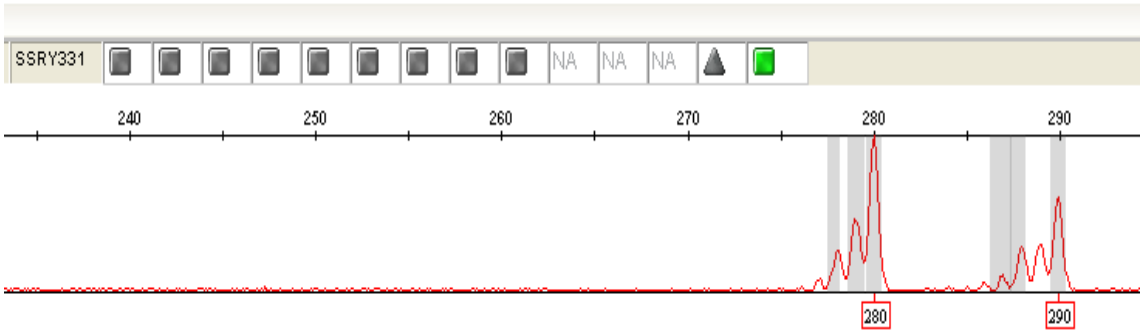


Figure 3: Marker SSRY 331 shows unexpected allele size 290bp

A total of 18 genotypes were regarded as self-cross because they show 100% similarity to the expected self cross to AR40-6 in marker (SSRY5, SSRY331, SSRY148, SSRY233), 56% in marker SSRY 309, 39% in marker NS911, 33% in marker SSRY119, 28% in marker SSRY64 and 11% in marker SSRY171 and NS193 respectively (Table 6).

Table 3: Marker name and parental alleles of AR40-6 X Albert F_1 cross lines

Marker	♀Allele for AR40-6	♂Allele for Albert	Expected Alleles in F_1	Number Allele expected	Number Allele observed
SSRY331	280/280	288/298	280/288, 280/298	3	5
SSRY5	127/127	106/106	127/106	2	2
SSRY64	196/198	190/191	196/190,196/191, 198/190, 198/191	4	5
SSRY148	113/113	111/113	113/111, 113/113	2	2
NS911	113/123	123/123	113/123, 123/123	2	2
SSRY171	303/289	287/289	303/287,303/289, 289/287, 289/289	3	3
SSRY309	225/221	218/231	225/218,225/231, 221/218, 221/231	4	4
NS193	250/258	250/256	250/250,250/256, 258/250, 258/256	3	3
SSRY119	157/169	157/171	157/157,157/171, 169/157, 169/171	3	3
SSSRY233	210/210	206/210	210/206, 210/210	2	2

3.3.3 Individual marker distribution

In marker SSRY331, expected recombinants had allele 280bp/288bp and 280bp/298bp (Table 4). Individuals with allele 280bp/280bp were regarded selfed plants of AR40-6 (female). Unexpected alleles were observed from individual with 290bp/280bp and 278bp/298bp of which allele 278bp and 290bp were likely to be an outcross (Table 5). True expected individual in SSRY 5 had allele 127bp/106bp. Allele 127bp/127bp was regarded as self's from AR40-6 and allele 106bp/106bp selfs from Albert because their allele was the same as that from the parents. Marker SSRY64 had expected alleles196/190, 196/191, 198/190 and 198/191 respectively (Table 5).

Individuals with unexpected allele (allele not in expected recombinants) such as 196bp/200bp and 198bp/200bp were regarded as off-types, outcross or mutants. Individual with allele 196bp/196bp and 196bp/198bp were considered selfed individuals from AR40-6. Marker polymorphism in SSRY 148 showed monomorphic in AR40-7 (113bp/113bp)

and polymorphic in Albert (111bp/113bp). Allele 113bp was shared between AR40-6 and Albert. In marker NS911 expected recombinants had individual with allele 113bp/123bp and 123bp/123bp however, individual with allele 123bp/123bp is likely to be self-cross of AR40-6. The results showed plants (individual) with allele 113bp/113bp was regarded as AR40-6 selfed individuals plants cross. In marker SSRY 171, unusual individual with allele 287bp/287bp was observed. This indicated there was selfed individual in male parent Albert. An individual with allele 103bp/103bp was regarded as AR40-6 selfed plant because it has all maternal alleles. Three types of alleles (221bp/221bp, 225bp/225bp and 221bp/225bp) were observed in AR40-6 female cross from marker SSRY 309. Unexpected individual with male cross (218bp/218bp) were also observed in this marker however a marker was highly polymorphic with different allele of 225bp/221bp and 218bp/231bp in AR40-6 and Albert respectively. Individual with male self-allele of 256bp/256bp was observed in marker NS 193. Selfed individual for AR40-6 with allele 258bp/258bp was also identified in marker NS 193. Expected individuals with allele 250bp/258bp were likely to be selfed plants from AR40-6 because both parents share allele 250bp. In marker SSRY 119, individual with allele 169bp/169bp were also regarded as self from AR40-6. Expected recombinants had alleles with 157bp/157bp, 157bp/171bp, 157bp/169bp and 169bp/171bp. Marker SSRY223 was observed to have 99% of expected true cross individuals (Table 4).

Table 4: Marker used for SSR genotyping, allele of each marker and true cross individuals (AA) from a cross between AR40-6 x Albert

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA100	288/280	-9/-9	196/191	111/113	123/123	287/289	218/225	-9/-9	157/157	210/210	True cross[1]
AA101	288/280	-9/-9	190/196	111/113	113/123	287/289	221/221	-9/-9	157/169	210/210	True cross
AA103	280/280	106/127	196/198	113/113	123/123	303/289	225/225	250/256	157/169	210/210	True cross
AA104	288/280	106/127	196/191	111/113	113/123	287/287	218/221	250/258	157/169	206/210	True cross
AA105	280/280	106/127	190/196	113/113	123/123	303/289	225/225	250/256	157/171	210/210	True cross
AA106	288/280	106/127	190/196	113/113	123/123	287/289	221/221	250/256	171/169	211/211	True cross
AA108	288/280	106/127	196/191	113/113	123/123	287/303	218/221	250/256	157/171	206/210	True cross
AA10	-9/-9	106/127	190/196	-9/-9	113/123	-9/-9	-9/-9	250/250	157/169	-9/-9	True cross
AA110	288/280	106/127	196/191	111/113	123/123	287/289	221/221	250/250	157/157	210/210	True cross
AA111	280/280	106/127	196/191	113/113	123/123	287/289	225/225	250/256	171/169	206/210	True cross
AA112	288/280	106/127	196/191	113/113	113/123	287/289	225/225	250/258	171/169	210/210	True cross
AA113	288/280	106/127	190/196	113/113	123/123	289/289	218/221	250/250	157/171	206/210	True cross
AA115	288/280	106/127	196/191	111/113	123/123	287/289	218/221	250/258	157/157	206/210	True cross
AA116	288/280	106/127	196/191	111/113	123/123	287/289	225/225	250/250	157/169	210/210	True cross
AA117	280/280	106/127	190/198	111/113	113/123	303/289	221/221	250/258	157/157	206/210	True cross
AA118	298/280	106/127	190/196	111/113	113/123	287/289	218/221	250/256	157/169	210/210	True cross
AA119	288/280	106/127	196/191	111/113	113/123	287/303	218/225	250/258	157/169	206/210	True cross
AA120	298/280	106/127	196/191	111/113	113/123	287/303	225/231	258/256	157/171	206/210	True cross
AA121	288/280	106/127	196/191	113/113	123/123	287/289	225/231	250/250	157/171	206/210	True cross
AA124	280/280	106/127	190/196	111/113	113/123	287/289	218/225	250/258	157/169	206/210	True cross
AA125	280/280	106/127	190/198	111/113	123/123	287/289	218/221	250/256	157/157	206/210	True cross
AA126	280/280	106/127	196/196	113/113	113/123	287/289	218/221	250/256	157/171	206/210	True cross
AA127	288/280	106/127	198/191	113/113	123/123	287/303	221/231	258/256	171/169	210/210	True cross
AA128	288/280	106/127	196/191	111/113	123/123	287/289	225/231	250/256	157/169	206/210	True cross

[1]True crosses are declared if more than four markers shows expected recombinants for true hybrid

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA129	298/280	106/127	196/191	111/113	113/123	287/289	218/225	250/256	157/169	210/210	True cross
AA133	288/280	106/127	198/191	111/113	123/123	287/303	218/218	250/250	157/169	210/210	True cross
AA134	288/280	106/127	196/191	113/113	113/123	287/289	218/221	250/256	171/169	210/210	True cross
AA136	298/280	106/127	198/191	113/113	113/123	287/303	221/231	250/256	157/171	210/210	True cross
AA137	298/280	106/127	190/196	113/113	123/123	287/289	221/231	250/250	157/171	210/210	True cross
AA138	288/280	106/127	198/191	111/113	123/123	287/303	218/225	250/250	157/169	210/210	True cross
AA13	288/280	106/127	190/196	111/113	123/123	287/289	-9/-9	258/256	157/169	210/210	True cross
AA142	287/280	106/127	196/191	111/113	123/123	287/289	225/231	250/250	157/169	210/210	True cross
AA146	298/280	106/127	190/198	113/113	123/123	303/289	221/231	250/256	157/171	206/210	True cross
AA147	288/280	106/127	196/191	111/113	113/123	287/289	218/225	250/256	157/157	206/210	True cross
AA148	288/280	106/127	190/198	111/113	113/123	303/289	221/231	250/250	157/169	206/210	True cross
AA14	-9/-9	106/127	-9/-9	113/113	123/123	-9/-9	-9/-9	250/256	171/169	206/210	True cross
AA150	288/280	106/106	196/191	113/113	123/123	287/289	221/231	250/256	171/169	210/210	True cross
AA151	298/280	106/127	198/191	111/113	113/123	287/303	221/231	250/250	157/157	206/210	True cross
AA152	298/280	106/127	196/191	111/113	123/123	287/289	218/221	258/256	157/169	206/210	True cross
AA153	298/280	106/127	198/191	111/113	123/123	287/303	218/221	250/256	157/157	210/210	True cross
AA154	298/280	106/127	198/191	111/113	113/123	287/303	218/221	258/256	157/169	206/210	True cross
AA155	288/280	106/127	198/191	113/113	123/123	287/303	221/231	250/250	157/171	210/210	True cross
AA156	280/280	106/127	190/198	111/113	113/123	303/289	221/225	250/250	157/169	206/210	True cross
AA157	288/280	106/127	196/191	113/113	113/123	287/289	225/231	250/250	171/169	210/210	True cross
AA158	280/280	106/127	196/191	113/113	123/123	287/289	225/231	250/250	157/169	210/210	True cross
AA15	298/280	106/127	198/191	113/113	123/123	287/303	221/231	250/250	157/171	210/210	True cross
AA161	298/280	106/127	198/191	111/113	123/123	287/303	218/225	250/258	157/157	210/210	True cross
AA162	280/280	106/127	190/196	111/113	123/123	289/289	218/225	250/258	157/171	206/210	True cross
AA163	298/280	106/127	198/191	111/113	123/123	-9/-9	221/231	250/256	157/157	210/210	True cross
AA164	288/280	106/127	198/191	113/113	123/123	287/303	218/221	250/250	171/169	210/210	True cross
AA165	288/280	106/127	190/196	111/113	123/123	289/289	225/231	250/256	157/157	206/210	True cross

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA166	298/280	106/127	198/191	113/113	123/123	287/303	218/225	250/256	157/171	206/210	True cross
AA167	279/279	106/127	196/191	111/113	113/123	289/289	225/231	250/256	157/157	210/210	True cross
AA168	287/280	106/127	196/191	113/113	123/123	287/303	231/221	250/258	157/171	210/210	True cross
AA169	298/280	106/127	198/191	111/113	123/123	303/289	218/221	250/256	157/169	210/210	True cross
AA16	-9/-9	106/127	-9/-9	113/113	113/123	-9/-9	-9/-9	250/256	-9/-9	206/211	True cross
AA170	288/280	106/127	196/191	111/113	113/123	303/289	231/221	250/250	157/157	206/210	True cross
AA171	288/280	106/127	198/191	111/113	123/123	287/303	218/225	250/250	157/157	206/210	True cross
AA172	288/280	106/127	196/191	113/113	123/123	287/289	231/225	250/258	171/169	206/210	True cross
AA173	288/280	106/127	198/191	113/113	123/123	287/303	218/225	250/258	157/171	210/210	True cross
AA174	288/280	106/127	190/198	113/113	113/123	303/289	231/225	250/258	171/169	206/210	True cross
AA175	288/280	106/127	196/191	113/113	113/123	287/303	221/221	250/258	157/171	206/210	True cross
AA179	298/280	106/127	198/191	113/113	123/123	287/303	218/225	250/256	171/169	206/210	True cross
AA17	288/280	106/127	196/190	111/113	123/123	-9/-9	231/221	-9/-9	157/169	206/210	True cross
AA182	288/280	106/127	198/191	113/113	113/123	287/303	218/225	250/256	171/169	206/210	True cross
AA183	298/280	106/127	196/191	111/113	123/123	287/289	231/225	258/256	157/169	210/210	True cross
AA184	288/280	106/127	198/191	113/113	113/123	287/303	218/225	258/256	169/169	210/210	True cross
AA185	288/280	106/127	196/191	113/113	113/123	287/289	231/225	250/258	171/169	210/210	True cross
AA186	298/280	106/127	196/191	113/113	113/123	287/289	218/225	258/256	169/169	206/210	True cross
AA187	-9/-9	106/127	190/196	113/113	113/123	303/289	221/225	250/250	157/169	206/210	True cross
AA189	280/280	106/127	198/191	113/113	123/123	287/303	231/225	258/256	157/157	206/210	True cross
AA18	298/280	106/127	198/191	111/113	123/123	287/303	231/225	258/256	157/171	206/210	True cross
AA191	280/280	127/127	196/191	113/113	113/123	289/289	221/221	258/258	157/171	206/210	True cross
AA194	280/280	127/127	190/196	113/113	123/123	287/289	221/221	258/258	157/171	206/210	True cross
AA195	280/280	106/127	190/196	111/113	113/123	287/289	221/221	250/250	157/169	206/210	True cross
AA196	288/280	106/127	196/191	111/113	113/123	287/289	231/225	250/250	157/169	210/210	True cross
AA197	288/280	106/127	198/191	111/113	113/123	287/289	218/221	258/256	157/157	206/210	True cross
AA199	287/280	106/127	196/191	113/113	113/123	287/289	231/225	250/258	157/171	206/210	True cross

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA19	298/280	106/127	196/191	113/113	123/123	287/289	218/225	250/258	157/171	206/210	True cross
AA200	280/280	106/127	196/198	113/113	113/123	287/289	221/221	250/258	157/171	206/210	True cross
AA201	288/280	106/127	198/191	111/113	113/123	287/303	218/221	250/256	157/169	206/210	True cross
AA202	280/280	106/127	196/191	113/113	123/123	287/289	218/225	250/258	157/171	206/210	True cross
AA203	288/279	106/127	198/191	111/113	113/123	287/303	-9/-9	250/250	157/169	210/210	True cross
AA207	280/280	106/127	196/198	113/113	123/123	303/289	218/225	250/258	157/171	206/210	True cross
AA208	288/280	106/127	198/191	111/113	123/123	287/303	231/221	250/250	157/169	210/210	True cross
AA209	288/280	106/127	198/191	111/113	113/123	287/303	231/221	258/256	171/169	206/210	True cross
AA20	288/280	106/127	191/196	111/113	113/123	287/303	231/225	258/256	157/169	206/210	True cross
AA211	298/280	106/127	196/196	113/113	123/123	287/289	218/221	250/256	157/171	206/210	True cross
AA212	288/279	106/127	196/191	111/113	123/123	287/289	218/221	250/250	157/169	210/210	True cross
AA213	298/280	106/127	198/191	113/113	113/123	287/303	231/225	250/256	157/171	210/210	True cross
AA214	298/280	106/127	191/196	111/113	123/123	287/289	231/221	250/258	157/169	206/210	True cross
AA215	288/280	106/127	191/196	113/113	113/123	287/303	231/225	250/250	157/171	206/210	True cross
AA216	288/280	106/127	191/196	111/113	113/123	287/289	231/221	250/250	157/169	206/210	True cross
AA21	288/280	106/127	190/198	113/113	113/123	303/289	218/225	250/250	157/171	206/210	True cross
AA221	278/298	106/127	196/191	113/113	123/123	287/287	231/225	250/256	171/169	210/210	True cross
AA224	298/280	106/127	198/191	113/113	123/123	287/303	218/225	250/256	171/169	210/210	True cross
AA22	288/280	106/127	198/191	111/113	113/123	287/303	218/225	250/250	157/169	206/210	True cross
AA23	287/280	106/127	190/196	111/113	123/123	287/289	231/225	250/250	157/157	206/210	True cross
AA25	298/280	106/127	196/191	113/113	113/123	287/289	218/225	250/256	157/171	210/210	True cross
AA26	288/280	106/127	190/196	113/113	123/123	287/289	231/225	250/256	157/171	210/210	True cross
AA27	298/280	106/127	198/191	113/113	123/123	287/303	231/221	250/258	157/171	206/210	True cross
AA29	280/280	106/127	198/191	111/113	113/123	287/303	218/221	250/250	-9/-9	206/210	True cross
AA30	280/280	106/127	190/198	113/113	123/123	303/289	218/225	250/256	-9/-9	206/210	True cross
AA31	298/280	106/127	190/198	113/113	123/123	303/289	231/221	250/258	-9/-9	206/210	True cross
AA32	298/280	106/127	198/191	111/113	123/123	287/303	218/225	250/256	-9/-9	206/210	True cross

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA33	288/280	106/127	198/191	111/113	123/123	287/303	231/225	258/256	-9/-9	210/210	True cross
AA35	298/280	106/127	196/191	113/113	113/123	287/289	231/221	258/256	-9/-9	206/210	True cross
AA36	280/280	106/127	190/196	113/113	123/123	287/289	221/221	258/256	-9/-9	206/210	True cross
AA38	298/280	106/127	198/191	113/113	123/123	287/303	218/225	250/256	157/171	206/210	True cross
AA39	288/280	106/127	198/191	113/113	113/123	287/303	218/221	250/258	157/169	206/210	True cross
AA3	-9/-9	106/127	-9/-9	-9/-9	113/123	-9/-9	-9/-9	250/250	157/169	206/210	True cross
AA40	298/280	127/127	190/196	113/113	123/123	287/289	218/221	258/256	157/169	210/210	True cross
AA41	298/280	106/127	198/191	113/113	123/123	287/303	218/221	258/256	157/157	206/210	True cross
AA42	288/280	106/127	190/196	113/113	-9/-9	287/289	218/225	250/258	157/169	206/210	True cross
AA43	298/280	106/127	196/191	113/113	123/123	287/289	218/221	250/258	157/169	210/210	True cross
AA44	288/280	106/127	196/191	111/113	113/123	287/289	231/221	250/250	157/169	210/210	True cross
AA45	288/280	106/127	198/191	111/113	113/123	287/303	218/225	250/258	157/157	206/210	True cross
AA48	288/280	106/127	190/196	113/113	123/123	287/289	218/221	250/258	157/171	210/210	True cross
AA4	288/280	106/127	190/196	113/113	113/123	287/289	231/225	250/258	157/171	210/210	True cross
AA50	288/280	106/127	190/198	111/113	113/123	303/289	218/225	250/256	157/157	210/210	True cross
AA51	280/280	106/127	190/198	111/113	123/123	303/289	218/225	250/256	157/157	206/210	True cross
AA53	288/280	106/127	190/198	113/113	113/123	303/289	218/225	250/250	171/169	206/210	True cross
AA54	280/280	106/127	196/191	113/113	123/123	287/289	218/221	250/256	157/171	206/210	True cross
AA56	288/280	106/127	196/191	111/113	123/123	287/289	218/221	250/250	157/169	210/210	True cross
AA57	280/280	106/127	190/196	111/113	113/123	303/289	218/221	250/256	157/157	206/210	True cross
AA58	288/280	106/127	190/196	113/113	113/123	287/289	231/221	250/258	171/169	206/210	True cross
AA59	288/280	106/127	198/191	111/113	123/123	287/303	218/225	250/256	157/169	206/210	True cross
AA5	-9/-9	106/127	-9/-9	-9/-9	113/123	-9/-9	-9/-9	-9/-9	171/169	206/210	True cross
AA60	288/280	106/127	196/191	111/113	123/123	287/289	231/221	250/250	157/157	210/210	True cross
AA62	298/280	106/127	198/191	111/113	123/123	287/303	231/225	250/258	157/169	206/210	True cross
AA63	298/280	106/127	196/191	113/113	123/123	287/289	218/218	250/258	157/171	206/210	True cross
AA64	280/280	106/106	190/198	111/113	123/123	-9/-9	218/221	258/256	157/157	206/210	True cross

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA65	298/280	106/127	198/191	113/113	113/123	287/303	218/225	250/250	171/169	206/210	True cross
AA66	288/280	106/127	190/196	113/113	113/123	287/289	231/225	250/250	171/169	206/210	True cross
AA67	298/280	106/127	196/191	111/113	113/123	287/289	231/225	258/256	157/169	210/210	True cross
AA6	298/280	106/127	196/191	113/113	113/123	287/289	231/221	258/256	157/171	206/210	True cross
AA70	288/280	106/127	190/198	113/113	123/123	303/289	218/221	250/258	157/171	206/210	True cross
AA71	288/280	106/127	190/196	111/113	123/123	289/289	218/221	250/258	157/157	206/210	True cross
AA72	298/280	106/127	198/191	113/113	123/123	287/303	218/221	250/256	157/171	210/210	True cross
AA73	280/280	106/127	196/190	113/113	113/123	287/289	218/221	258/258	157/171	206/210	True cross
AA74	298/280	106/127	196/191	113/113	113/123	287/289	218/221	258/256	157/171	210/210	True cross
AA75	288/280	106/127	198/191	111/113	113/123	287/303	218/221	250/250	157/157	206/210	True cross
AA78	298/280	106/127	198/191	111/113	123/123	287/303	218/225	250/250	157/157	210/210	True cross
AA79	298/280	106/127	198/191	113/113	113/123	287/303	218/221	258/256	157/171	210/210	True cross
AA7	-9/-9	106/127	198/191	113/113	123/123	-9/-9	-9/-9	258/256	157/171	206/210	True cross
AA80	288/280	106/127	190/198	113/113	113/123	303/289	231/225	250/250	171/169	210/210	True cross
AA85	288/280	106/127	198/191	111/113	123/123	287/303	231/221	250/258	157/157	210/210	True cross
AA87	280/280	106/127	190/196	111/113	123/123	287/289	218/221	258/256	157/169	206/210	True cross
AA88	288/280	106/127	198/191	111/113	123/123	287/289	231/221	250/258	157/157	206/210	True cross
AA89	298/280	106/127	198/191	111/113	113/123	287/303	231/221	258/256	157/169	206/210	True cross
AA8	280/280	106/127	-9/-9	111/113	113/123	-9/-9	-9/-9	256/256	157/157	206/210	True cross
AA90	298/280	106/127	198/191	113/113	123/123	287/303	218/221	250/250	171/169	210/210	True cross
AA91	298/280	106/127	198/191	113/113	123/123	287/303	218/221	250/256	171/169	206/210	True cross
AA92	298/280	106/127	198/191	111/113	123/123	287/303	231/225	250/256	157/157	206/210	True cross
AA95	298/280	-9/-9	198/191	113/113	123/123	287/303	231/221	-9/-9	157/171	210/210	True cross
AA96	287/280	-9/-9	198/191	113/113	123/123	287/303	231/225	-9/-9	171/169	206/210	True cross
AA97	280/280	-9/-9	190/198	111/113	113/123	303/289	218/225	-9/-9	157/157	206/210	True cross
AA98	288/280	-9/-9	198/191	111/113	123/123	287/303	231/225	-9/-9	157/157	206/210	True cross
AA99	-9/-9	-9/-9	198/191	113/113	113/123	-9/-9	-9/-9	-9/-9	171/169	206/210	True cross

Table 5: Marker used for SSR genotyping, allele of each marker and off-types individuals (AA) from a cross between AR40-6 x Albert

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA102	280/280	-9/-9	196/200	111/113	123/123	287/289	218/225	-9/-9	157/169	206/210	Off-type[2]
AA107	290/280	106/127	196/200	113/113	113/123	287/289	221/221	250/256	157/171	206/210	Off-type
AA109	290/280	106/127	200/196	113/113	113/123	287/289	218/225	250/256	157/171	206/210	Off-type
AA114	280/280	106/127	200/198	111/113	113/123	303/289	221/221	250/250	157/169	206/210	Off-type
AA122	290/280	106/127	200/198	111/113	123/123	303/289	221/221	250/256	157/157	206/210	Off-type
AA132	280/280	106/127	196/200	113/113	113/123	287/289	221/221	258/256	171/169	206/210	Off-type
AA139	290/280	106/127	200/198	111/113	123/123	303/289	221/225	250/250	157/157	206/210	Off-type
AA140	280/280	106/127	200/198	111/113	113/123	303/289	225/225	250/258	157/169	206/210	Off-type
AA141	280/280	106/127	200/198	113/113	113/123	303/289	218/225	250/258	171/169	206/210	Off-type
AA143	290/280	106/127	200/198	113/113	123/123	303/289	221/221	250/256	171/169	206/210	Off-type
AA144	290/280	106/127	200/198	113/113	113/123	303/289	221/221	258/256	171/169	206/210	Off-type
AA145	290/280	106/127	196/200	113/113	113/123	289/289	218/221	250/256	171/169	206/210	Off-type
AA176	280/280	106/127	200/196	113/113	113/123	289/289	221/221	250/256	157/169	206/210	Off-type
AA177	280/280	106/127	200/196	113/113	113/123	289/289	218/221	250/256	157/171	206/210	Off-type
AA178	290/280	106/127	200/196	111/113	123/123	303/289	221/221	258/256	157/157	206/210	Off-type
AA190	298/280	106/127	198/200	113/113	113/123	287/303	231/221	250/256	157/169	206/210	Off-type
AA192	280/280	127/127	196/200	113/113	113/123	287/289	221/225	258/250	157/171	206/210	Off-type
AA193	290/280	106/127	198/191	113/113	113/123	287/289	221/221	258/250	157/169	206/210	Off-type
AA198	290/280	106/127	196/198	111/113	123/123	303/289	221/225	258/256	157/169	206/210	Off-type
AA1	280/280	106/127	200/198	113/113	123/123	303/289	218/221	250/258	157/171	206/210	Off-type
AA204	290/280	106/127	196/191	111/113	113/123	287/289	218/221	250/250	157/157	206/210	Off-type
AA205	290/280	106/127	196/191	113/113	123/123	287/289	218/225	250/256	171/169	206/210	Off-type

[2]Individual is declared to be off-type if one marker shows unexpected recombinants from true hybrid due to insertion and deletion of one or more allele

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA206	290/280	106/127	190/198	111/113	123/123	303/289	218/225	250/258	157/169	206/210	Off-type
AA223	290/280	106/127	190/196	113/113	123/123	287/289	221/225	250/250	171/169	206/210	Off-type
AA226	280/280	106/127	198/200	111/113	113/123	303/289	221/221	250/250	157/157	206/210	Off-type
AA24	280/280	106/127	200/198	111/113	123/123	303/289	218/225	250/258	157/169	206/210	Off-type
AA28	280/280	106/127	198/200	113/113	123/123	303/289	218/225	250/258	171/169	206/210	Off-type
AA2	296/280	106/127	190/198	113/113	113/123	303/289	231/221	250/258	157/169	206/210	Off-type
AA34	290/280	106/127	198/200	111/113	123/123	303/289	218/225	258/256	-9/-9	206/210	Off-type
AA37	290/280	106/127	198/200	111/113	123/123	303/289	221/225	258/256	157/169	206/210	Off-type
AA46	280/280	106/127	198/200	113/113	113/123	303/289	218/221	250/258	171/169	206/210	Off-type
AA47	280/280	106/127	198/200	113/113	113/123	303/289	218/225	250/250	157/169	206/210	Off-type
AA52	290/280	106/127	200/196	113/113	123/123	287/289	218/225	250/250	157/171	206/210	Off-type
AA55	280/280	106/127	200/198	111/113	113/123	303/289	221/225	258/256	157/169	206/210	Off-type
AA61	280/280	106/127	198/200	111/113	113/123	303/289	221/225	250/256	157/157	206/210	Off-type
AA76	280/280	106/127	200/196	113/113	113/123	287/289	218/225	258/256	157/171	210/210	Off-type
AA77	-9/-9	106/127	200/196	111/113	113/123	-9/-9	-9/-9	250/250	157/169	206/210	Off-type
AA82	290/280	106/127	190/196	111/113	113/123	303/289	221/221	250/250	157/157	206/210	Off-type
AA83	280/280	106/127	200/198	113/113	113/123	303/289	221/225	250/256	157/171	206/210	Off-type
AA93	280/280	106/127	200/196	113/113	123/123	287/289	218/225	250/258	157/171	206/210	Off-type
AA94	280/280	106/127	198/200	111/113	113/123	303/289	221/221	250/256	157/169	206/210	Off-type

Table 6: Marker used for SSR genotyping, allele of each marker and self-cross individuals (AA) from a cross between AR40-6 x Albert

Individual	SSRY331	SSRY5	SSRY64	SSRY148	NS911	SSRY171	SSRY309	NS193	SSRY119	SSRY233	Result
AR40-6	280/280	127/127	196/198	113/113	113/123	303/289	221/225	250/258	157/169	210/210	**Female parent**
ALBERT	288/298	106/106	190/191	111/113	123/123	287/289	218/231	250/256	157/171	206/210	**Male parent**
AA11	280/280	127/127	196/196	113/113	113/113	-9/-9	-9/-9	250/258	157/157	210/210	Self[3]
AA123	280/280	127/127	196/196	113/113	113/123	287/289	221/221	250/258	169/169	210/210	Self
AA12	-9/-9	127/127	-9/-9	-9/-9	113/113	-9/-9	-9/-9	250/258	157/157	210/210	Self
AA130	280/280	127/127	196/198	113/113	113/123	303/289	225/225	250/250	157/157	210/210	Self
AA159	280/280	127/127	198/196	113/113	113/113	303/289	221/225	250/250	169/169	210/210	Self
AA180	280/280	127/127	196/198	113/113	113/113	303/289	221/225	250/258	157/169	210/210	Self
AA181	280/280	127/127	196/196	113/113	123/123	287/289	221/225	250/258	157/171	210/210	Self
AA188	280/280	127/127	196/191	113/113	123/123	303/289	221/221	258/258	157/169	210/210	Self
AA210	280/280	127/127	196/198	113/113	123/123	303/289	221/225	250/258	157/169	210/210	Self
AA217	280/280	127/127	196/198	113/113	113/123	303/289	-9/-9	258/258	169/169	210/210	Self
AA218	280/280	127/127	196/198	113/113	123/123	303/303	225/225	250/258	157/157	210/210	Self
AA219	280/280	127/127	196/198	113/113	113/123	-9/-9	221/221	250/258	157/169	210/210	Self
AA220	280/280	127/127	196/198	113/113	113/123	303/289	225/225	250/250	157/169	210/210	Self
AA222	280/280	127/127	196/196	113/113	123/123	303/289	221/225	250/258	157/157	210/210	Self
AA49	280/280	127/127	198/196	113/113	113/113	303/289	221/221	250/258	169/169	210/210	Self
AA81	280/280	127/127	198/196	113/113	113/113	303/303	225/225	258/256	169/169	210/210	Self
AA84	280/280	127/127	196/196	113/113	113/123	287/289	221/221	-9/-9	157/157	210/210	Self
AA86	280/280	127/127	198/196	113/113	113/113	303/289	221/221	258/256	169/169	210/210	Self

[3]Individual declared to be self if two markers shows martenal recombinants

3.4 Discussion

Verification of F1 is more efficient for clonally propagated crops because in F1's is where the crossing over occurred. Some progenies are not genetically easy to be classified because their genotypes have some single –allele which is common in both parents. Such type of individuals are likely to be female self-cross or true expected cross. In this case DNA sequencing of each alleles of similar size will provide the true polymorphic information. High polymorphic markers are important to verify population crosses and inability to identify selfs from true cross makes a marker to lack insufficient information (low polymorphism) and weak as a marker tool for verification of population crosses.

SSRs Marker was used to genotype the population crosses and identified the true F_1 hybrids. Ten SSR markers identified 159 individuals plants as true cross of AR40-6 and Albert indicating true recombinant alleles from the two parents for each marker. Microsatellite genotyping data was used for selection of true to type individual based on expected recombinants from parental allele. Individuals with unexpected allele (allele not in parents) were regarded as non-hybrids (off-types). These individuals were observed in marker SSRY331, SSRY64 and SSRY233. In SSRY331, allele 290bp/280bp and 278bp/298bp were off-types because 278bp and 290bp was not in either of the two parents. In SSRY64, individuals with allele 196bp/200bp and 198bp/200bp were also off-types because of unexpected allele 200bp which was not found in either of the parents (AR40-6 and Albert) and SSRY233 had unexpected allele 211bp which is not in either of the parents. The variation in allelic sizes observed from off-types progenies indicated either insertion of one, or two nucleotides residues in each of allelic marker. The insertion or deletion of some nucleotides residues might be due to somaclonal variation, DNA mutation or DNA damage as a result of clonal propagation of genetic recombination. Other reasons for such off-types might be seed contamination/seed mixture, wrongly

selection of pure Albert line as parent or Albert line has some heterozygous region to the clone used for DNA genotyping.

Non-parental allele observed in this study caused occurrence of off-types. Occurrence of off-types was caused by genetic variation in the parental planting materials (cuttings) used for the crosses. Although cassava is clonally propagated crop, it is also out crossing species and farmers grow several varieties in one field which might cross one another. These plants produce seeds which germinate in the farmers' fields. These seedlings (offspring) can look phenotypically similar to the parents but genetically difference. Wrongly selection of pure Albert plants for crossing purposes, accidentally collection of seeds from non-labelled fruit and seed mixture might be expected as human errors which causes occurrence of off-types. The validation which proves true identity of the genotype in this study, genotyping is by using microsatellite markers. In this study, two more population occurred; self-cross population (sharing same parent) and non-hybrids population (non-parental allele) instead of one full-sib population as expected in controlled cross.

Another unexpected individuals were found in SSRY 5 (106bp/106bp), SSRY171 (287bp/287bp), SSRY309 (218bp/218bp) and NS193 (256bp/256bp). These individuals were selfs from male parent Albert which indicated that, there might be out-crossing of Albert genotype that had occurred from farmers' fields. Marker SSRY331 and marker SSRY309 had high number of selfs individuals of 54% and 43% respectively. This indicates that markers were weak to show difference in parental alleles or is a tightly linked marker that does not allow high recombination frequency. In marker NS193, 12 individuals were observed to lack allele information. More than two markers were considered more reliable for verification of genotypes than selection based on single

marker (Collard *et al.,* 2005). In addition, SSR markers allow the easy, fast, inexpensive, accurate, reliable, and simultaneous detection of polymorphisms at multiple loci in the genome using low quantities of DNA. This study therefore, revealed the use of SSRs markers in screening breeding materials for identification of likely genotype identity mix-ups, analysis of parent-offspring and possible crossing errors. Several writers have also reported SSRs have been successfully used to detect off-type genotypes, confirm true hybrids, high confidence authentication of genotypes and even differentiate full siblings (Mbanjo *et al.,* 2012; Takrama *et al.,* 2005; Dongre *et al.,* 2011, Brunings *et al.,* 2010). Sartie and Asiedu (2011) were able to use SSR markers to confirm successful hybridization of yam mapping population parents and the true identity of off springs.

3.5 Conclusion

Rapid and early identification of genotypes by using molecular markers can provide convenient way of accurate selection of true F1 hybrids of population cross in cassava improvement. Although controlled cross is practiced by many breeding programs, genotype mix-ups are a frequent problem. Admixtures population in this study was likely caused by variations in parents. A variety which has been in farmers fields for many years like Albert is likely to be affected by out-crossing. So far, there are no studies which clearly reported genetic purity of parental landraces and therefore out-crossing might be a cause in genotype admixture for local landrace varieties in other crops. Validation of parents by using molecular tools prior to crossing is a vital as admixture populations causes cost, time consuming and sometimes errors in genetic studies hence affect rapid improvements of the crop. The study have allowed for selection of the true hybrid population that was eventually use to provide good stock for mapping studies and linkage mapping of CBSD.

3.6 References

Al-Doss, A. A., Elshafei, A. A., Moustafa, K. A., Saleh, M. and Barakat, M. N. (2011). Comparative analysis of diversity based on morphoagronomic traits and molecular markers in durum wheat under heat stress. *African Journal Biotechnology* 10(19): 3671 – 3681.

Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. and Colvin, J. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant Disease* 91: 24 – 29.

Alves, A. A. C. (2002). Cassava Botany and Physiology. In: Hillocks, R. J., Thresh, J. M. and Bellotti, A. C. (Eds.), Cassava: Biology, Production and Utilization. CABI Publishing, pp. 67 – 89.

Arif, I. A., Khan, H. A., Shobrak, M., Al Homaidan, A. A., Al Sadoon, M., Al Farhan, A. H. and Bahkali, A. H. (2010). Interpretation of electrophoretograms of seven microsatellite loci to determine the genetic diversity of the Arabian Oryx Genet. *Molecular Research* 9(1): 259 – 265.

Asif, M., Mehboob, U. R. and Zafar, Y. (2006). Genotyping analysis of six maize (*Zea mays* L.) hybrids using DNA fingerprinting technology. *Pakistan Journal Botanic* 38: 1425 – 1430.

Banhos, A., Hrbeck, T., Gravena, W. and Sanaiotti, T. (2008). Genomic resources for the conservation and management of the harpy eagle (*Harpiaharpyja*, Falconiformes, Accipitridae). *Genetics Molecular Biology* 31: 146 – 154.

Brunings, A. M., Moyer, C., Peres, N. and Folta, K. M. (2010). Implementation of simple sequence repeat markers to genotype Florida strawberry varieties. *Euphytica* 173: 63 – 75.

Bull, S. E., Ndunguru, J., Gruissem, W., Beeching, J. R. and Vanderschuren, H. (2011). Cassava: constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Reproduction* 30: 779 – 787.

Campo B. V. H., Hyman, G. and Bellotti, A. (2011). Threats to cassava productions: known and potential geographical distribution of four key biotic constraints. *Journal Article* 3: 329 – 345.

Chan, C. H., Zhao, Y., Cheung, M. Y. and Chambers, G. K. (2008). Isolation and characterization of microsatellites in the kakerori (*Pomareadimidiata*) using feathers as source of DNA. *Conservation Genetics* 9: 1067 – 1070.

Charles, A. L., Sriroth, K. and Huang, T. C. (2005). Proximate composition, mineral contents, hydrogen cyanide and phytic acid of 5 cassava genotypes. *Food Chemistry* 92: 15 – 20.

Chavarriaga-Aguirre, P., Maya, M. M., Bonierbale, M. W., Kresovich, S., Fregene, M. A., Tohme, J. and Kochert, G. (1998). Microsatellites in cassava (*Manihotesculenta* Crantz): discovery, inheritance and variability, *Theoretical and Applied Genetics* 97: 493-501.

Chavarriaga-Aguirre, P. and Halsey, M. (2005). Cassava (*Manihotesculenta* Crantz): Reproductive biology and practices for confinement of experimental field trials. Report prepared for the Program for Biosafety Systems. Washington, D.C. Program for Biosafety Systems. 25 pp.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142: 169 - 196.

Dellaporta, S. L., Wood, J. and Ticks, J. B. (1983). A plant molecular DNA minipreparation version 2. *Plant Molecular Biology Reporter* 1: 19 - 22.

Dongre, A. and Pakri, V. (2005). Identification of cotton hybrids through the combination of PCR based RAPD, ISSR and microsatellite markers. *Journal Plant Biochemistry Biotechnology* 14: 53 – 55.

Evans, K. M., Patocchi, A., Rezzonico, F., Mathis, F., Durel, C. E., Fernández-Fernández, F., Boudichevskaia, A., Dunemann, F., Stankiewicz-Kosyl, M., Gianfranceschi, L., Komjanc, M., Lateur, M., Madduri, M., Noordijk, Y. and Van de Weg, W. E. (2011). Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. *Molecular Breeding* 28(4): 535 – 547.

FAO (2004). Food and Agriculture Organization's, Global Cassava Development Strategy [http://www.fao.org/ag/agp/agpc/gcds/index_en.html] site visited on 18/5/2013.

Fen, L. J., Guo-Bin, M. and Ling, X. (2008). SSR markers for identification of purity of melon hybrids. Chin. J. *Agriculture. Biotechnology*. 5(3): 223 - 229.

Gomez, S. M., Denwar, N. N., Ramasubramanian, T., Simpson, C. E., Burow, G., Burke, J. J. and Puppala, N. (2008). Identification of Peanut Hybrids Using Microsatellite Markers and Horizontal Polyacrylamide Gel Electrophoresis. *Peanut Science* 35: 123 – 129.

Kapinga, R., Mafuru, J., Jeremiah, S., Rwiza, E., Kamala, R., Mashamba, F. and Mlingi, N. (2005). *A Review of Cassava in Africa with Country Case Studies on Nigeria, Ghana, the United Republic of Tanzania, Uganda and Benin.* Status of Cassava in Tanzania. FAO and IFAD, Rome, Italy. 24pp.

Kawka, M., Horbanczuk, J. O., Sacharczuk, M. and Zieba, G. (2007). Genetic characteristics of the Ostrich population using molecular methods. *Poultry Science* 86: 277 – 281.

Khasa, D. P., Nadeem, S., Thomas, B., Robertson, A. and Bousquet, J. (2003). Application of SSR markers for parentage analysis of *Populus*clones. *Forest Genetics* 10(4): 273 – 281.

Li, Y., Wongprasert, K., Shekhar, M. and Ryan, J. (2007). Development of two microsatellite multiplex systems for black tiger shrimp *Penaeusmonodon* and its application in genetic diversity study for two populations. *Aquaculture* 266: 279 – 288.

Lin, X. C., Lou, Y. F., Liu, J. and Peng, J. S. (2010). Crossbreeding of *Phyllostachys* species (Poaceae) and identification of their hybrids using ISSR markers. *Genetic Molecular Research* 9: 1398 – 1404.

Mba, R. E. C., Stephenson, P., Edwards, K., Melzer, S., Mkumbira, J., Gullberg, U., Apel, K., Gale, M., Tohme, J. and Fregene, M. (2001). Simple sequence repeat (SSR) markers survey of the cassava (*Manihotesculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theory Applied Genetic* 102: 21–31.

Mbanjo, E. G. N., Tchoumbougnang, F., Mouelle, A. S., Oben, J. E., Nyine, I. M., Dochez, C., Ferguson, M. E. and Lorenzen, I. J. (2012). Development of expressed sequence tags-simple sequence repeats (EST-SSRs) for *Musa* and their applicability in authentication of a *Musa* breeding population. *African Journal of Biotechnology* 11(71): 13546 – 13559.

Mbanzibwa, D. R., Tian, Y. P., Mukasa, S. B. and Valkonen, J. P. (2009). Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology* 83: 6934 – 6940.

McCouch, S., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y. G., Huang, N., Ishii, T. and Blair, M. (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology* 35(2): 89 – 99.

Ntawuruhunga, P. and Legg, J. (2007). New spread of Cassava Brown Streak Virus Disease and its implications for the movement of cassava germplasm in the east and central African region. Report prepared for Crop Crisis Control Project. 3pp.

Powell, W., Machray, G. and Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Science* 1: 215 – 222.

Rabello, E., Nunes de Souza, A., Saito, D. and Tsai, S. M. (2005). *In silico* characterization of microsatellites in *Eucalyptus* spp: abundance, length variation and transposon associations. *Genetic Molecular Biology* 28(3): 582 – 588.

Reddy, M. P., Sarla, N., Siddiq, E. A. (2002). Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica* 128: 9 – 17.

Sartie, A. and Asiedu, R. (2011). Development of mapping populations for genetic analysis in yams (*Dioscorearotundata* Poir. *and Dioscoreaalata L.*). *African Journal of Biotechnology* 10(16): 3040 - 3050.

Takrama, J. F., Cernantes-Martinez, C., Phillips-Mora, W., Brown, J. S., Motamayor, J. C. and Schnell, R. J. (2005). Determination of off-types in a cacao breeding programme using microsatellites. *INGENIC Newsletter* 10: 2 - 8.

CHAPTER FOUR

A High Density SNP based Genetic Map of Cassava from AR40-6 X Albert population

4.0 Abstract

Cassava suffers from a threatening cassava brown streak disease (CBSD) which affects quality and quantity of its storage roots. To increase quality and productivity of the cassava crop, genetic improvement is essential. Single nucleotide polymorphism (SNP) markers provide a powerful molecular tool for construction of genetic linkage map. In this study, a high density SNP based genetic map is presented. High throughput Genotyping by sequencing (GBS) approach was used to sequence 151 F_1population from a cross between AR40-6 x Albert cultivars. A genetic linkage map consisted of 4784 SNP markers encompassing 4250.4 cM, distributed on 19 linkage groups was constructed. The mean distance between markers was 1.26 cM, and the map covers 815 scaffolds on cassava reference genome sequence. The genetic map had 3449 SNP markers in integrated map, 2159 and 2169 markers in female and male map respectively. The first high density SNP based genetic linkage map developed here in cassava will lay a solid foundation for several applications in the genetic improvement of cassava including the QTL mapping, marker-assisted breeding and comparative genomic studies. The genetic map obtained here can provide the basis for mapping and analysis of genes/QTL related to cassava brown streak disease resistance and development of a high density consensus map for SNPs in cassava.

4.1 Introduction

Cassava (*Manihotesculenta* Crantz) is a perennial crop originating from Latin America with a diploid chromosome 2n=36. The DNA of cassava has 1.67 picograms (pg) per cell with 772 mega-base pairs (Mbp) in the haploid genome (Awoleye *et al*., 1994). Cassava is a food crop for more than 800 million people in tropical and subtropical Africa, Asia and Latin America (Utsumi *et al.,* 2011, FAOSTAT 2010). The main value of the crop is in its storage roots with dry matter containing more than 80% starch (FedBase.com). Roots of cassava are either consumed fresh when they are low in cyanogens or in many processed forms including flour and for animal feed. Cassava can also be used for agro-industrial processing such a starch and biofuel. Subsistence farmers are the main growers of cassava crop in Africa. The current global cassava production is more than 240 million tons a year, of which more than 50% occurs in Africa, Asia contributes about 33% and America 15% (FAOSTAT 2010). Globally Tanzania is the sixth largest producer of cassava with annual production of fresh roots ranging from 5.4 to 7 million tonnes during 2008 and 2010. It is the largest producer of cassava in East Africa (FAO, 2011).

Cassava brown streak disease (CBSD) is currently a major threat to cassava production in Tanzania and in East Africa region. The disease is caused by two distinct virus *Cassava brown streak virus* (CBSV) and *Uganda Cassava brown streak Virus* (CBSUV) (Monger *et al.,* 2010; Winter *et al.,* 2010). The complexity of the viruses and their symptoms make breeding for resistance difficult. If quantitative trait loci (QTL) associated with resistance can be tagged, then molecular breeding can be applied leading to more informed and accurate selection of resistant genotypes. Genetic maps are basis for identification of QTL related to CBSV resistance. DNA based markers such as Single Nucleotide polymorphisms (SNPs) are useful in developing a map which can be used for QTL

mapping. Single nucleotide polymorphisms, or SNPs, are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered.

Genetic linkage maps have facilitated the identification of markers associated with many traits of importance. In cassava, genetic linkage maps developed by Fregene (1997) using a population of TMS30572 X CM2177-2 assisted in identification of markers associated with cassava bacterial blight (Lopez *et al*., 2007; Wydra *et al.,* 2004) and markers associated with plant architecture and yield (Okogbenin *et al*., 2003). SSR based genetic linkage maps developed by using Huay Bong 60 x Hanatee F1 population (Whankeaw *et al.,* 2011; Sraphet *et al.,* 2011; Kunkeaw *et al.,* 2011) assisted in identification of marker associated with Cynogenic Potential (Whankaew *et al.,* 2011). Other cassava genetic linkage maps have been constructed by Chen *et al.* (2010) using EST-SSRs, SSRs and AFLPs. The first SNP based map (Rabbi *et al.,* 2012) was developed from a full-sib family (F1) from a cross between two Tanzanian varieties, Namikonga (CBSD resistant) and Albert (CBSD susceptible). The map consisted of 19 linkage groups, encompassing 568 markers (134 SSRs and 434 SNPs) spanning 1837 cM with an average density of 3.4 cM.

This study aimed at developing a genetic linkage map from an F1 population with SNPs markers. The map intended to use more than 1000 SNPs using advanced techniques of Genotyping-by-Sequencing (Elshire *et al.,* 2011). This is the highest-density linkage map so far published for cassava that will facilitate anchoring and ordering of the reference genome sequences and hence providing important tools for basic and applied genomics in assisted breeding for cassava. The map will also be used for QTL identification for CBSD resistance.

4.2 Materials and Methods

4.2.1 Mapping population

The mapping population used to construct the genetic linkage map consisted of 151 F1 genotypes. The genotypes were derived from an intra-specific cross between two cassava cultivars; "AR40-6" (female parent) an elite cassava cultivar originating from the Centro Internacional de Agricultura Tropical (CIAT) in Colombia and "Albert" (male parent) a local landrace originating from Mtwara, southern Tanzania. Mapping population was selected based on studies for genetic integrity of F_1 progenies in chapter three of this dissertation. AR40-6 is one of the best sources of resistance to Cassava Brown Streak Disease while Albert is a cultivar, susceptible to Cassava brown streak disease (CBSD) and has high flowering ability making it a good pollen donor in the crossing programme. Both are tolerant to cassava bacterial blight (CBB) and Cassava Mosaic Disease (CMD).

4.2.2 SNP genotyping

SNP genotyping was carried out using 'Genotyping by Sequencing (GBS)' (Elshire *et al.,* 2011) that makes use of a Next-generation sequencing platform. The GBS yielded a total of 5381 SNPs markers across the population of 151 F1 individual and their parents AR40-6 and Albert. Genotyping was done by using 96-plex protocol according to Elshire *et al.* (2011). In summary; Genomic DNA is first quantified using a fluorescence –based method followed by normalization in a 96-plate. A master mix with ApeKI digest enzyme and buffer is added to the plate and incubated with DNA to generate genomic fragments. ApeKI is a methylation sensitive restriction endonuclease that recognizes a degenerate 5bp motif (GCWGC, where W is A or T), creates a 5' overhang of 3bp and has few recognition sites. Barcoded adapters are added along with ligase and ligation buffers to each sample after digestion. The purpose of the barcode is to facilitate multiplexing, and the purpose of the adapter is to provide a template for PCR primers, to enable attachment

to the flow cell on the sequencer, and to initiate amplification during the sequencing process. Equal volumes of each sample are then pooled and cleaned and size fractioned on an agarose gel. Fragments of size 400bp – 800bp are excised. This provides a greater depth of reads hence increasing the confidence of SNP calling. Samples are then taken for amplification in PCR. The library is then sequenced using 100bp pair end reads Illumina HiSeq 2000. Following a sequencing run, FASTQ files containing raw data from the run are used to analyze sequencing reads to samples using the DNA barcode sequence. Once assigned to individual samples, the reads are aligned to a reference genome sequences. Filtering algorithms are used to distinguish true bialellic SNPs from sequencing errors. GBS provides output SNP marker genotyping data with loci nomenclature as number of scaffold in reference genome and position of specific SNP in the scaffold (e.g. s0247:61142). This approach made easy foridentification of unique scaffolds that were used in linkage groups without searching in at Phytozome (cassava genome website http:// www.phytozome.net/ cassava).

4.2.3 Linkage analysis

Marker genotyping data was cleaned prior to linkage analysis. Markers with high proportion of missing data (>10%) were removed from the dataset using Microsoft Excel.). Genotypic data was scored based on study population model *Cross Pollinator*, (CP) described by the JoinMap® version 3.0 (Van Ooijen and Voorrips 2001). Markers segregation was denoted as *lm* and *ll* for markers segregating in female parent AR40-6, *nn* and *np* for markers segregating in male parent Albert lastly *hh, hk* and *kk* for markers segregating in both parents (anchor markers). In a CP population, the Chi-square tests of the Locus Genotype Frequencies were classified according to genotype classes (Table 7). A total of 4787 SNP loci remained after data cleaning and were loaded into the software for linkage analysis. In JoinMap, data was calculated according to individual genotype

frequency and nine individual were found to have large proportion of missing loci (>10%) and were excluded in analysis. Identical markers and individuals were analyzed and identified for observation using Similarity of loci and Similarity of individual Joinmap function. Marker genotyping data were calculated by using chi-square test in Locus genotype frequency (JoinMap function) to assess the goodness-of-fit according to Mendelian Segregation ratio. Loci showing highly significant deviation from Mendelian segregation ratio (e.g $P \leq 0.01$) were excluded from subsequent mapping analysis. Other markers which showed slight deviations were maintained for observation to see if their inclusion could inflate the map distance.

Table 7: Classification type code; Ratio is the expected Mendelian segregation ratio

Code	Ratio	Classification into genotype classes
(hh, hk, kk)	1:2:1	hh, hk and kk
(ll, lm)	1:1	ll and lm
(nn, np)	1:1	nn and np

Source: JoinMap®4

4.2.4 Linkage mapping

A total of 4728 SNPs were used to construct genetic linkage groups using JoinMap 4.1 (Kyzma, Wageningen, and The Netherlands). Markers in the population were assigned to groups based on recombination frequency using the independence LOD test Regression mapping for map construction was set as follows: Recombination frequency > 0.4 and Goodness-of-fit jump for removal of loci = 5.0. Map units (centMorgans) were determined using the Kosambi mapping function (1944) and threshold level of the test (LOD) for map calculation ranged from 4 to 10. Linkage maps were constructed using identified linkage groups LOD 5.0 to 8.0. Three maps were generated. A female map denoted 'P1'

consisting of markers that segregate in the female ie. *lm* x *ll* and *hk* x *hk*. A male map denoted as 'P2' consisting of markers segregating in the male parent ie. *nn* x *np* and *hk* x *hk*, and an integrated map denoted as 'I' consisting of all markers in female and male map. During map construction markers that were suspected to increase map distance were manually checked and removed in order to assure quality and accuracy of the genetic map. In this study, high density genetic map was referred as Comprehensive integrated map (CM) because of large number of SNP markers (4728) used during construction.

4.2.5 Framework map

Another set of genetic linkage maps was constructed using the same methodology as the one used to construct high-density linkage maps but using greater stringency in the choice of SNPs to be included. The data set used to develop groupings was filtered to remove SNP sites on the same locus so that only one SNP site represented a locus. A number of SNP sites were excluded including; SNP sites that showed similarity value of 1.00 (identical locus), SNP sites with large proportion of missing data (>10%) and SNP sites that showed highly distortion from Mendelian segregation ratio. The data was narrowed down using binning approach which excluded SNP sites from the same parental genotypes encoding (*lm, hk, np*) within 150kb bins. In individual genotype frequency (JoinMap function), seven individuals showing high proportion of missing loci (>15%) were excluded before groupings. A total of 2041 SNP sites and 144 individuals remaining were used for groupings. An independence LOD threshold for map calculation ranged from 4 to 10 and linkage maps were constructed using identified linkage groups with LOD 4.0 to LOD 8.0.

4.3 Results

4.3.1 Markers segregation

A total of 4787 SNP markers were used for studying segregation analysis of these 40% (1921) markers were observed to segregate in male parent Albert *(nnxnp)* and 36% (1703) of markers in the female parent AR40-6 *(llxlm)*. About 1163 (24%) of markers segregated in both parents *(hkxhk)*. Marker segregation type was not evenly distributed across the linkage groups (Fig. 4). Linkage group four (LG4) had the lowest number of markers segregating in male parent and the highest number of markers (105) in LG8.

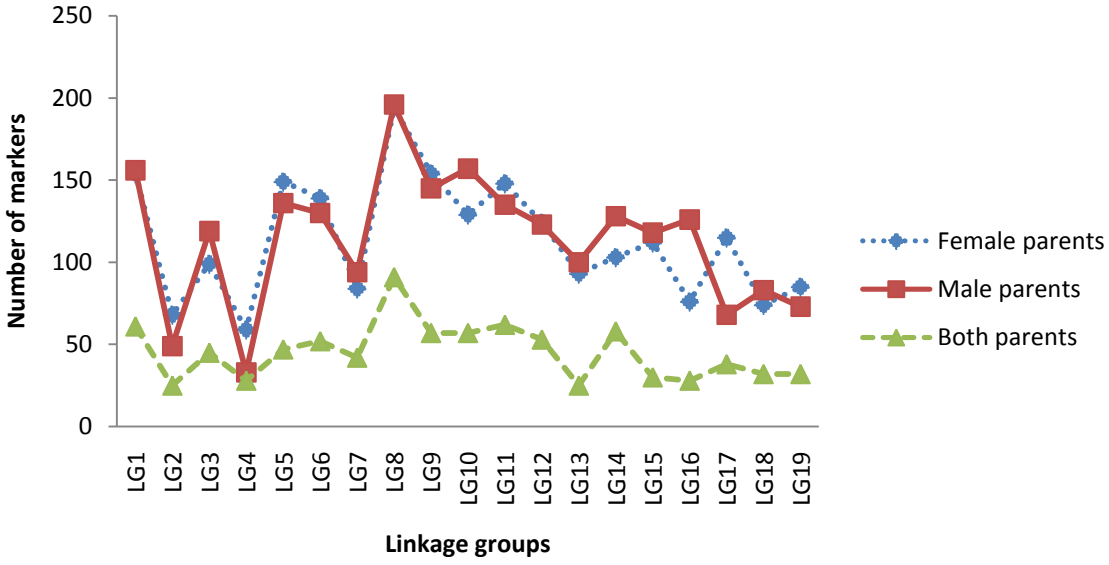


Figure 4: Distribution proportion of number of markers segregating in female parent (*ll*x*lm*), male parent (*nn*x*np*) and in both parents (*hk*x*hk*) across cassava linkage groups from cassava F_1 population (AR40-6 X Albert)

Segregation distortion was evaluated by using chi-square test classification type code (Table 7) where 82% (3928) of loci concurred with the expected Mendelian Segregation ratios of 1:1 and 1:2:1. The results showed a total of 59 (1 %) markers with highly

distorted segregation of highly significance level (p = 0.0001). These markers were excluded from the analysis. From the observation all highly segregating markers had *"hk"* genotype classes which indicated that parents were "*hh"* and *"kk"* hence no segregation occurred in progenies and their inclusion in the initial data set was due to SNP calling errors. The remaining 800 (17%) of loci with low to moderate segregation distortion ($0.1 \leq P \geq 0.0005$) were included in genetic linkage map construction. Interestingly, all markers with segregation distortion of significance level less than p=0.0005, occurred in Female linkage groups LG 3, LG8, LG9, LG14 and LG19 and varied from 1 to 4. Linkage analysis results revealed that markers with low to moderate deviation from Mendelian segregation ratio had no effect on the map position marker order and were distributed evenly across the linkage groups.

4.3.2 Markers in linkage groups

In the integrated comprehensive genetic map, 4784 SNPs markers were allocated in 3449 loci along 19 linkage groups. The results of linkage groups are presented in figure 6 as descriptive for linkage group one (LG1), then LG2-LG19 are presented in appendix figures I. In the linkage groups, scaffolds refer to overlapping fragments of DNA and gaps of known length. The results revealed that, genetic map (LG1-LG19) had an average of 114 markers in female and male map respectively. An average of 118 markers was observed in the integrated map. In total, identical loci encompassed 26.3% of the total marker genotype data. The result showed that among two identical markers, one marker was included in the group and its identical marker not included in linkage groups construction. The results also showed 100 identical loci from the same scaffold but differed in position of SNPs. A total of twenty markers were removed because they inflated the distance of the map.

4.3.2.1 Integrated comprehensive genetic map

The integrated Comprehensive genetic map (CM) consisted of 19 linkage groups which varied in length from 88.5cM (LG4) to 358.6cM (LG8) (Fig. 5). The short linkage group (LG4) has few numbers of markers (64) and the longest group (LG8) has largest number of markers (299) from a total of 3449 markers in the integrated maps. Single Nucleotide Polymorphism markers were extensively distributed across the genome with high correlation between linkage distance and number of markers (r = 0.91, P = 0.05). However, LG5 and LG11 appeared to have a greater number of markers in relation to their distance which indicated clustering of markers in the same loci position and greater confidence in their map position. Marker clustering was observed in the integrated comprehensive genetic map where by 2 to 12 markers were clustered in the same locus (Fig.6). Eleven markers clustered together at locus 43.81cM. The largest cluster of markers was observed in LG19 where 12 markers were located at 159.75cM (Fig. 8).

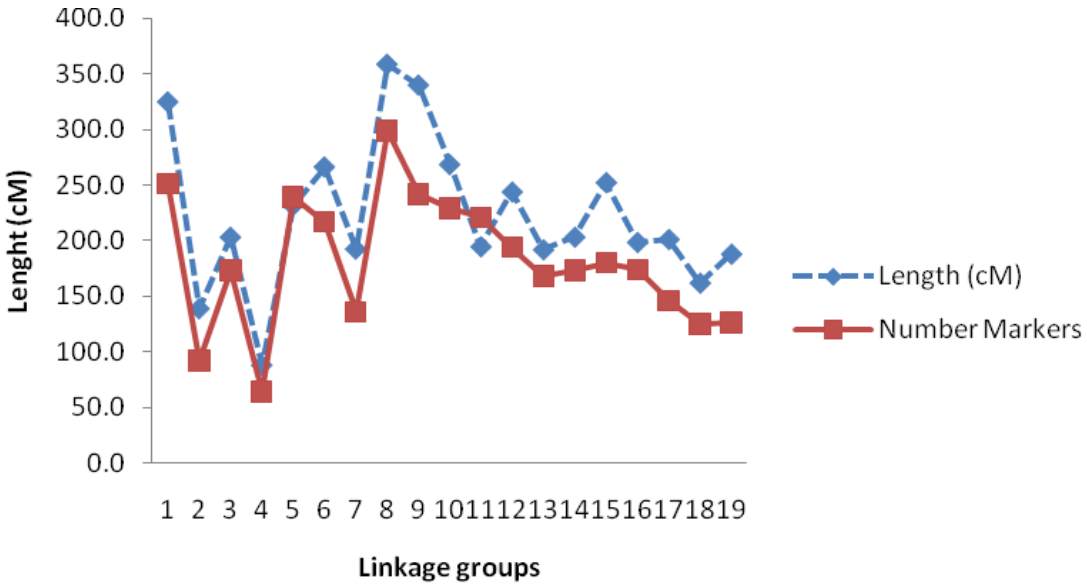


Figure 5: Relationship between map length and number of markers in 19 linkage groups of the comprehensive integrated map (CM)

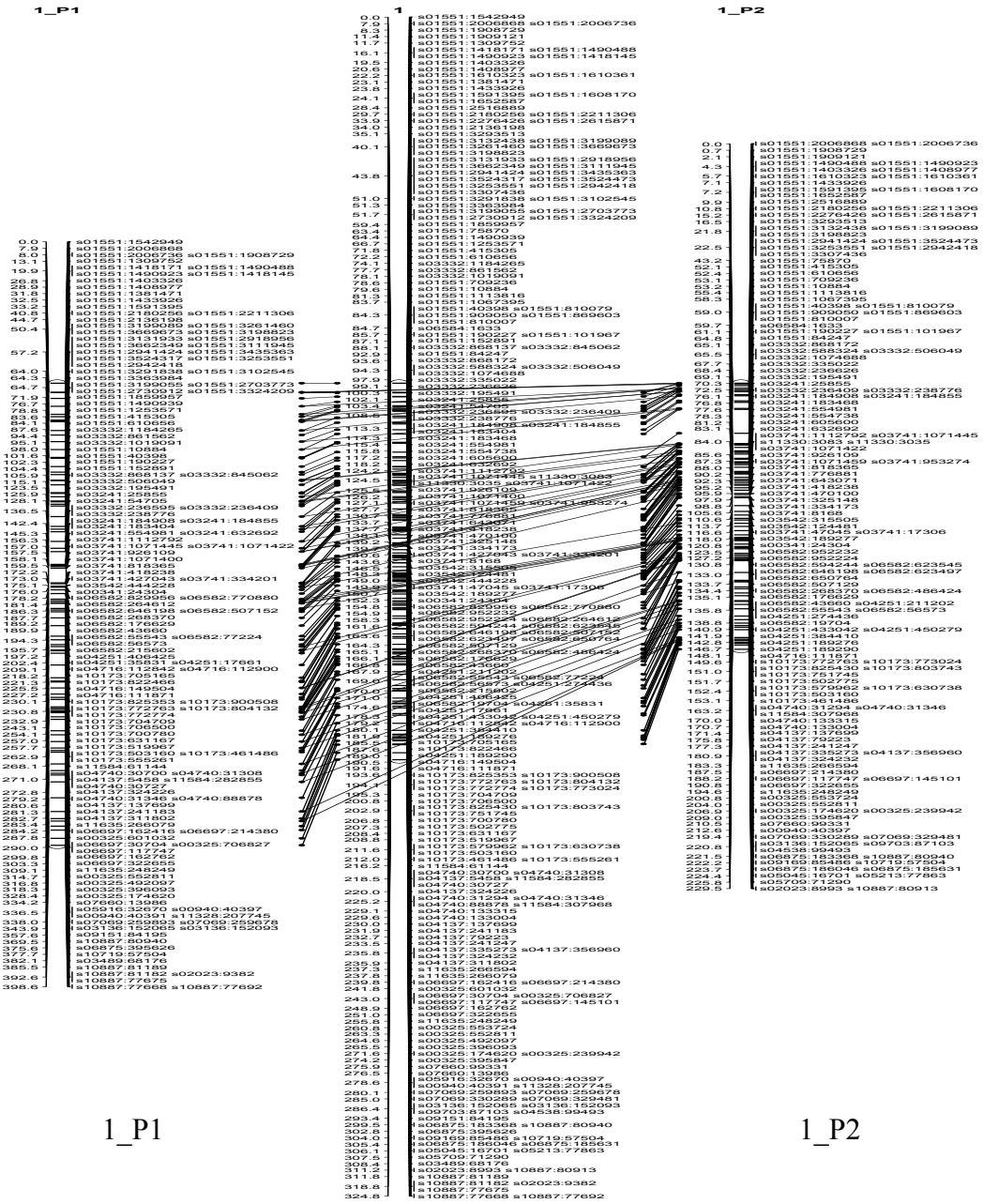


Figure 6: An alignment of integrated comprehensive genetic map of cassava derived from AR40-6 x Albert F1 population. Total number of markers for LG1 is 251. The maps on the left (designated by P1) are female maps, on the right (designated by P2) are male maps and at the Centre integrated maps combining markers in female and male maps. Arrows represents homologous markers between female map, integrated and male maps. Scaffold name (designated by s followed by 5-digit identity) is followed by SNP name of the right and on the left bar are distance between markers.

Note: For illustration purpose LG1 is presented here, however the remaining 18 LG are presented in Appendix figures I (page 99 – 116)

Average marker interval per linkage group ranged from 0.88 to 1.51 with an average marker interval across all linkage groups of 1.26 cM (Table 8). Unique loci ranged from 40 (LG4) to 200 (LG8) with an average distance of 2.00 cM in all nineteen linkage groups. The results showed one to two gaps of more than 15 cM in LG3, LG7, LG11, LG13, LG17, LG18 and LG19 respectively. The largest gap of 29.4 cM was observed in LG13.

Table 8: Summary of linkage groups, length, marker distribution, average marker density, unique loci, average unique loci and number of gaps greater than 15 cM in Comprehensive integrated map.

Linkage	Length (cM)	Number of Markers	Average marker density (cM)	Unique loci	Average unique loci (cM)	No. gap>15cM
LG1	324.8	251	1.29	155	2.1	0
LG2	139.2	92	1.51	59	2.36	0
LG3	203.3	173	1.17	99	2.05	1
LG4	88.5	64	1.38	40	2.21	0
LG5	231.7	239	0.97	138	1.68	0
LG6	266.4	217	1.23	133	2	0
LG7	192.7	136	1.42	81	2.38	1
LG8	358.6	299	1.2	200	1.79	0
LG9	339.9	242	1.4	193	1.76	0
LG10	268.6	229	1.17	127	2.11	0
LG11	194.8	221	0.88	108	1.8	1
LG12	243.9	194	1.26	121	2.02	0
LG13	192	168	1.14	95	2.02	1
LG14	203.4	173	1.18	110	1.85	0
LG15	252.3	180	1.4	127	1.99	0
LG16	198.6	174	1.14	106	1.87	0
LG17	201.3	146	1.38	91	2.21	1
LG18	162.2	125	1.3	72	2.25	1
LG19	188.1	126	1.49	126	1.49	2
Total	**4250.4**	**3449**	**23.91**	**2181**	**37.94**	**8**
Average	**223.7**	**182**	**1.26**	**115**	**2.00**	

4.3.2.2 Female parent genetic linkage map

A summary of length, number of markers, number of unique loci, average marker interval, average of unique loci and number of gaps with greater than 15 cM in each linkage group

of the AR40-6 genetic map are presented (Table 9).The AR40-4 map consisted of 2159 SNP markers on 19 linkage groups. The AR40-6 linkage map encompassed 4466.3 cM with an average marker interval of 2.1 cM and average cM per unique locus of 3.4 cM. The length of the linkage groups ranged from 114.2 cM of LG4 to 430.8 cM of LG8. This was longer than the integrated (4250.4cM) and male maps (3702.5cM). The number of markers per linkage group varied from 59 to 194. The largest gap between markers was 52.2 cM in LG13. Each linkage group had loci with more than one marker however LG8 had a maximum of 13 markers at loci 26.98cM.

Table 9: Summary of linkage groups, length, marker distribution, average marker interval, unique loci, average unique loci and number of gaps greater than 15 cM in Female derived map

	Female Map (AR40-6)					
Linkage	**Length (cM)**	**Number of Markers**	**Average Marker Density (cM)**	**Unique loci**	**Average Unique loci (cM)**	**gap > 15 cM**
LG1	398.6	154	2.59	98	4.07	0
LG2	147.7	68	2.17	44	3.36	1
LG3	212.2	99	2.14	58	3.66	2
LG4	114.2	59	1.94	36	3.17	0
LG5	239.9	149	1.61	90	2.67	0
LG6	334.1	139	2.4	86	3.88	1
LG7	186.4	84	2.22	43	4.33	2
LG8	430.8	194	2.22	134	3.22	2
LG9	355	154	2.31	103	3.45	0
LG10	280.4	129	2.17	103	2.72	0
LG11	203.9	148	1.38	71	2.87	1
LG12	240.1	124	1.94	68	3.53	0
LG13	209.4	93	2.25	56	3.74	1
LG14	156.3	103	1.52	69	2.27	0
LG15	271.8	112	2.43	78	3.48	1
LG16	188	76	2.47	47	4	3
LG17	209.2	115	1.82	60	3.49	1
LG18	157.8	74	2.13	36	4.38	1
LG19	130.5	85	1.53	43	3.03	0
Total	**4466.3**	**2159**	**39.24**	**1323**	**65.32**	**16**
Average	**235.1**	**114**	**2.07**	**70**	**3.44**	

4.3.2.3 Male parent genetic linkage Map

The male map consists of 2159 SNP markers on 19 linkage groups spanning 3701.5 cM, with an average marker interval of 1.8cM (Table 10). The markers are located on 1314 loci with an average space of 2.9 cM ranged from 2.4 cM of LG16 to 3.7 cM of LG19. Linkage group nine (LG9) was the longest group with 145 markers spanning 299.8 cM and LG8 was had the largest number of markers (196), spanning 285.6 cM. The shortest group (61.3 cM) had the lowest number of markers in the map (33). Albert genetic map has the highest number of gaps (> 15cM) between markers in all of the groups. The largest gap distance of 48.3 cM appeared in LG17 while LG 19 has three gaps 18.2cM, 17.2cM and 26.3 cM respectively. Single locus mapping occurred in LG8 whereby 16 markers mapped at locus162.65cM.

Table 10: Summary of linkage groups, length, marker distribution, average marker interval, unique loci, average unique loci and number of gaps greater than 15 cM in male derived map

	Male Map (Albert)					
Linkage	**Length (cM)**	**Number of Markers**	**Average. Marker density (cM)**	**Unique loci**	**Average Unique loci (cM)**	**gap > 15 cM**
LG1	229.5	156	1.47	97	2.37	1
LG2	110.2	49	2.25	30	3.67	0
LG3	174.8	119	1.47	68	2.57	2
LG4	61.3	33	1.86	20	3.07	0
LG5	208.2	136	1.53	80	2.6	1
LG6	198.8	130	1.53	73	2.72	1
LG7	199.1	94	2.12	64	3.11	1
LG8	285.6	196	1.46	115	2.48	0
LG9	299.8	145	2.07	109	2.75	0
LG10	252.8	157	1.61	89	2.84	0
LG11	173.6	135	1.29	69	2.52	2
LG12	242.3	123	1.97	81	2.99	0
LG13	162.5	100	1.62	61	2.66	1
LG14	182.2	128	1.42	67	2.72	1
LG15	223.6	118	1.9	72	3.11	0
LG16	178.4	126	1.42	73	2.44	0
LG17	179.7	68	2.64	49	3.67	1
LG18	164.4	83	1.98	50	3.29	2
LG19	175.8	73	2.41	47	3.74	3
Total	**3702.5**	**2169**	**34.02**	**1314**	**55.32**	**16**
Average	**194.9**	**114**	**1.79**	**69**	**5.53**	

4.3.3 Correlation between parental maps

Relationship between AR40-6 (female) parent map and Albert (male) parent indicated significantly difference in size of the maps, slightly difference in number of markers, marker interval and distinct positions of loci in relation to the markers embedded in such loci for all nineteen linkage groups. Female map had the longest length of 4466.3 cM (Table 9) compared to male map which had 3702.5cM (Table 10). Male map has 10 more number of markers and almost 10 less number of distinct position of markers (loci).

Number of homologs markers (shared markers between female and male parent) are large (91) in linkage group 8 and lowest (25) in linkage group 13 (Table 11).

Table 11: Summary of number of recombinant markers in female parent, male parent and homologs

	Number of Markers		
Linkage group	Female	Male	Homologs
LG1	154	156	61
LG2	68	49	25
LG3	99	119	45
LG4	59	33	28
LG5	149	136	47
LG6	139	130	52
LG7	84	94	42
LG8	194	196	91
LG9	154	145	57
LG10	129	157	57
LG11	148	135	62
LG12	124	123	53
LG13	93	100	25
LG14	103	128	58
LG15	112	118	30
LG16	76	126	28
LG17	115	68	38
LG18	74	83	32
LG19	85	73	32
TOTAL	**2159**	**2169**	**863**

4.3.4 Framework genetic linkage map

A total of 2041 SNP markers were used to construct a framework genetic map of 18 linkage groups spanning 3668 cM (Appendix 2). The map consisted of Female, Male and integrated linkage maps with an average of 63markers in female, 70 markers in male and 109 markers in integrated map respectively. Average distance ranges from 1.5cM/marker to 4.9cM/marker with an average of 2.6 cM/marker. A total of 86% of the 2041 (Table

12) SNPs used for linkage analysis in the Framework map showed the expected Mendelian segregation ratio and 13.6% showed slight distortion with significant levels $0.01 \leq P \leq 0.1$. Observation of these markers from the map did not show any significant effect in marker order nor map distance, and were distributed evenly across the linkage groups. A high percent of segregation occurred in male <*nn x np*> with 43% and the lowest (22%) in anchor markers < *hk x hk* > (Table 12).

Table 12: Classification of segregation and number of markers in Framework Genetic Linkage map

Segregation Type	Markers
Alleles from both parents <hkxhk>	452
Alleles from female <lmxll>	712
Alleles from male <nnxnp>	877
Total	**2041**

4.3.5 Comparison of the framework and comprehensive genetic linkage maps

The Framework map had the expected 18 linkage groups for cassava, compared to 19 linkage groups in the Comprehensive maps. Although the numbering of linkage groups were different, linkage groups of Comprehensive genetic map highly corresponds with linkage groups of Framework genetic map but difference occurred in positions of markers (Table 13). The results showed that, markers in comprehensive map for LG2 and LG4 are constituted in LG3 in the framework map. Slight variations occurred in the order of markers between CM and FM genetic maps (Fig. 7). Few groups observed to have similar marker orders and locus position around the centre of chromosome but differ at one end or both ends of the group (Telomere) (Fig. 7). In average both maps showed high degree of correlation ($r = 0.99$). In other hand, markers arrangement from the framework map was identical in LG5, LG8, LG9, LG11, LG12, LG13, LG14 and LG16 of comprehensive map; however few markers showed shift of less than 30cM. In LG1, LG2, LG4, LG6,

LG7, LG10, LG15, LG17 and LG18. There were inversions in position of markers started from high ends to lower ends of the framework linkage map (Appendix 2). The results also showed some close by linked markers in the same loci stayed together in both comprehensive and framework genetic maps; example markers s01551:3131933, s01551:3 307436 s01551:2918956, s01551:3662349, s01551:2941424, s01551:352 4317 and were assigned at 43.8cM position in LG1 of comprehensive map (Fig. 6) and position 281.7cM in LG1 of the framework map (Appendix 2). Others markers were positioned separately with less than five centimorgans apart in the framework genetic map (e.g. s06697: 117747 and s06697:145101 are in position 243.05cM of CM and position 87.7cM and 83.4cM respectively in the FM.

Table 13: Relationship between comprehensive genetic map and framework and number of markers

Comprehensive Map (CM)		Framework map (FM) Map		Correlation between CM & FM
Linkage Group	No. Markers	Linkage Group	No. Markers	
1	251	1	136	0.994
2	92	3	87	0.993
3	173	14	92	0.925
4	64	3	87	0.987
5	239	2	133	0.991
6	217	4	119	0.997
7	136	5	88	0.997
8	299	6	152	0.995
9	242	7	140	0.997
10	229	8	131	0.984
11	221	9	119	0.994
12	194	10	120	0.974
13	168	17	79	0.99
14	173	11	109	0.995
15	180	12	101	0.998
16	174	13	100	0.997
17	146	18	93	0.998
18	116	15	88	0.995
19	126	16	87	
Total/Average	3440		1974	0.989

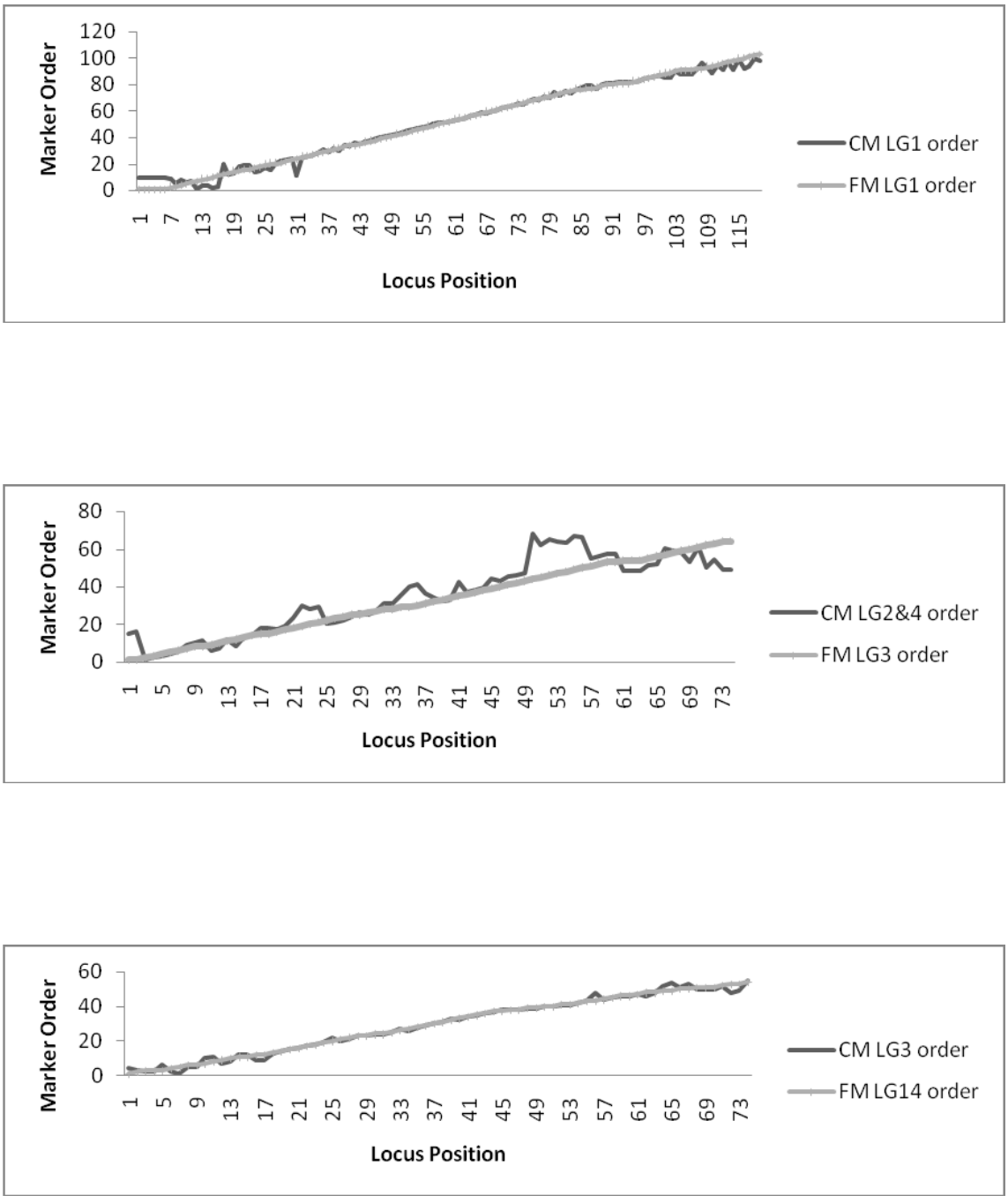
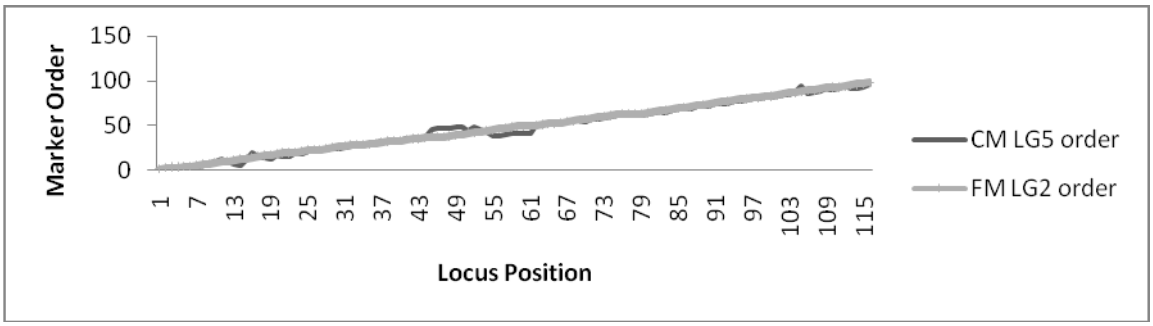
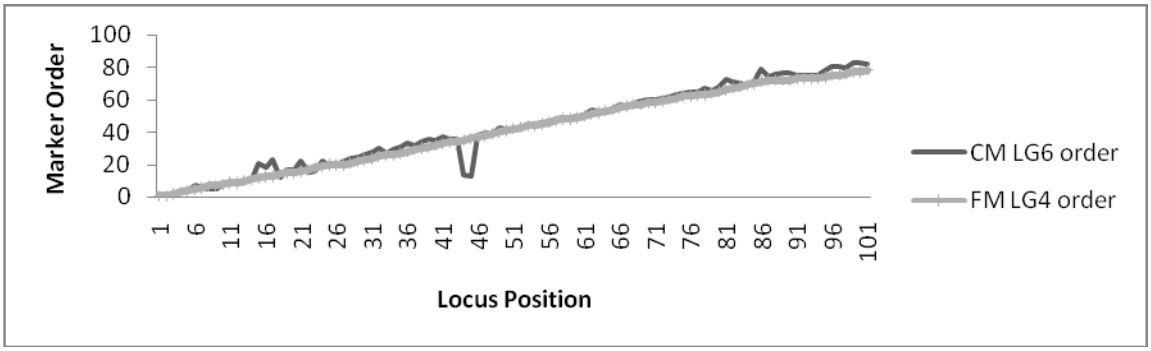
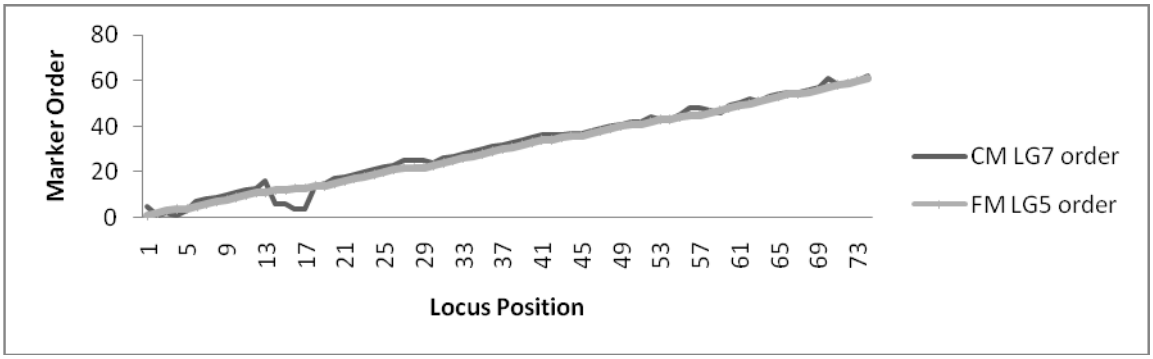


Figure 7: Relationship between nineteen linkage groups in Comprehensive genetic map (CM) and eighteen linkage groups in Framework genetic map (FM) from cassava F_1 population (AR40-6 x Albert). Order in Y-axis represents markers and Locus in X-axis represents position of markers (cM)

Figure 7 continue







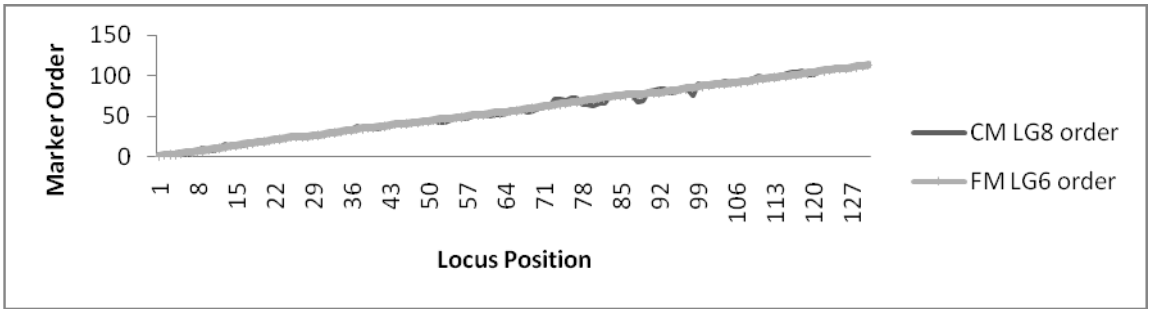
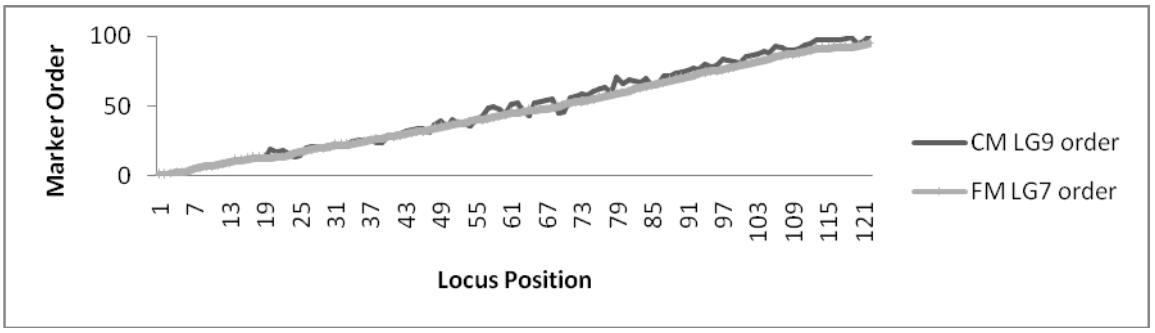
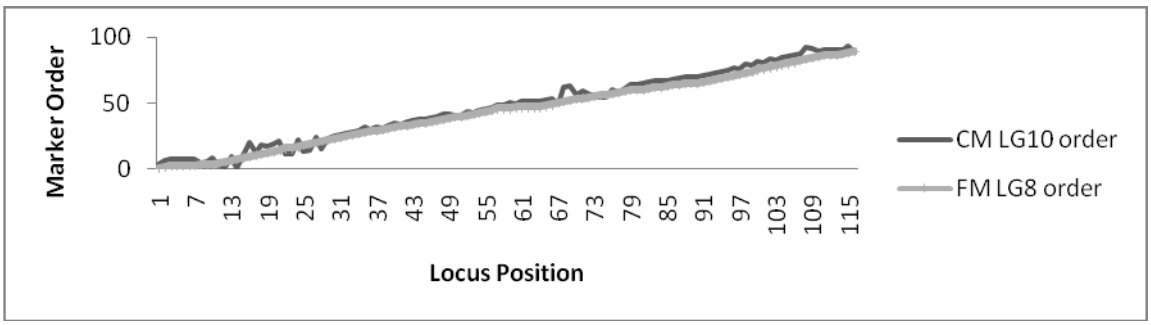
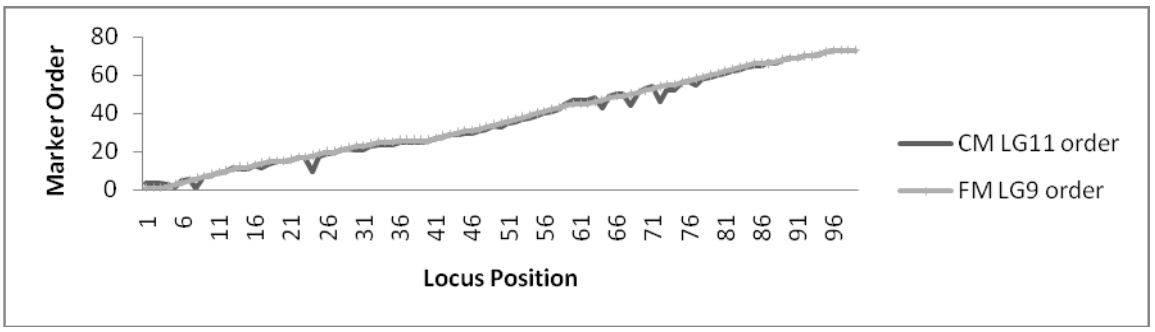


Figure 7 continue







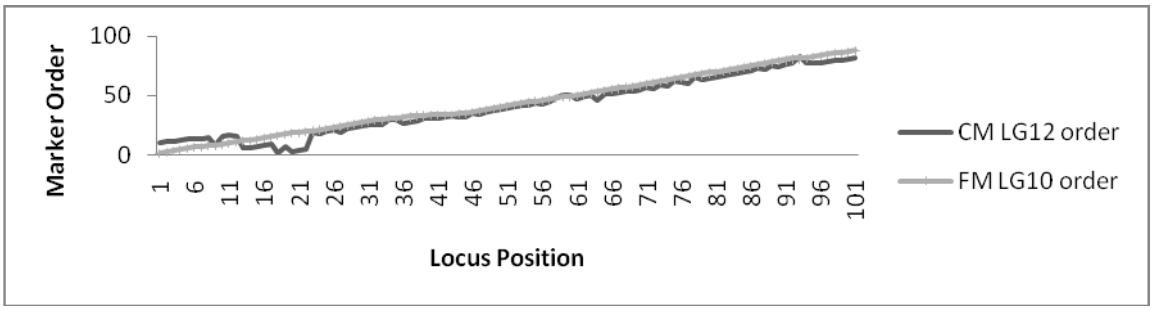
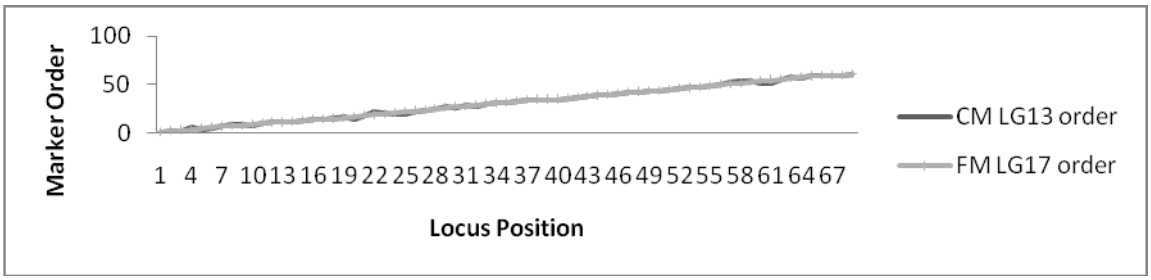
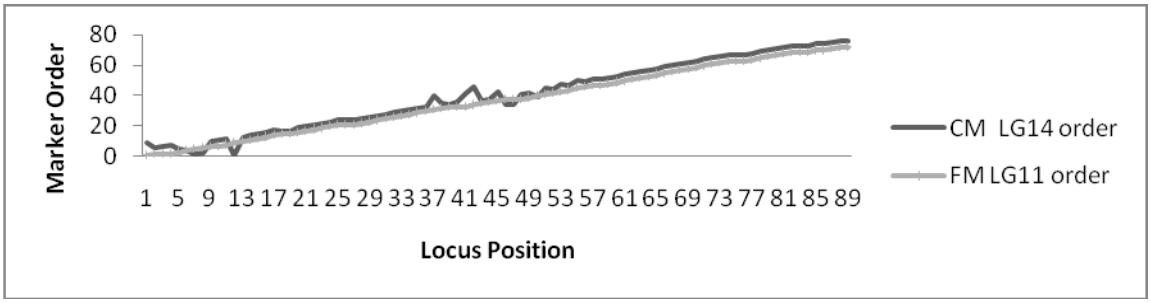
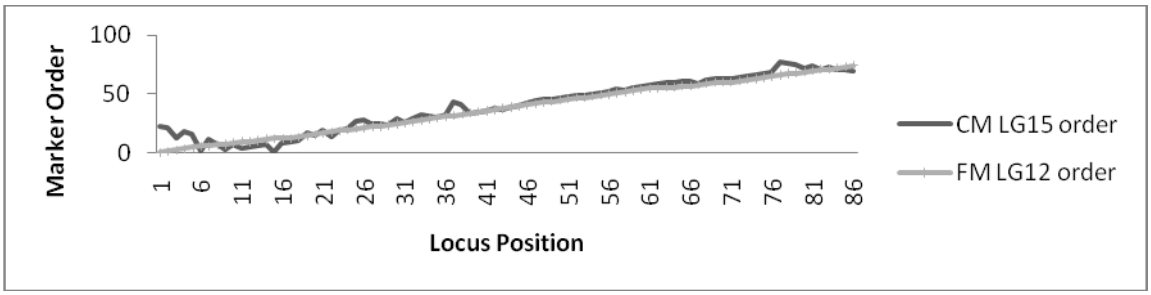
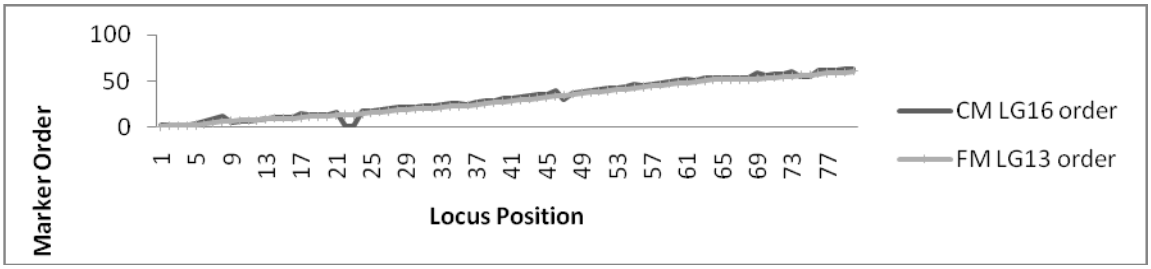


Figure 7 continue









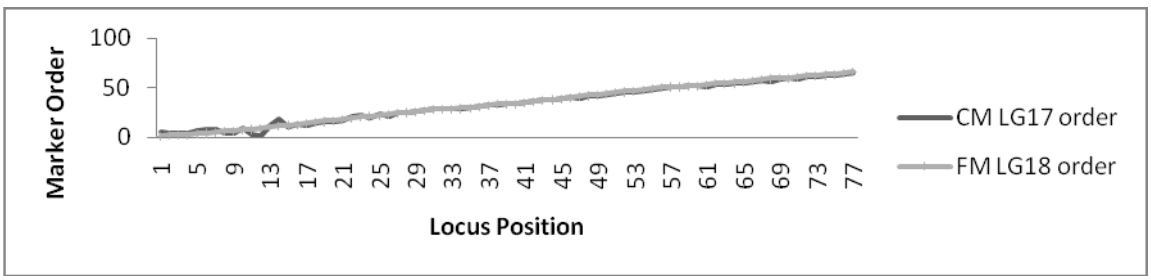
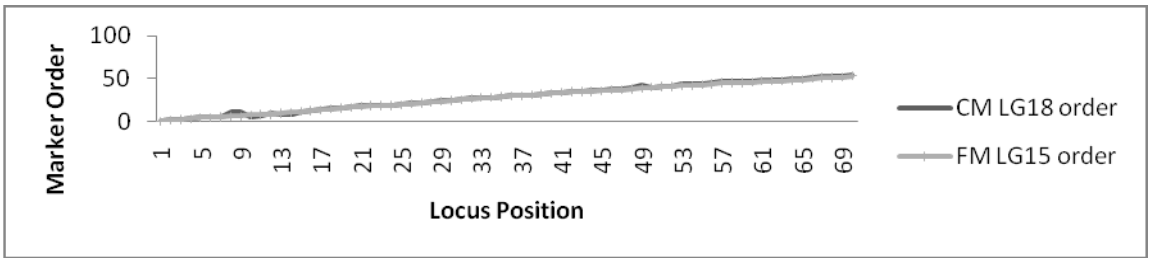
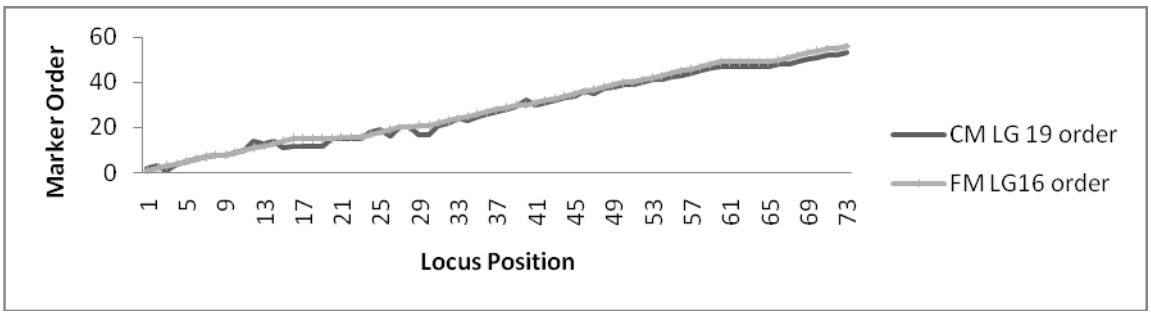


Figure 7 continue





4.3.6 Genome coverage

In this study, location of all SNP markers observed to cover 851 unique scaffolds of the cassava genome. The largest 62 and lowest number 18 scaffolds were observed in LG4 and LG10 respectively (Table 14). A unique scaffold in the table below refers distinctive marker of the known length. The unique scaffolds were demonstrated in the linkage maps (Appendix I) by arrows cross cut from P1 to P2.

Table 14: Linkage groups and number of unique scaffolds

Linkage groups	Unique scaffolds
LG1	36
LG2	29
LG3	50
LG4	18
LG5	47
LG6	44
LG7	41
LG8	49
LG9	47
LG10	62
LG11	49
LG12	44
LG13	37
LG14	61
LG15	56
LG16	44
LG17	40
LG18	54
LG19	43

4.4 Discussion

Advanced development of sequencing technologies based on DNA markers and availability of discovered SNPs markers, make population genotyping indispensable need in crop improvement. In this study, the first high density SNP-based linkage map of cassava is reported. This genetic map used F1 mapping population developed from two cassava parents, one resistant and another susceptible to CBSD. Population was genotyped by using advance sequencing technology of GBS which is convenient for more number of SNPs markers (> 1000) hence increase the genome coverage. This approach has not been used before in cassava but used in other crops like maize and barley (Elshire *et al.,* 2011) and wheat and oats (Poland *et al.,* 2012) to develop high density genetic linkage maps. The GBS approach uses enzyme-based complexity reduction (APEeKI) coupled with DNA barcoded adapters to produce multiplex libraries of samples ready for NGS

sequencing (Poland *et al.,* 2012). Other sequencing approach like Illumina Golgengate approach has fixed arrays of markers and genotypes and KBioscences' KASPar platform which requires primers for each SNP are more expensive as greater number of markers used to construct genetic linkage maps increase.

Two types of genetic maps are constructed and presented in this study; the comprehensive and the framework genetic linkage maps. The framework genetic map was obtained by thinning down the number of loci from the comprehensive genetic map. The comprehensive genetic map presented in this study has the longest length of 4250.4 cM for integrated map, 4466.3 cM for female-derived map and 3702.5 cM for male-derived map in comparison to other earlier genetic maps for cassava (Rabbi *et al.,* 2012; Sraphet *et al.,* 2011; Kunkeaw *et al.,* 2011). The maps reported here have a much higher marker density of 1.3 cM/marker and greater number of markers (> 1000) which has not been reported in earlier genetics maps of the crop. Large number of markers has currently used to generate high-density genetic map in other crops like grape (Wang *et al.,* 2012) and cotton (Yu *et al.,* 2012).

4.4.1 Correlation with other cassava genetic maps

Single Nucleotide Polymorphism based genetic linkage maps developed in this study comprise of parental maps and integrated maps. An integrated cassava map containing SNP markers has been reported by Rabbi *et al.* (2012) where 434 markers (348 SNPs and 128 SSRs) were linked in 18 linkage groups spanning 1541 cM. Integrated map differs significantly with previous SNP based genetic map developed by Rabbi (2012) by number of markers, size and marker density. The map has 3449 markes spanning 4250.4 cM with marker density of 1.5 cM. Another difference occurred in sizes of linkage groups and number of markers in each group whereas the longest group had 143.43 cM (LG6-I) while

in this map the longest group had 358.6 cM (LG8). Occurrences of gaps which are greater than 20cM have been reported in several genetic maps (Sraphet *et al*., 2011, Kunkeaw *et al.,* 2010, Okogbenin *et al.,* 2006. This study reports a maximum gap of 52.2cM in LG 13 of the female map (scaffold 0247:61142 and scaffold 07901:32951) which is higher than that observed from earlier genetic map of 46.08cM in LG 11 (Kunkeaw *et al.,* 2010).

The present study showed the lowest number of markers with severe distortion (1.2%) when compared to other reported maps of cassava of which 20% and 27% were observed in genetic maps developed by Okogbenin (2006) and that developed by Fregene (1997). These markers were not used in construction of linkage groups. The reasons mainly came from genetic factors (pollen tube competition, pollen lethal, preferential fertilization, sterility and chromosome translocation (Liu *et al.,* 2010). Marker genotype data revealed a high number (40%) of male segregating allele which indicates that Albert is highly heterozygous genotype. A similar situation was reported by Chen *et al.* (2010) in the SC6xMianbao population in which 243 (46.6%) of AFLP, SRAP, SSR and EST-SSR markers segregated in the male compared to 116 (22.3%) in the female (Chen *et al.,* 2010). On the other hand presence of 23.4% of common markers helped in identification of homologous linkage groups and construction of the integrated map. Significant gap within linkage groups signifies insufficient polymorphic markers in some position within a chromosome hence unsaturated genetic linkage map. The situation also explains the slight increase in the number of linkage group to 19 as opposed to 18 linkage groups which correspond to cassava haploid chromosomes (n = 18) in the high stringency genetic map. The higher correlation in the order of markers between genetic maps presented in this study makes the framework genetic map reliable to further genetic studies. Cassava is heterozygous in nature. The crop has long growing cycle (up to 12 years), low seed set and characterized by inbreeding depression which make difficulty in developing second

population (F_2) for use in genetic mapping (Rojas *et al.,* 2009). The F1 mapping population in cassava has been observed to be convenient and has been reported in all previous map with exceptional to first SSR cassava genetic map where F2 was used (Okogbenin *et al.,* 2006). The use of similar parents in generating F1 population seems to be common for developing genetic map in cassava and other crops; example in cassava, TMS30572 x CM2177-2 (Fregene *et al.,* 1997; Okogbenin *et al.,* 2006), Huay Bong 60 x Hanatee (Kunkeaw *et al.,* 2011; Sraphet *et al.,* 2011). In the current map, one parent (Albert male) is similar to that used in previous map of (Rabbi *et al.,* 2012).

4.5 Conclusion

The high density cassava genetic map of cassava consisting of 4784 SNP markers spanning 4250.4cM with an average of marker density of 1.26cM/marker was constructed in this study. This map paves the way for future genetic and genomic studies including QTL mapping and marker assisted breeding approach for genetic improvement in cassava.

4.6 References

Awoleye, F., Duren, M., Dolezel, J. and Novak, F. J. (1994). Nuclear DNA content and in vitro induced somatic polyploidization cassava (*Manihotesculenta* Crantz) breeding. *Euphytica* 76: 195 – 202.

Chen, X., Xia, Z., Fu, Y., Lu, C. and Wang, W. (2010). Constructing a genetic linkage map using an F1 population of non-inbred parents in cassava (*Manihotesculenta* Crantz). *Plant Molecular Biology* 28: 676 – 683.

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 (5): 19 - 21.

FAO (2011). Food outlook: Global market analysis. FAO, Rome. [http://faostat.fao.org] site visited on 20/5/2012.

FAOSTAT (2010). Food and agriculture organizations statistics database. FAO, Rome. [http://faostat.fao.org] site visited on 20/5/2012.

Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J. and Bonierbale, M. (1997). A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theory Applied Genetics* 95: 431 – 441.

Kunkeaw, S., Yoocha, T., Sraphet, S., Boonchanawiwat, A., Boonseng, O., Lightfoot, D. A., Triwitayakorn, K. and Tangphatsornruang, S. (2011). Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihotesculenta* Crantz). *Molecular Breed* 27: 67 – 75.

Liu, X., Guo, L., You, J., Liu, X., He, Y., Yuan, H., Liu, G. and Feng, Z. (2010). Progress of segregation distortation in genetic mapping of plants. *Research Journal of Agronomy* 4(4): 78 – 83.

Lopez, C. E., Quesada-Ocampo, L. M., Bohorquez, A., Duque, M. C., Vargas, J., Tohme, J. and Verdier, V. (2007). Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihotesculenta*. *Genome* 50: 1078 – 1088.

Monger, W. A., Alicai, T. and Ndunguru, J. (2010). The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. *Archives of Virology* 155(3): 429 – 433.

Okogbenin, E. and Fregene, M. (2003). Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non in bred parents in cassava (*Manihotesculenta* Crantz). *Theory Applied Genetics* 107: 1452 – 1462.

Okogbenin, E., Marin, J. and Fregene, M. (2006). An SSR-based molecular genetic map of cassava. *Euphytica* 147: 433 – 440.

Poland, J. A., Brown, P. J., Sorrells, M. E. and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*. 7(2): 1 - 8.

Rabbi, I. Y., Kulembeka, H. P., Masumba, E., Marri, P. R. and Ferguson, M. (2012). An EST-derived SNP and SSR genetic map of cassava (*Manihotesculenta* Crantz), *Theory Applied Genetics* 125(2): 329 – 342.

Rojas M. C., Pérez J. C., Ceballos, H., Baena, D., Morante, N., Calle, F. (2009). Analysis of inbreeding depression in eight S_1 cassava families. *Crop Science* 49: 543 – 548.

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., Shirasawa, K., Isobe, S., Lightfoot, D. A., Tangphatsornruang, S. and Triwitayakorn, K. (2011). SSR and EST-SSR based genetic linkage map of cassava (*Manihotesculenta* Crantz). *Theory Applied Genetics* 122: 1161 – 1170.

Utsumi, Y., Sakurai, T., Umemura, Y., Ayling, S., Ishitani, M., Narangajavana, J., Sojikul, P., Triwitayakorn, K., Matsui, M., Minabe, R., Shinozaki, K., Seki, M. (2011). RIKEN Cassava Initiative: Establishment of a Cassava Functional Genomics Platform. *Tropical Plant Biology* 1: 110 - 116.

Van Ooijen, J. W. (2006). *Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Kyazma, B. V., Wageningen: 63pp.

Whankaew, S., Poopear, S., Kanjanawattanawong, S., Tangphatsornruang, S., Boonseng, O., Lightfoot, D. A. and Triwitayakorn, K. (2011). A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an out bred population. *BMC Genomics* 12: 261 – 266.

Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. and Butgereitt, A. (2010). Analysis of *Cassava brown streak virus*es reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology* 91(5): 365 – 376.

Wydra, K., Zinsou, V., Jorge, V. and Verdier, V. (2004). Identification of pathotypes of *Xanthomonas axonopodis* PV. Manihotis in Africa and detection of quantitative trait loci and markers for resistance to bacterial blight of cassava. *Phytopathology* 94: 1084 – 1093.

Yu, J. Z., Kohel, R. J., Fang, D. D., Cho, J., Van Deynze, A., Ulloa, M., Hoffman, S. M., Peper, A. E., Stelly, D. M., Jenkins, J. N., Saha, S., Kumpatla, S. P., Shah, M. R., Hugie, W. V. and Percy, R. G. (2012). A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. *G3 (Bethesda)* 2: 42 – 58.

CHAPTER FIVE

GENERAL CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Authentification of F1 hybrids resulted from controlled crossing is very important in providing accurate and efficiency way in selection of true to type hybrids. Breeding crossing errors, seed contamination and admixture population is one of the challenges to breeding in many crops. Cassava is one of the crops which face these breeding errors when controlled crossing is used for development of F1 population. This study has applied molecular markers SSR for verification of true F1 hybrids of population cross. Early detection of false hybrids can prevent further propagation in subsequent generations thereby enhancing the effectiveness and efficiency of breeding program. The study have allowed for selection of the true hybrid population that was eventually use to provide good stock for mapping studies and linkage mapping of CBSD.

In this study, high-throughput SNP markers were used to develop high density genetic linkage map for CBSD resistance using AR40-6 F1's population. The high density cassava genetic map of cassava consisting of 4784 SNP markers spanning 4250.4cM with an average of marker density of 1.26cM/marker was constructed. Single Nucleotide Polymorphism genotyping by using advanced GBS approach offers high cost effective and efficiency method for high-throughput genotyping.

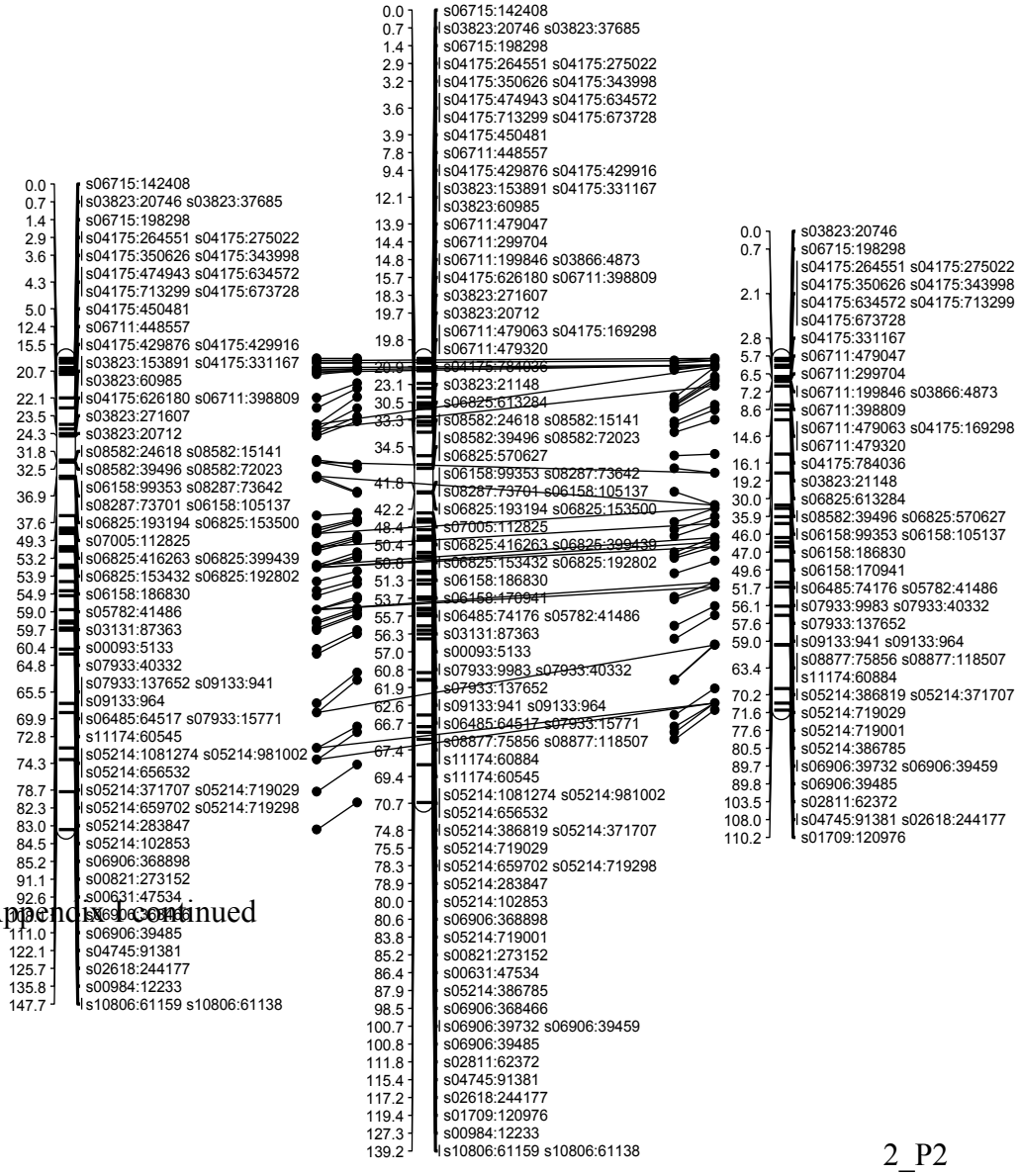
5.2 Recommendations

Early identification of true F1 hybrids and validation of parents prior to crossing is important for rapid improvement of the crop as it reduce the cost of field management, time consuming in breeding and errors in genetic studies. The principle used in this study for verification of F1 population using SSR markers can be extended to other markers with corresponding routine of polymorphism in any population of the crop. More studies can be done in developing efficient markers such as SNP and other gene-based markers and use them for verification of populations integrity in breeding scheme.

The genetic linkage map developed in this study, paves the way for future genetic and genomic studies including QTL mapping and marker assisted breeding approach for genetic improvement in cassava. The GBS approach used in this study is useful for marker assisted genomics and assessment of genetic resources in wide range of species. It allows large numbers of markers to be identified an mapped in very short time an at a lower cost therefore call for genotyping of many crops.

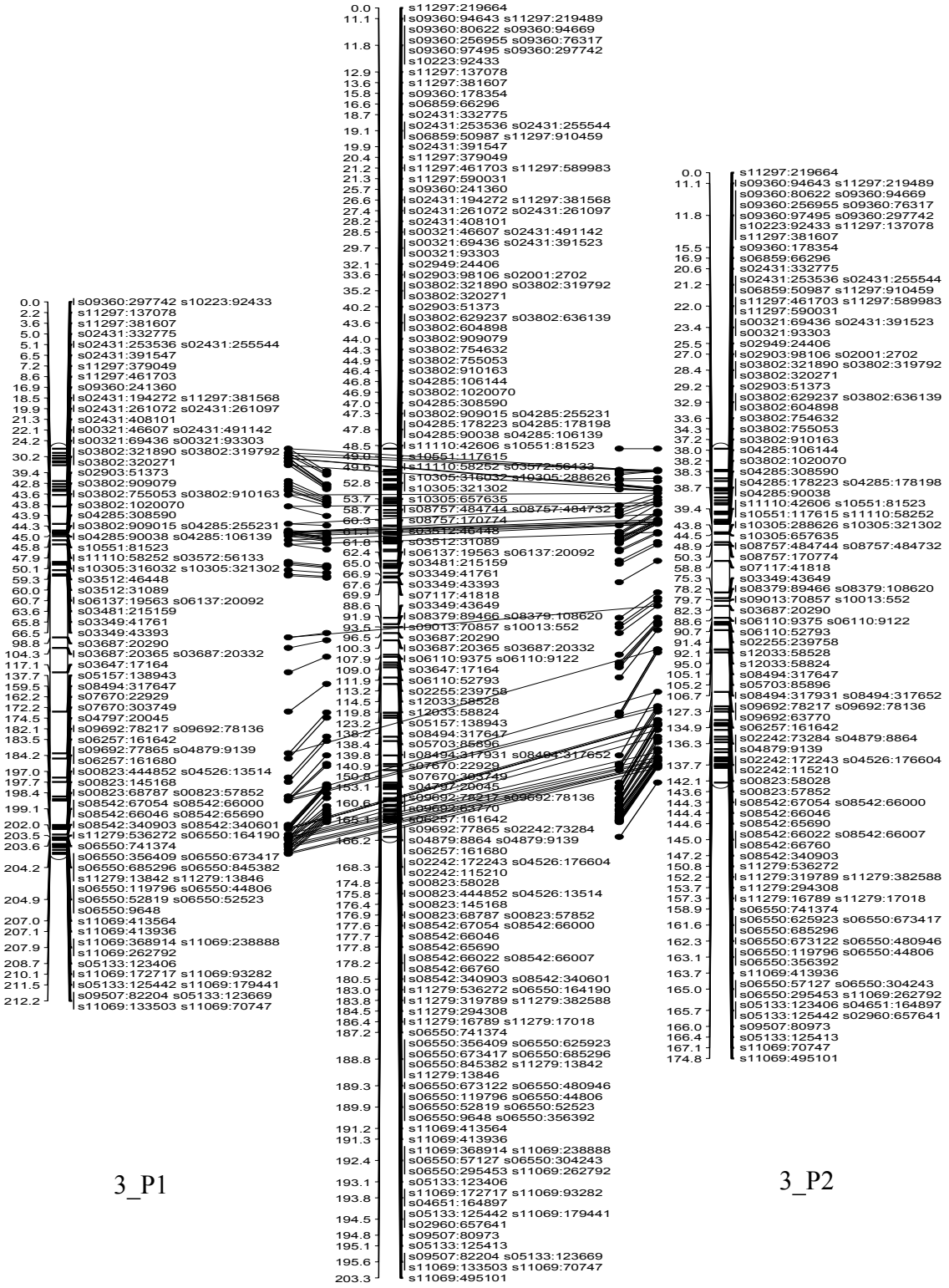
APPENDICES

Appendix 1: An alignment of Comprehensive genetic map (CM) map derived from AR40-6 x Albert F_1 population. The linkage group map on the left (designated by P1) are female maps, on the right (designated by P2) are male maps and at the centre, are the integrated map combining markers in female and male maps. Arrows represents homologous markers between female map, integrated and male maps. Scaffold name (designated by s followed by 5-digit identity) is followed by SNP name of the right of bar and distance between markers on the left bar of each map

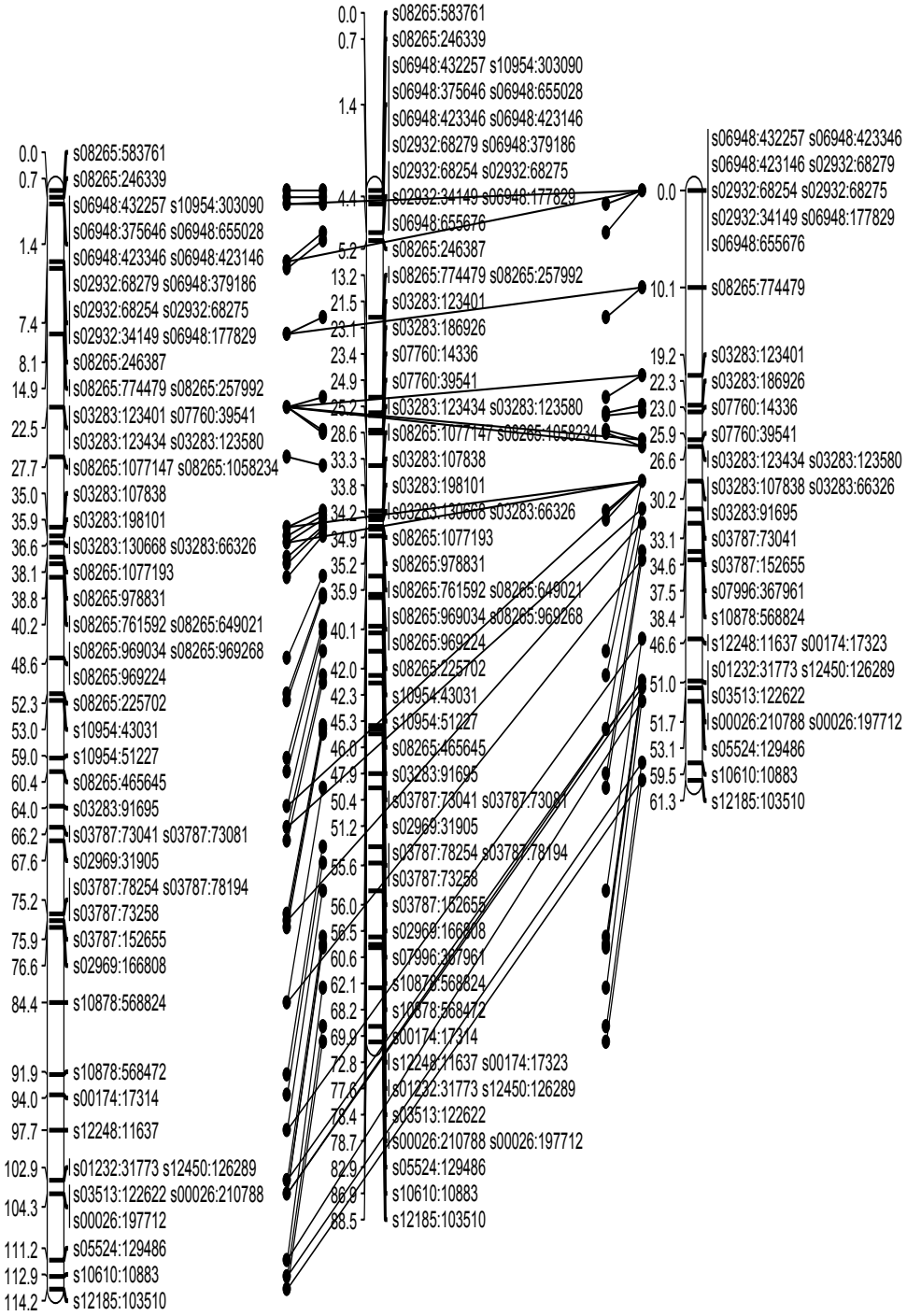


Appendix 1 continued

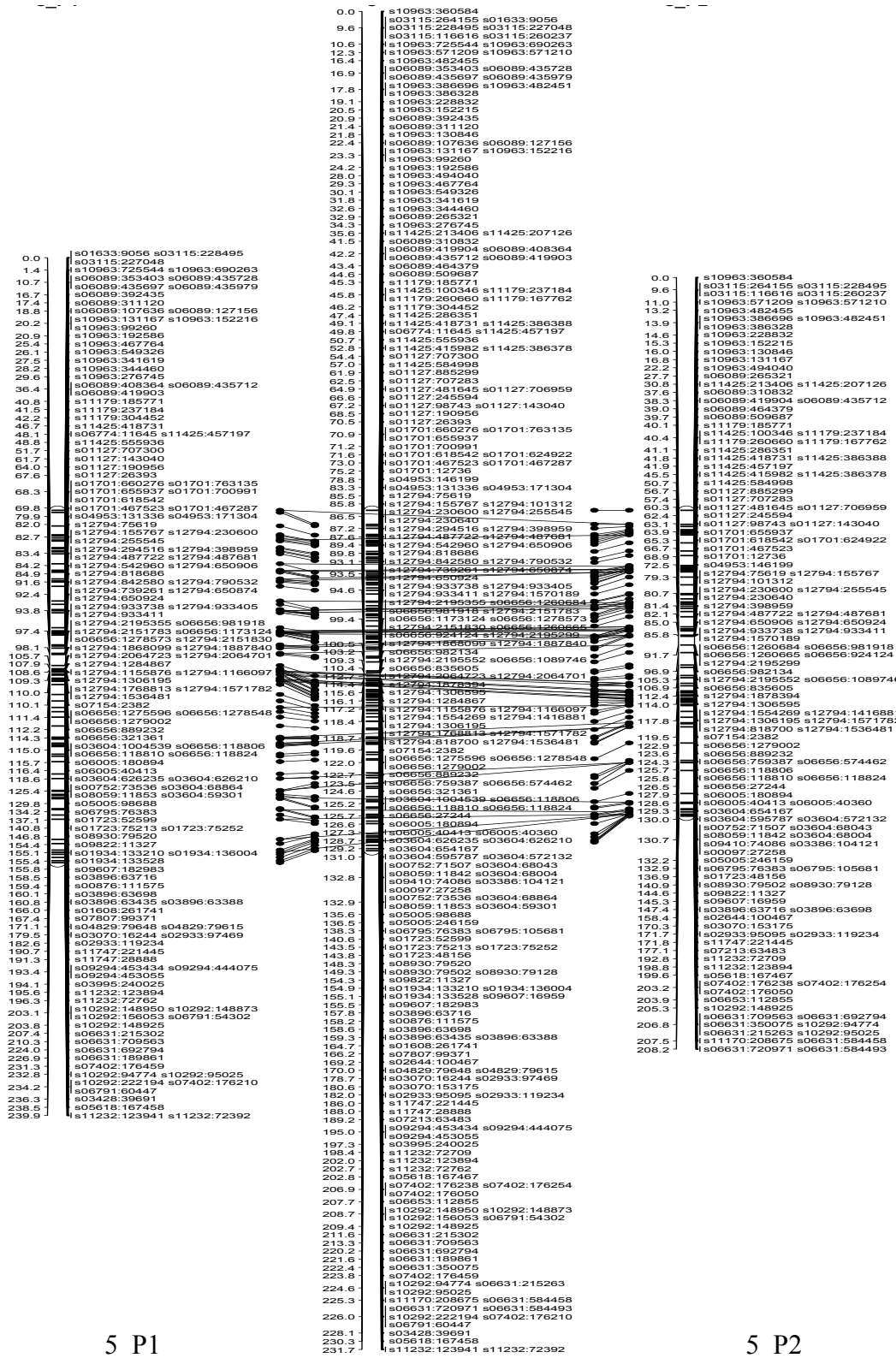
Appendix I continued



Appendix I continued

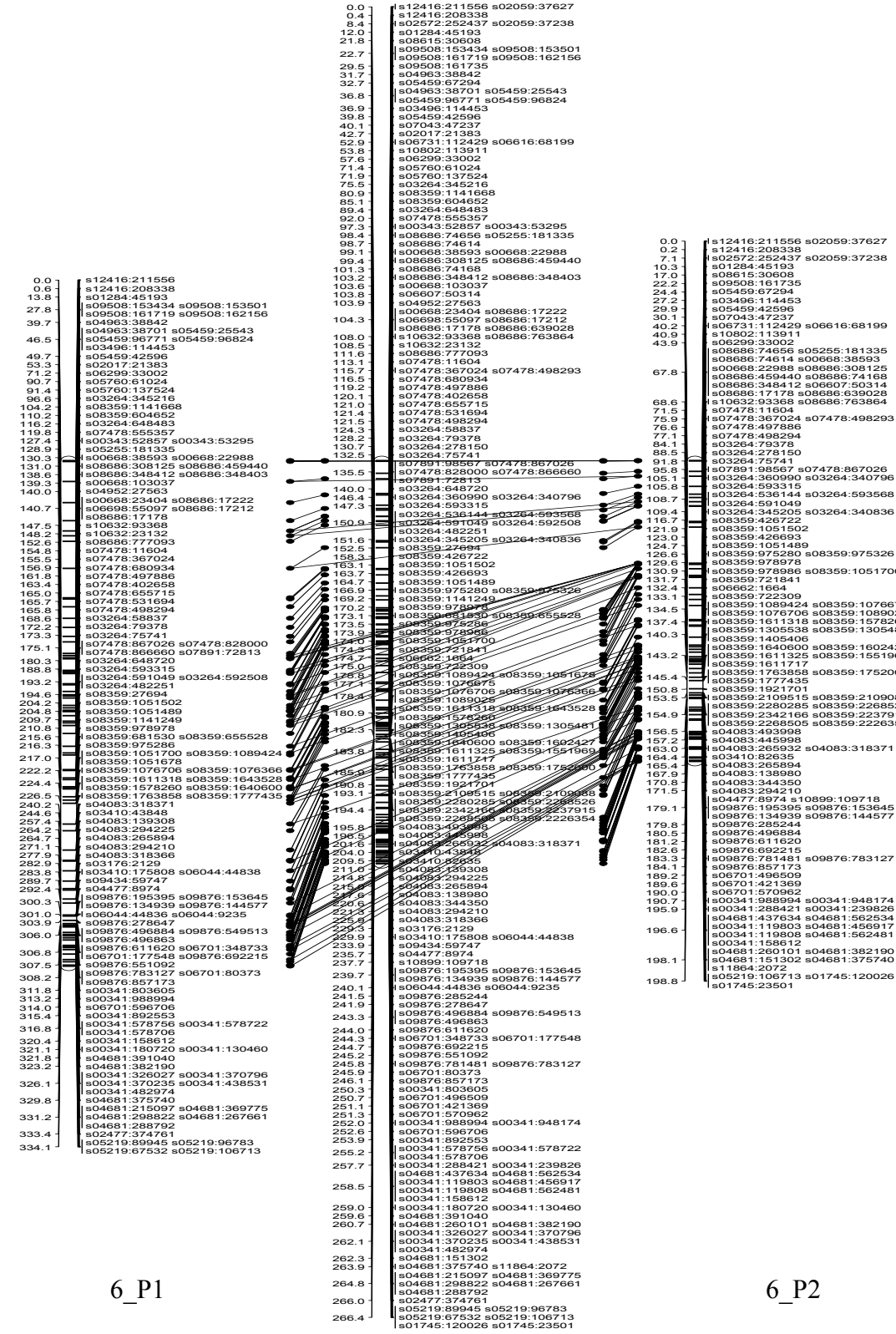


Appendix I continued

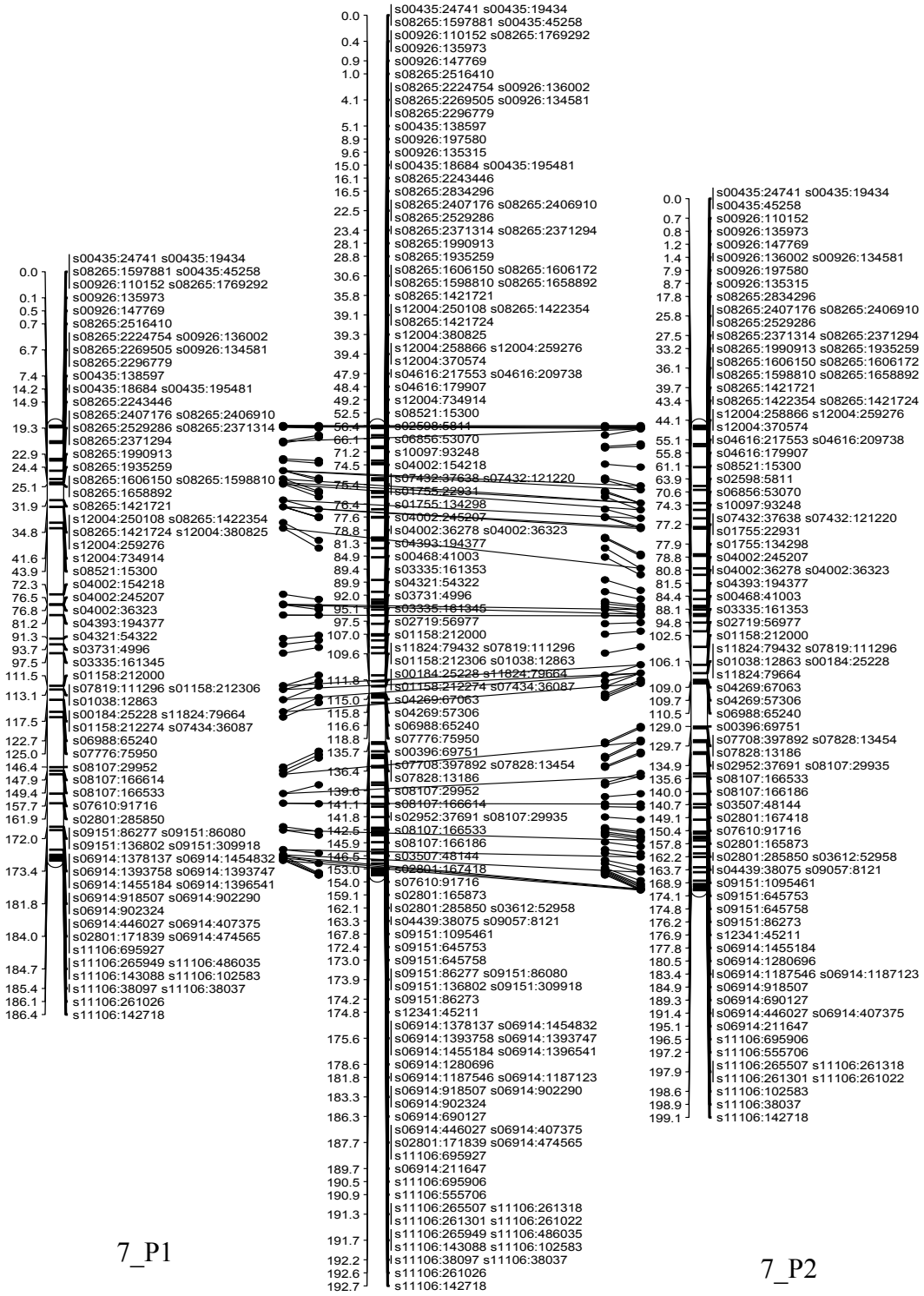


LG 5

Appendix I continued

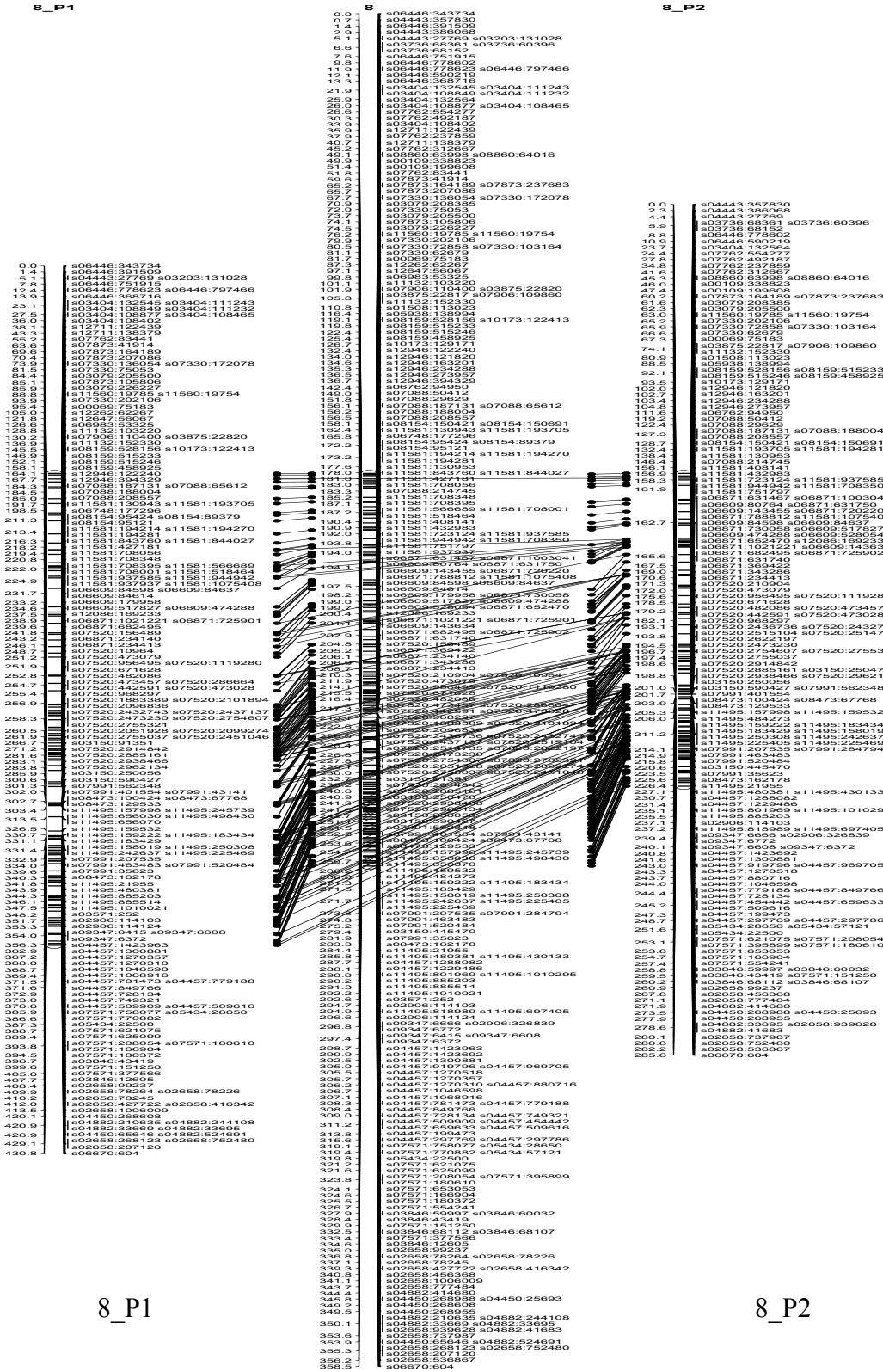


Appendix I continued

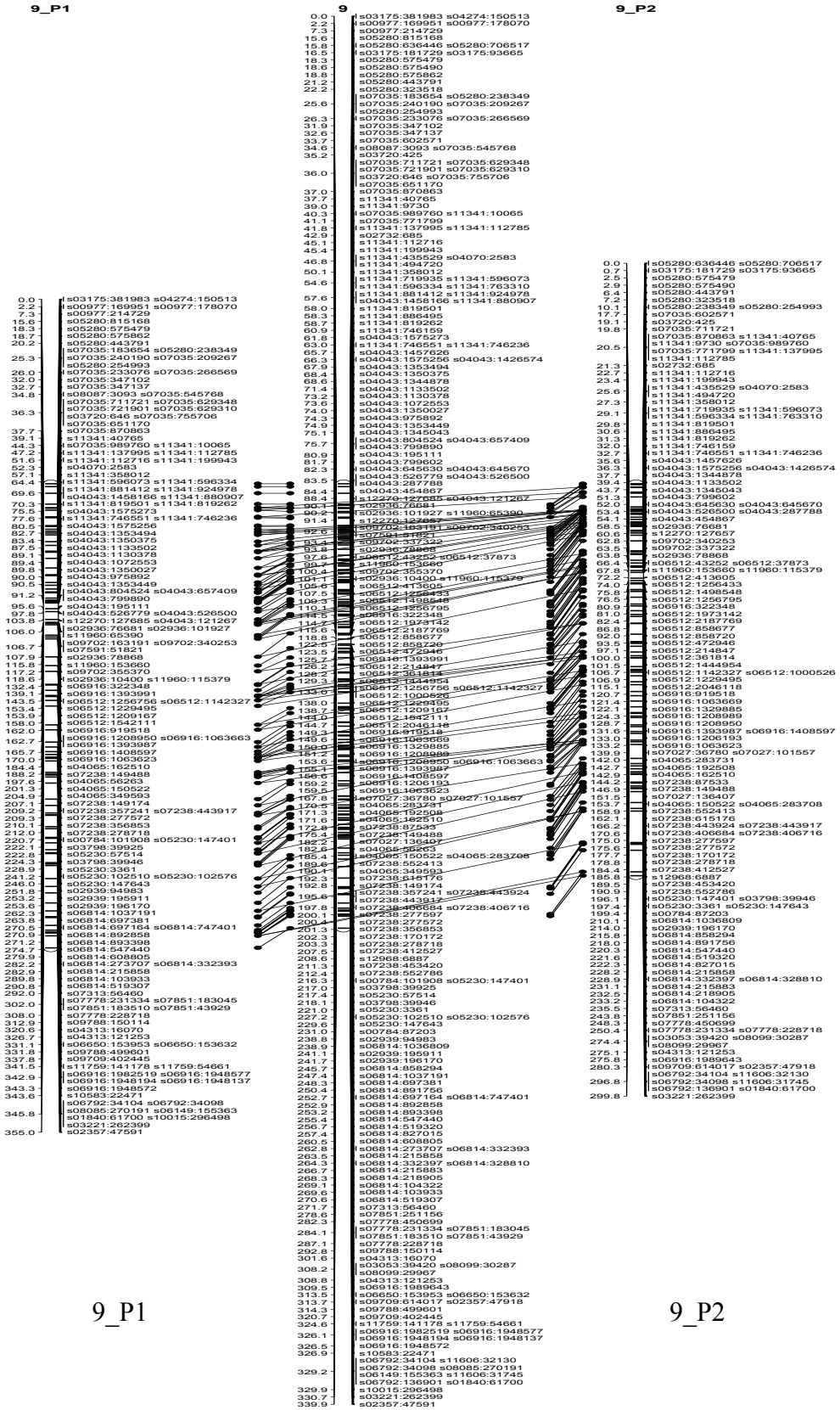


LG 7

Appendix I continued



Appendix I continued

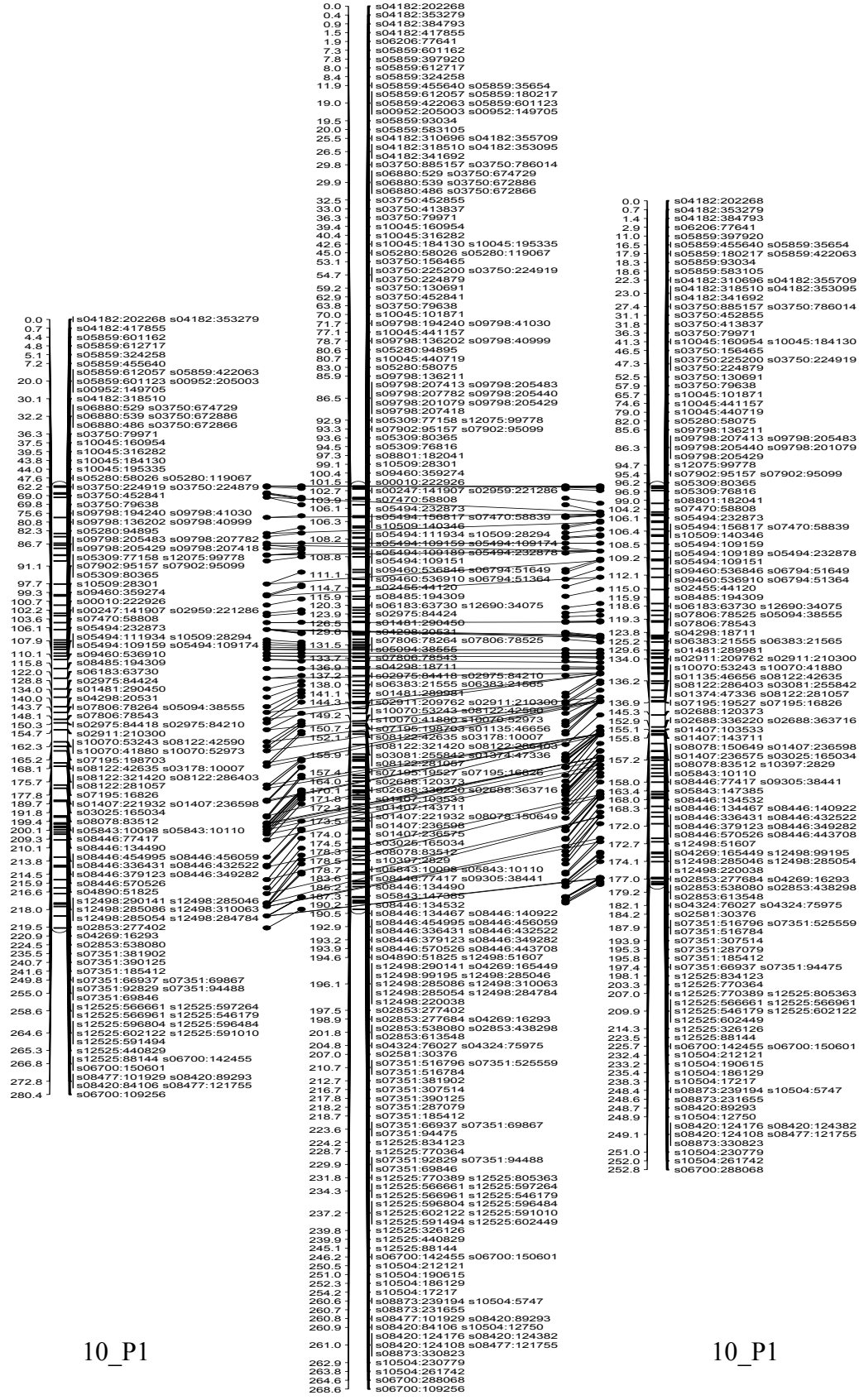


9_P1

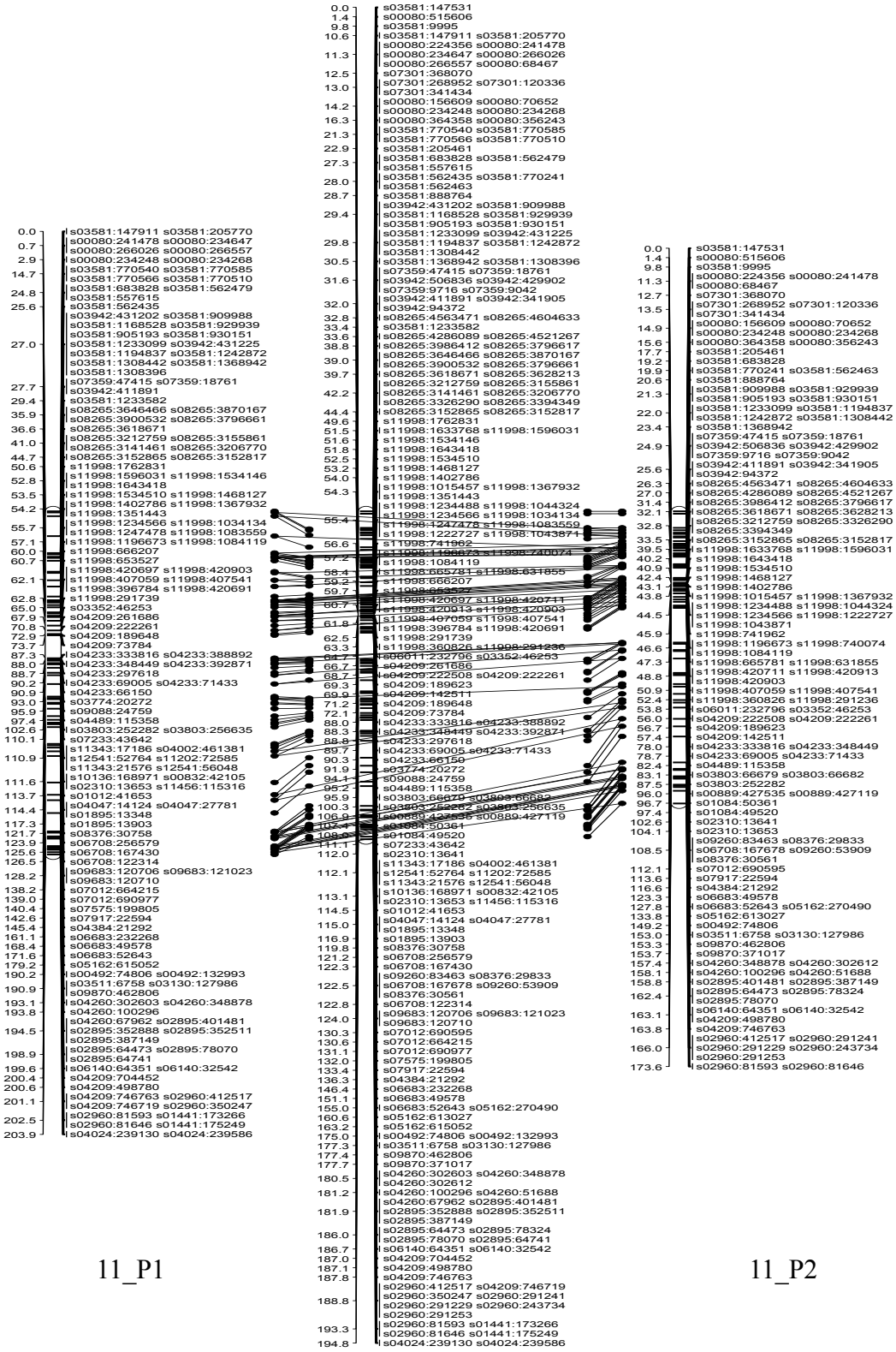
LG 9

9_P2

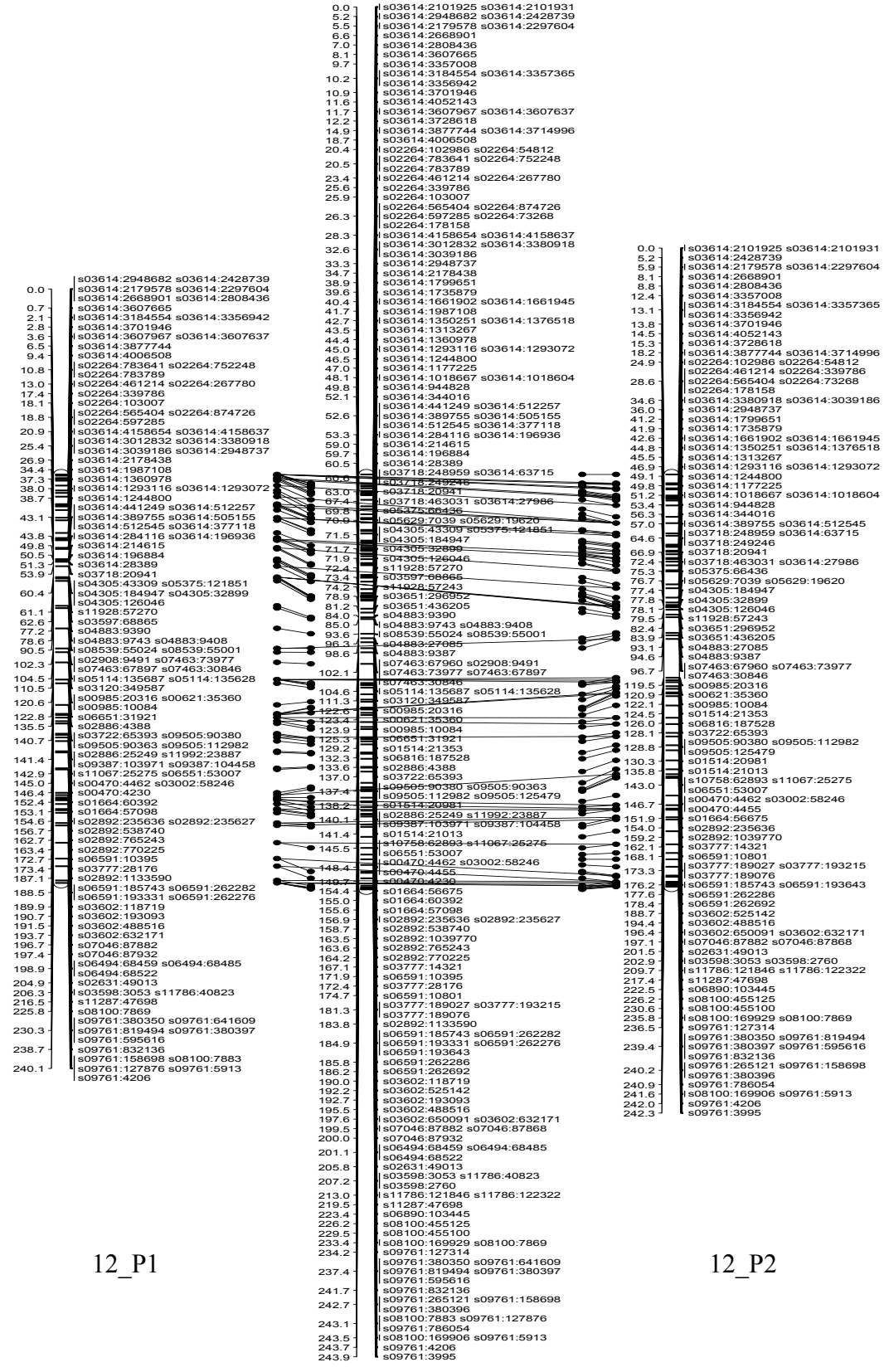
Appendix I continued



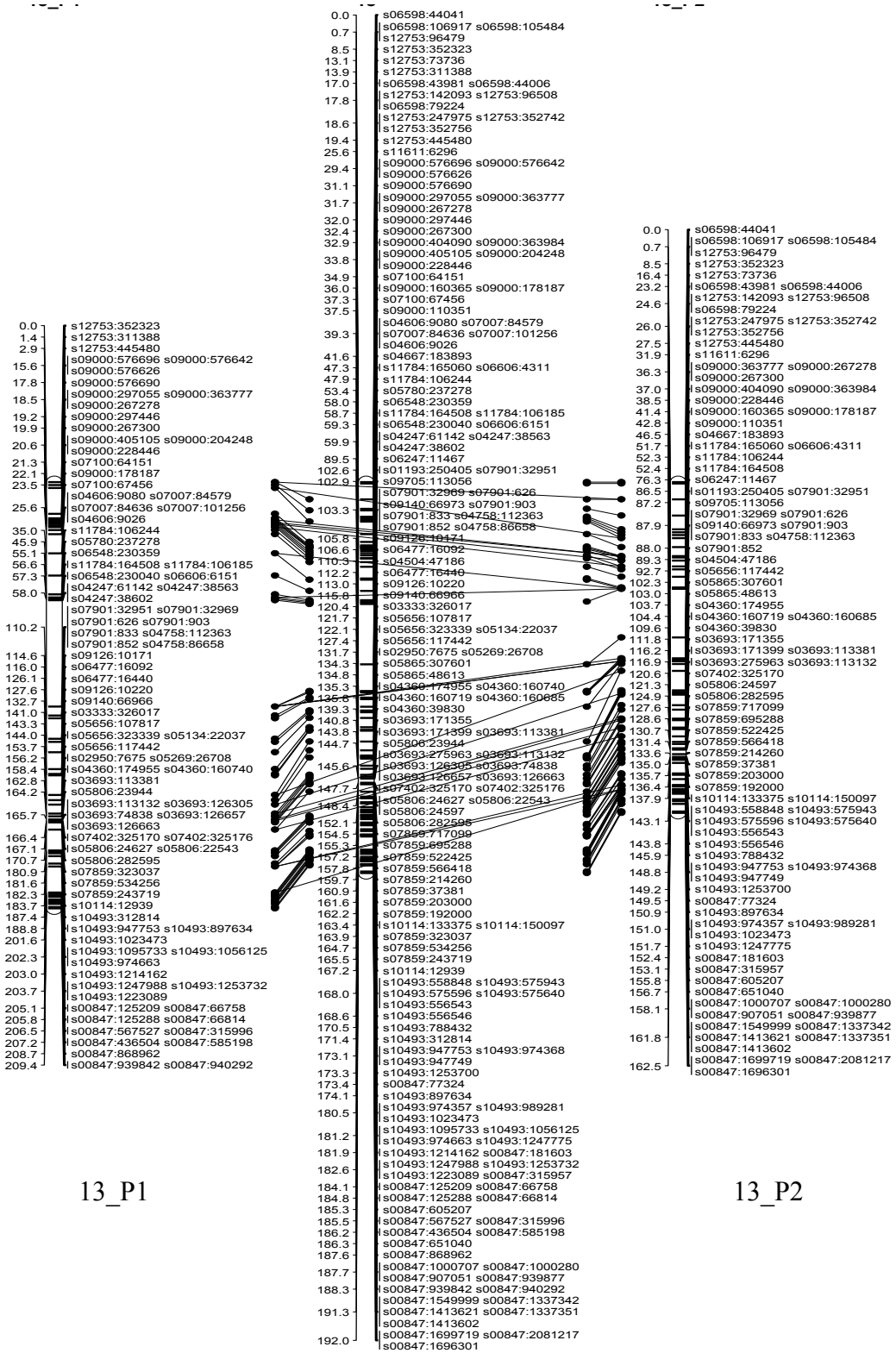
Appendix I continued



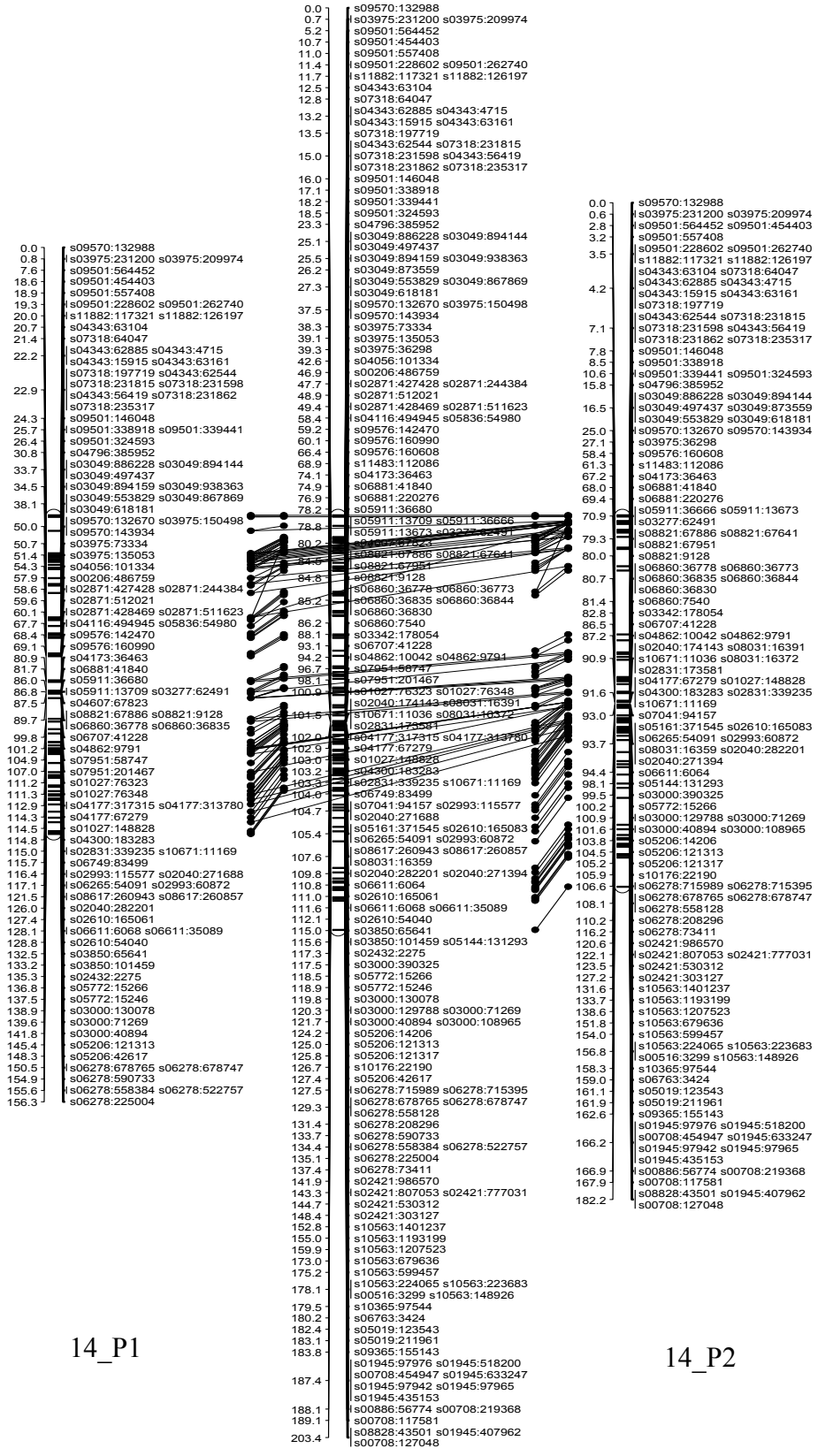
Appendix I continued



Appendix I continued

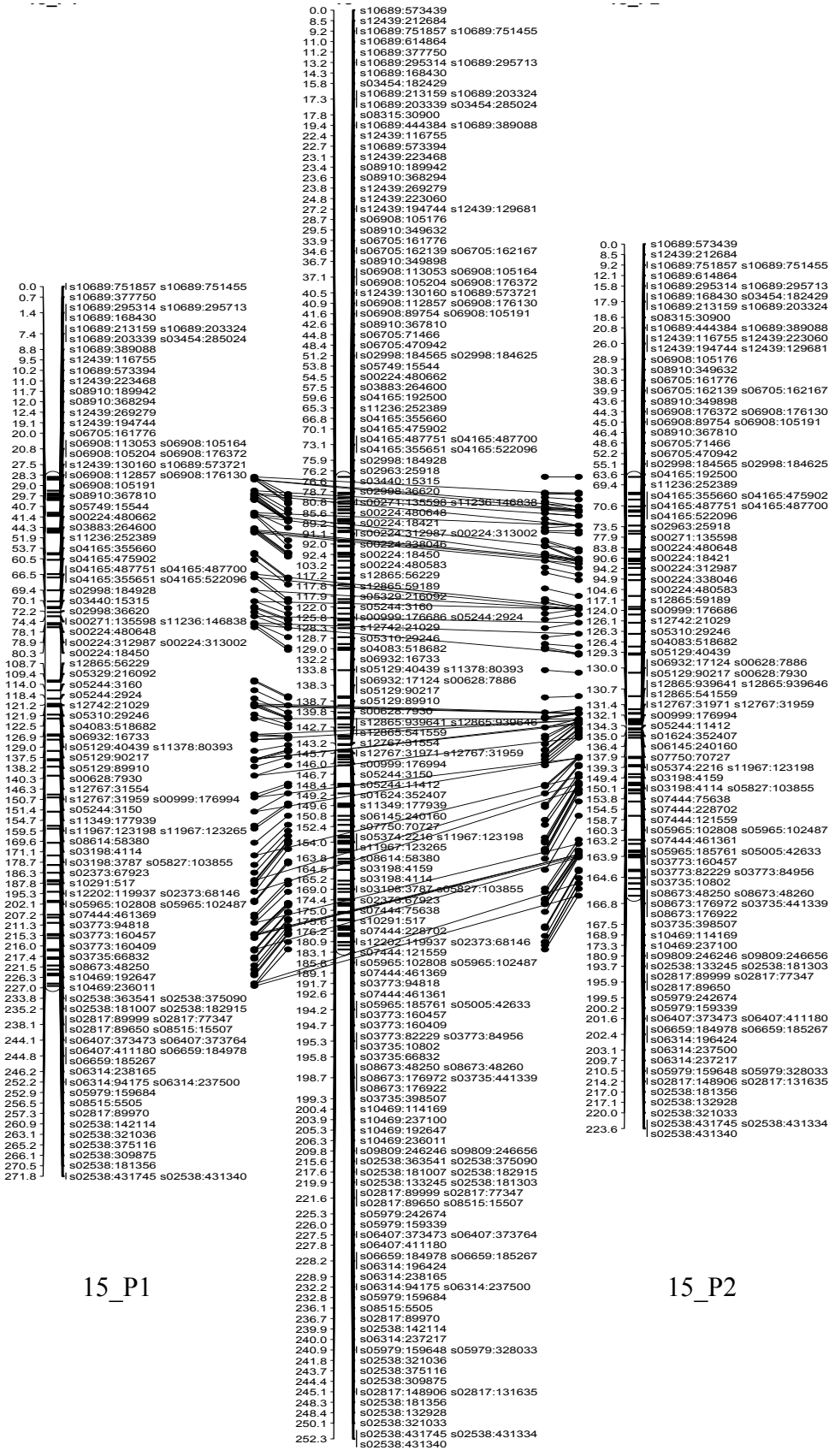


Appendix I continued

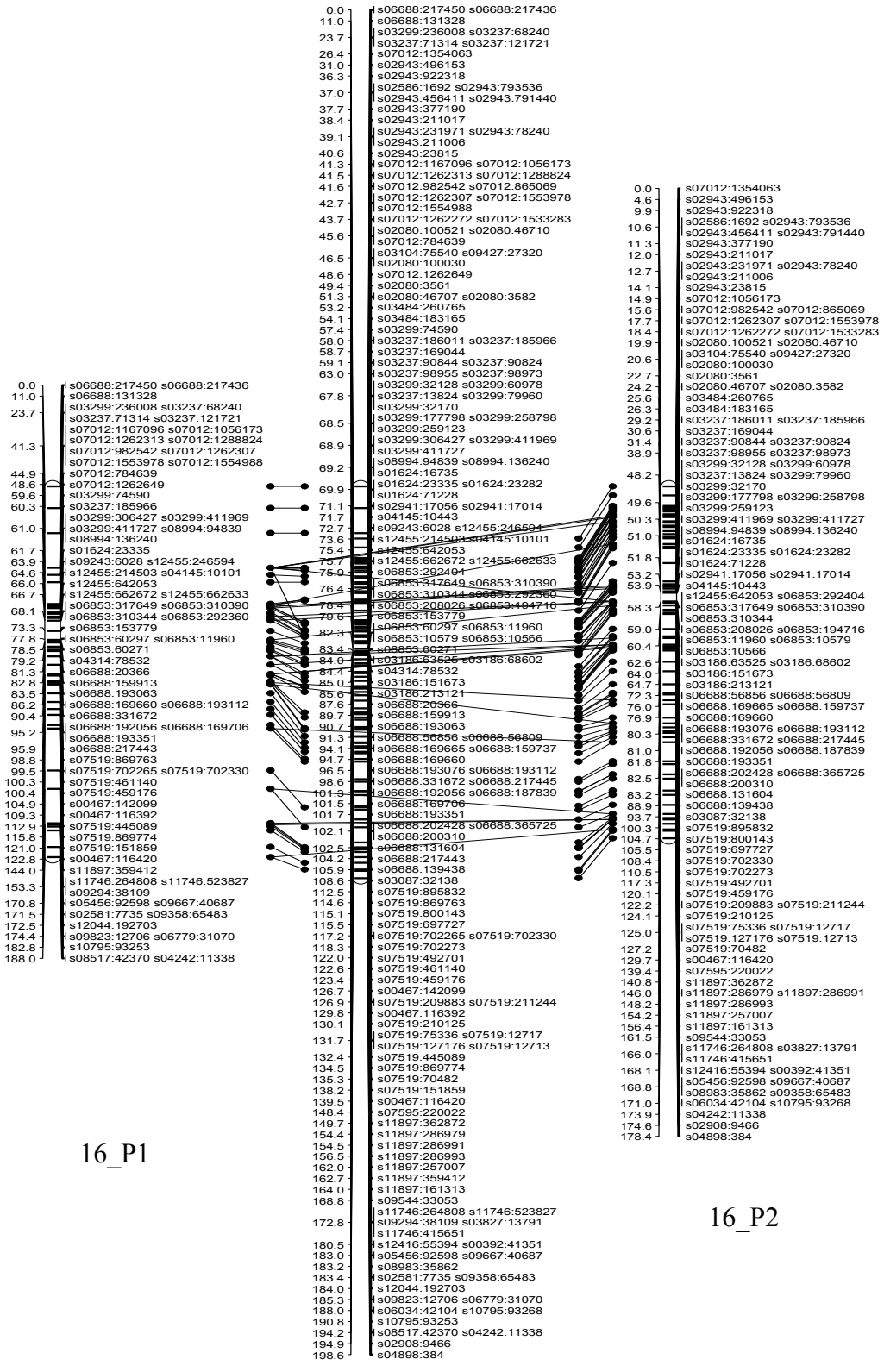


LG 14

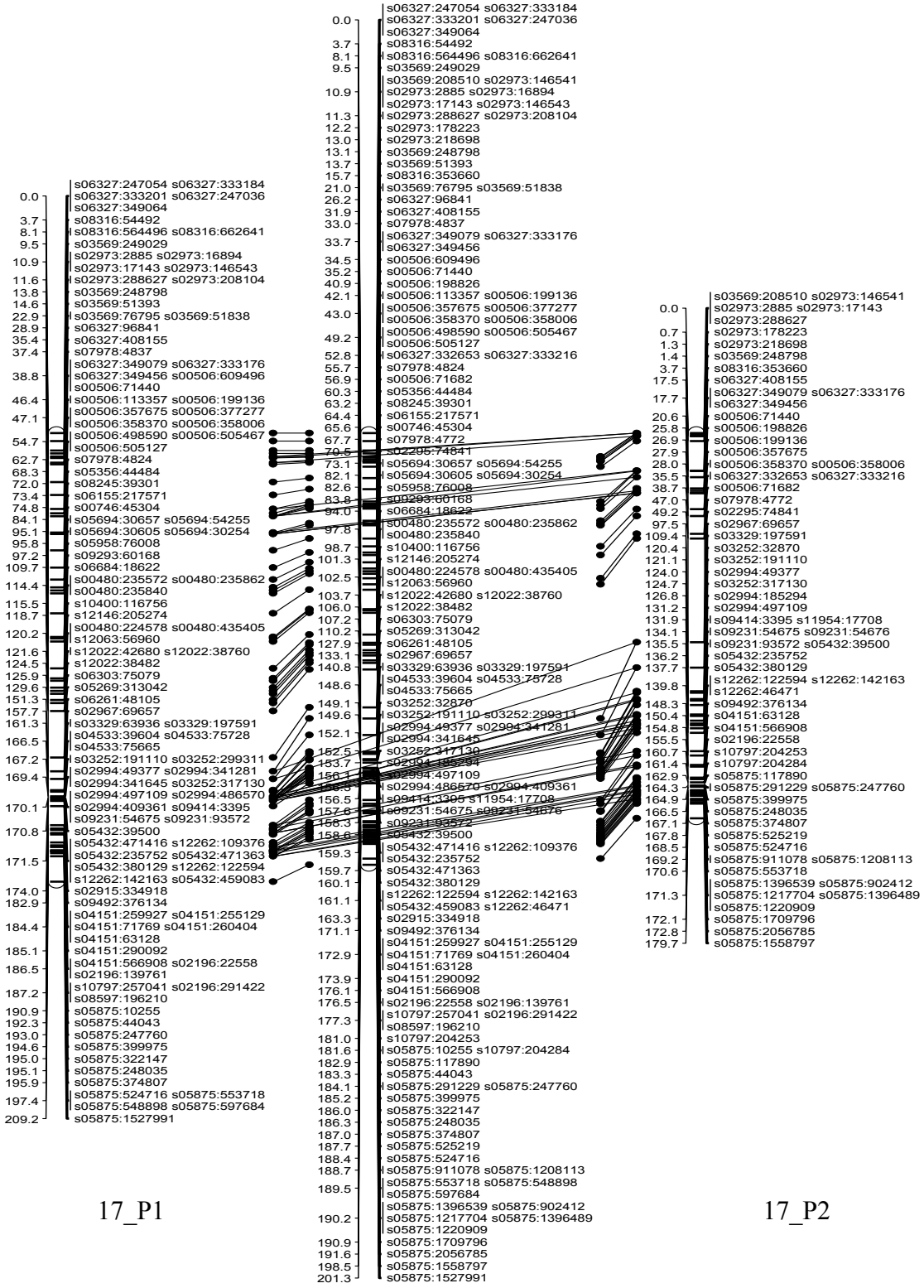
Appendix I continued



Appendix I continued

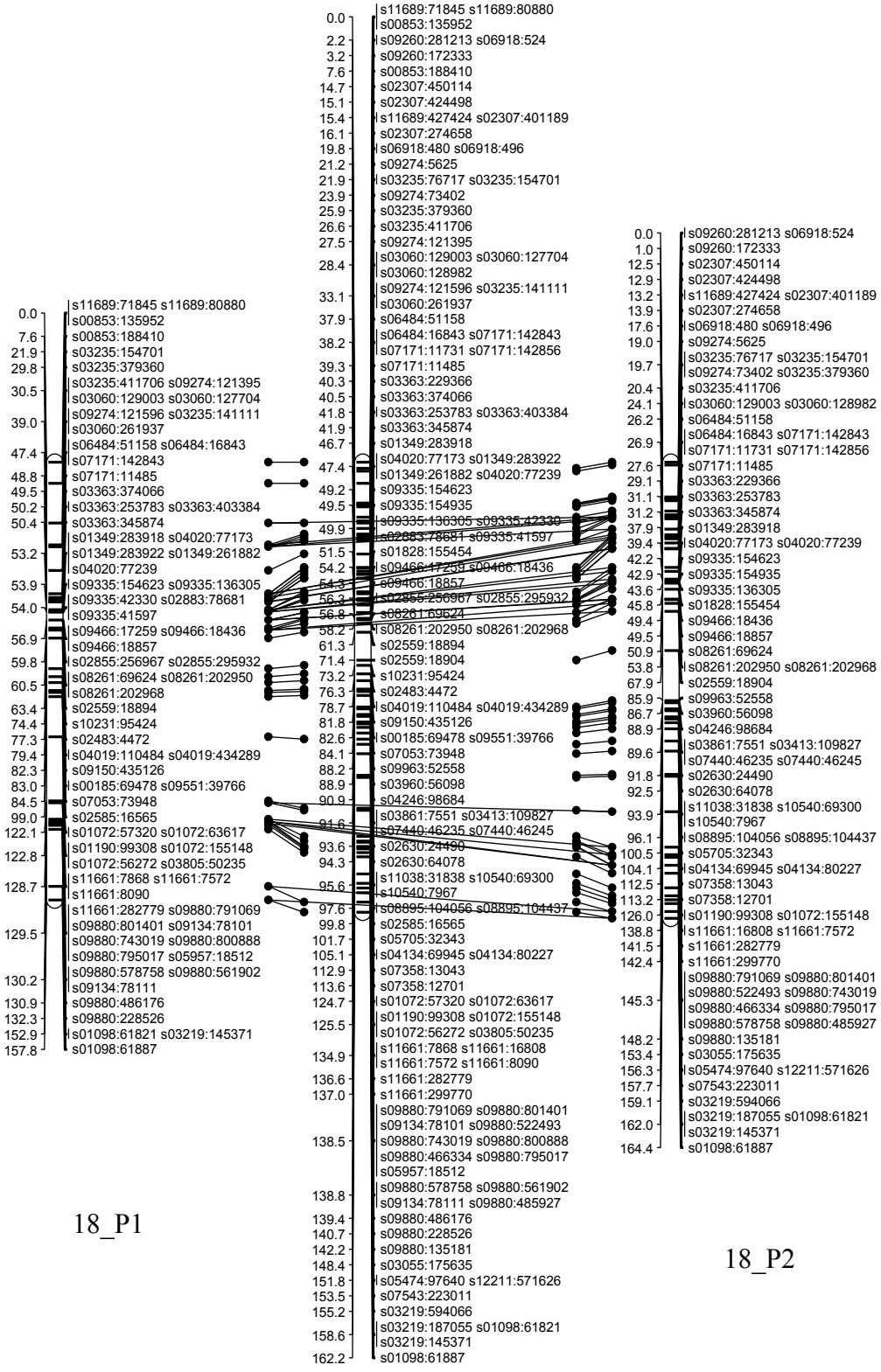


Appendix I continued



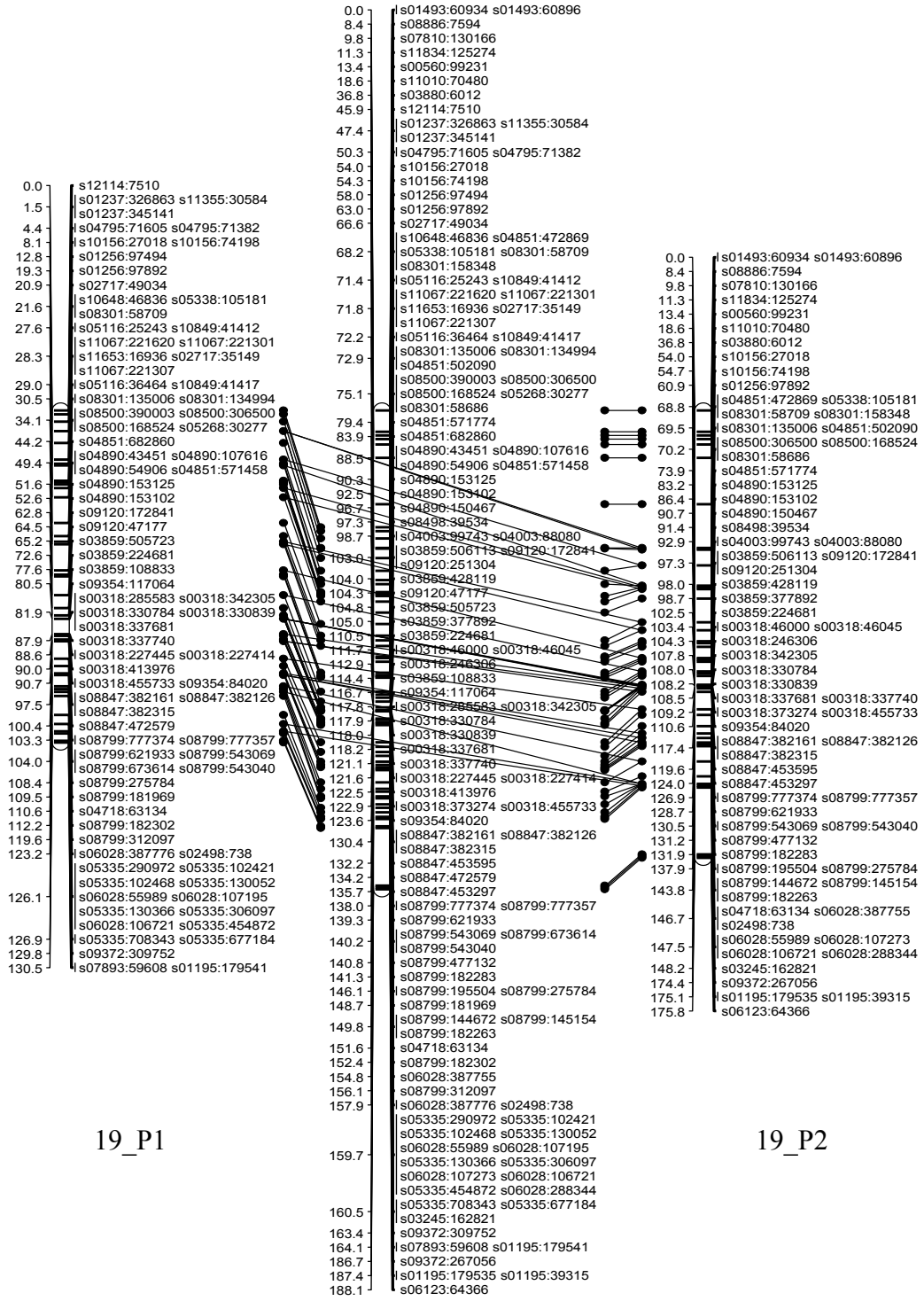
LG 17

Appendix I continued

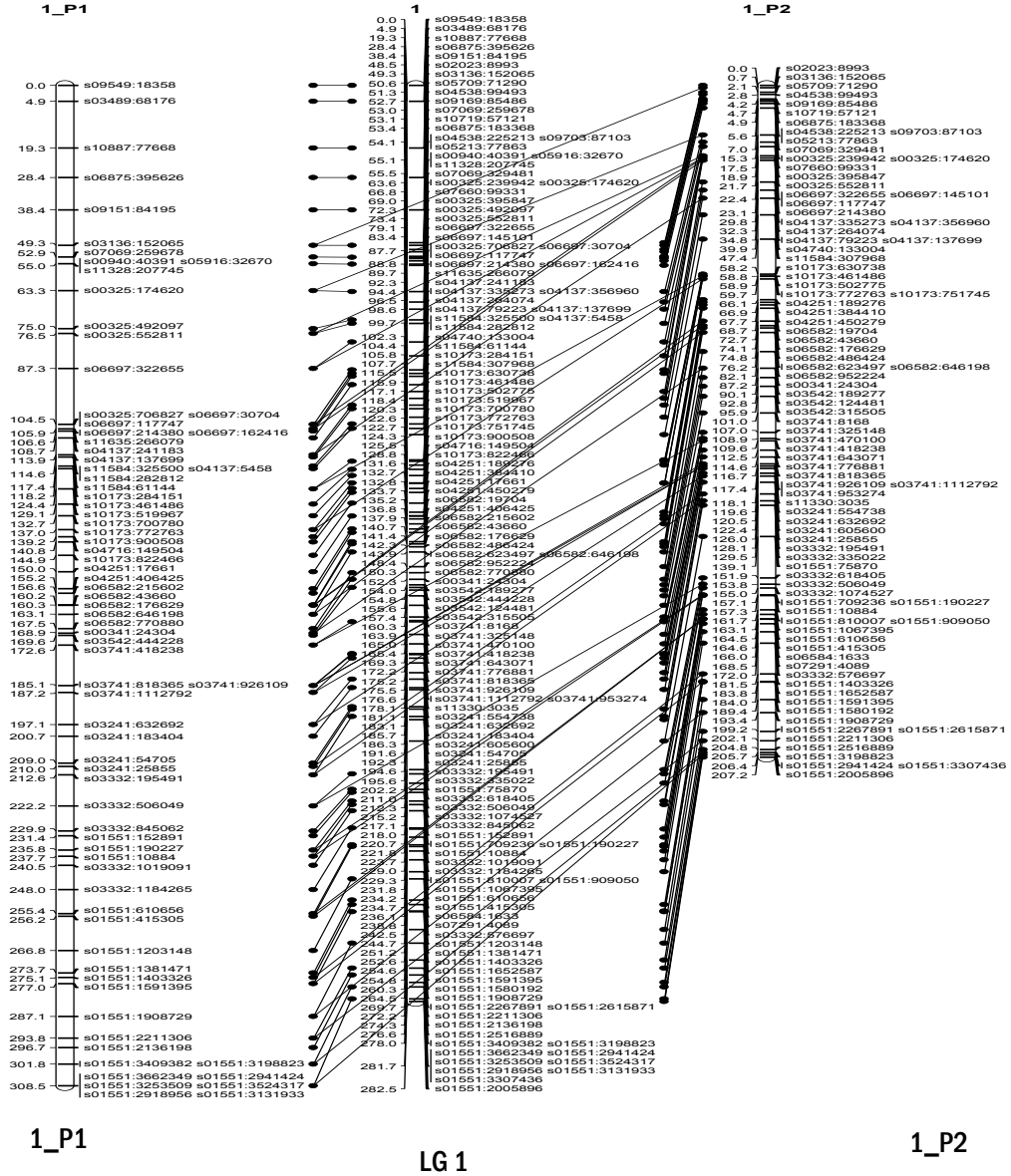


LG 18

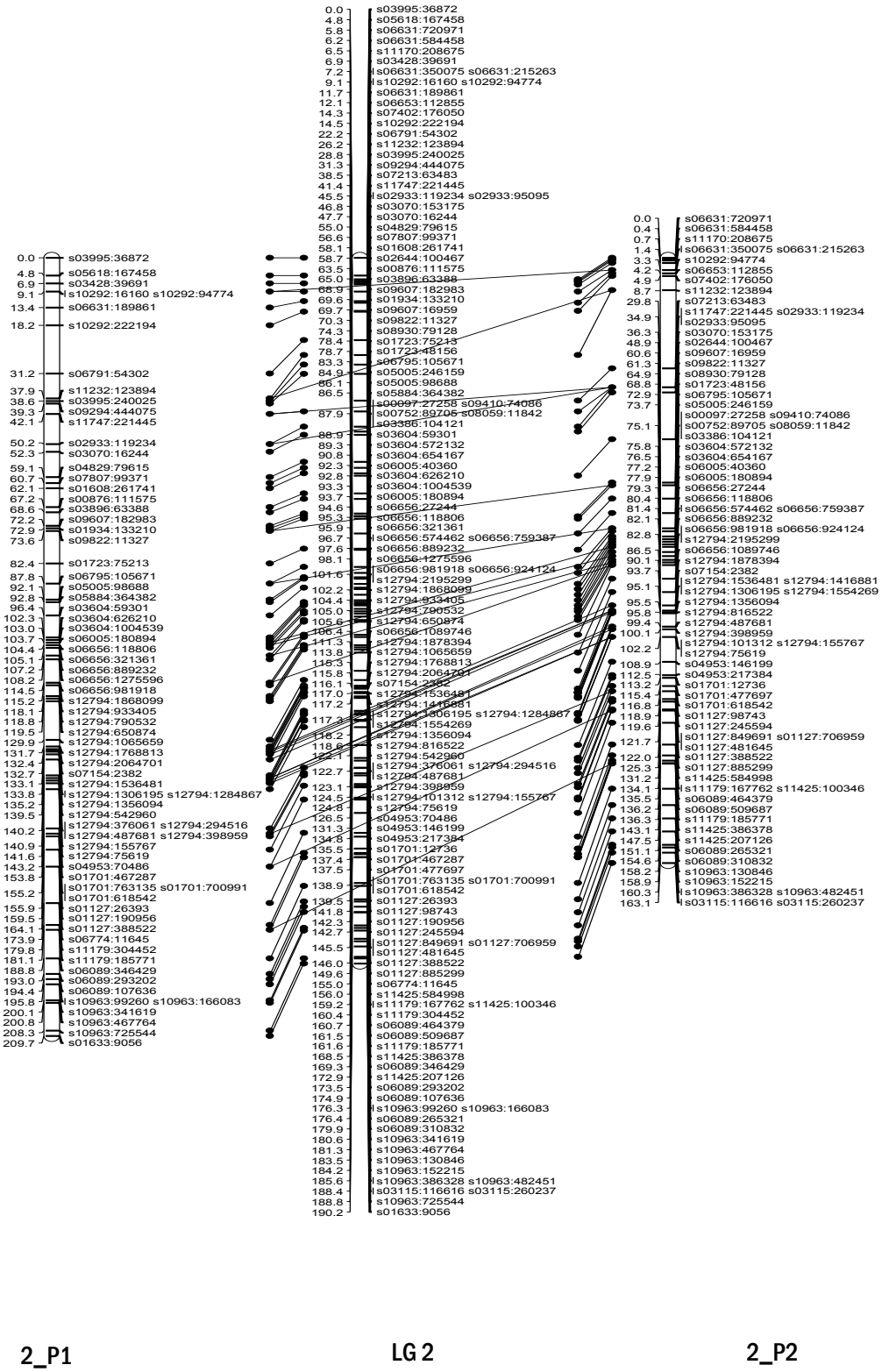
Appendix I continued



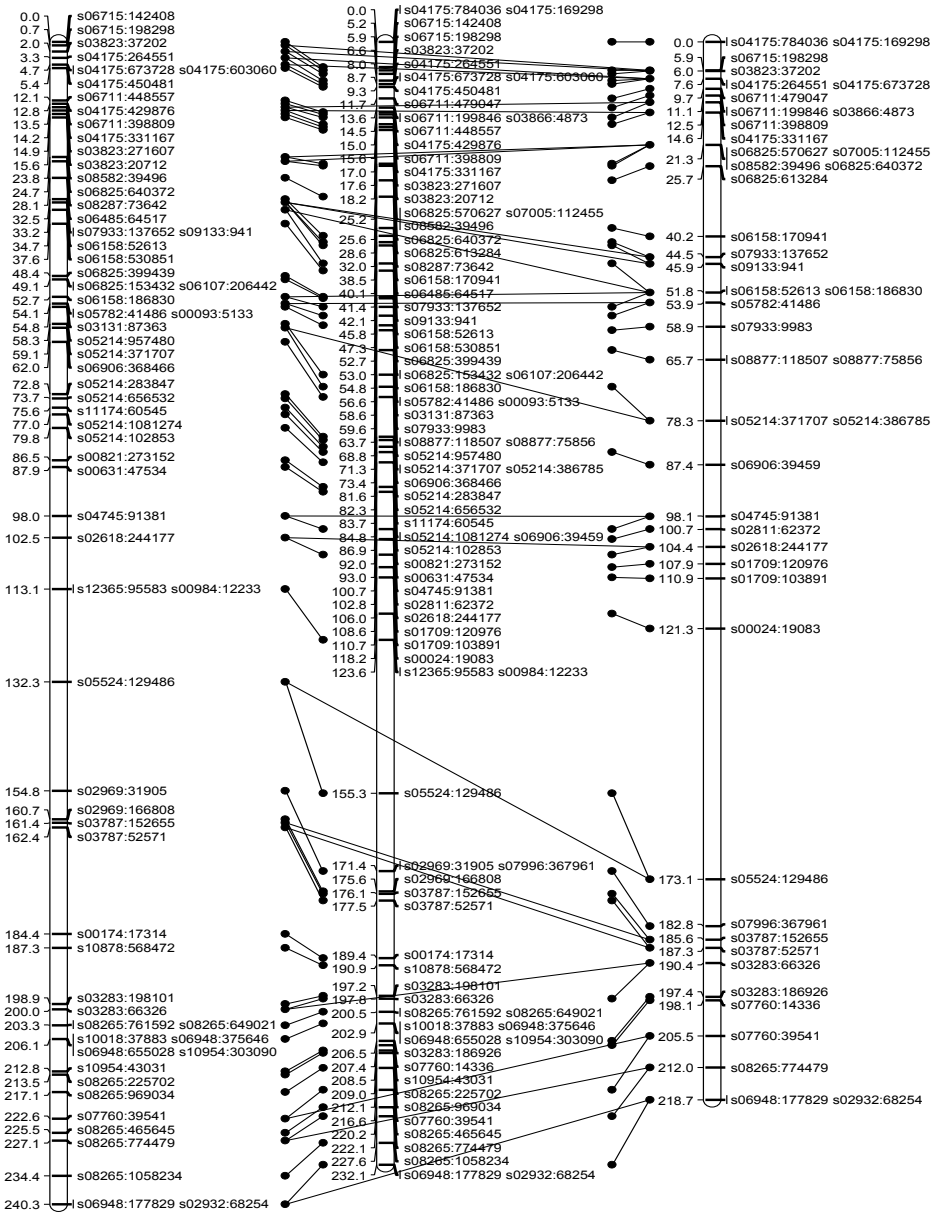
Appendix I1: An alignment of cassava Framework genetic map derived from AR40-6 x Albert F_1 population constructed using 2041 SNP markers. The linkage groups mapped on the left (designated by P1) are female maps, on the right (designated by P2) are male maps and at the Centre are the integrated map combining markers in female and male maps. Arrows represents homologous markers between female map, integrated and male maps. Scaffold name (designated by s followed by 5-digit identity) is followed by SNP name of the right of bar and distance between markers on the left bar of each map



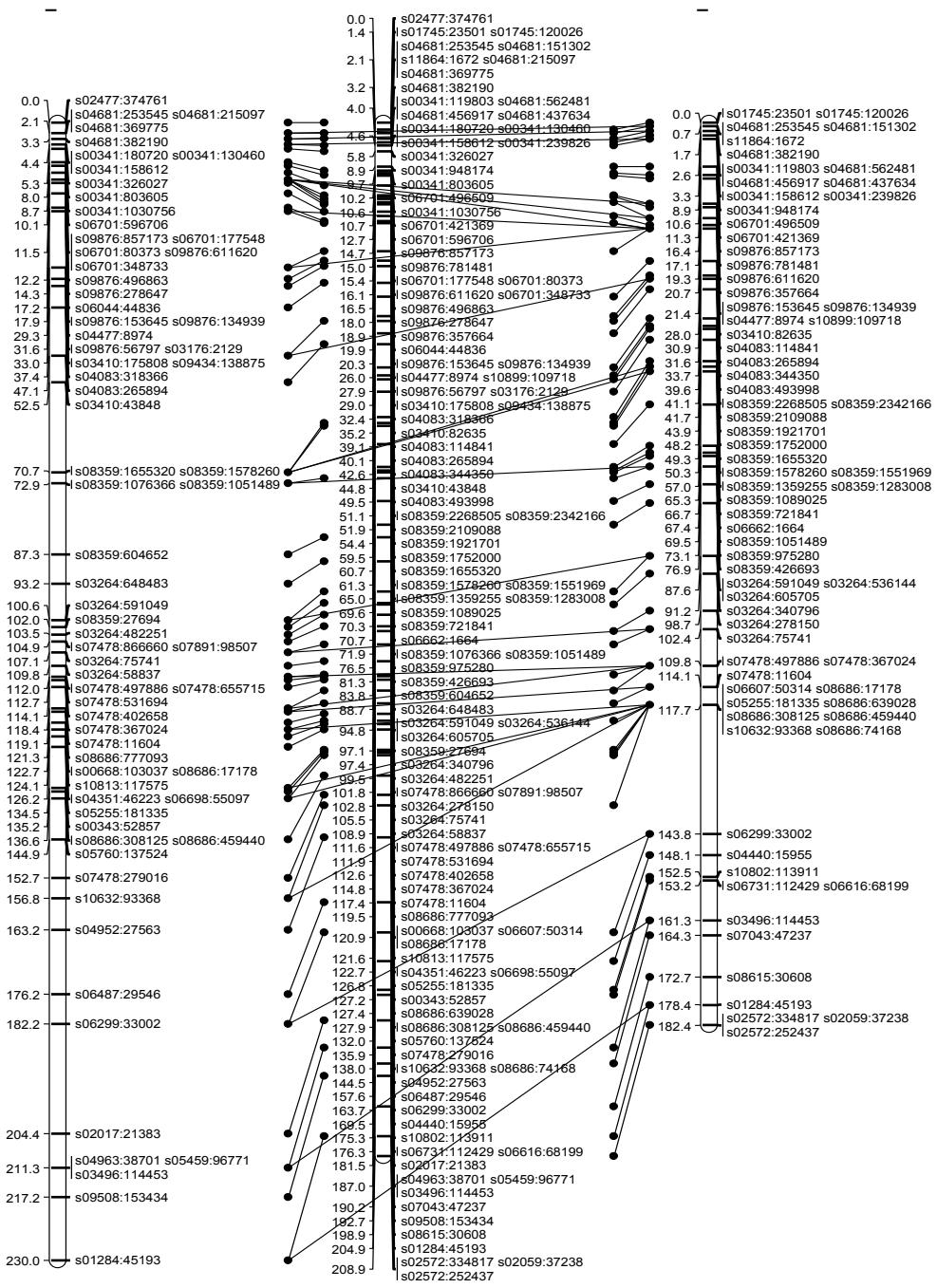
Appendix II continued



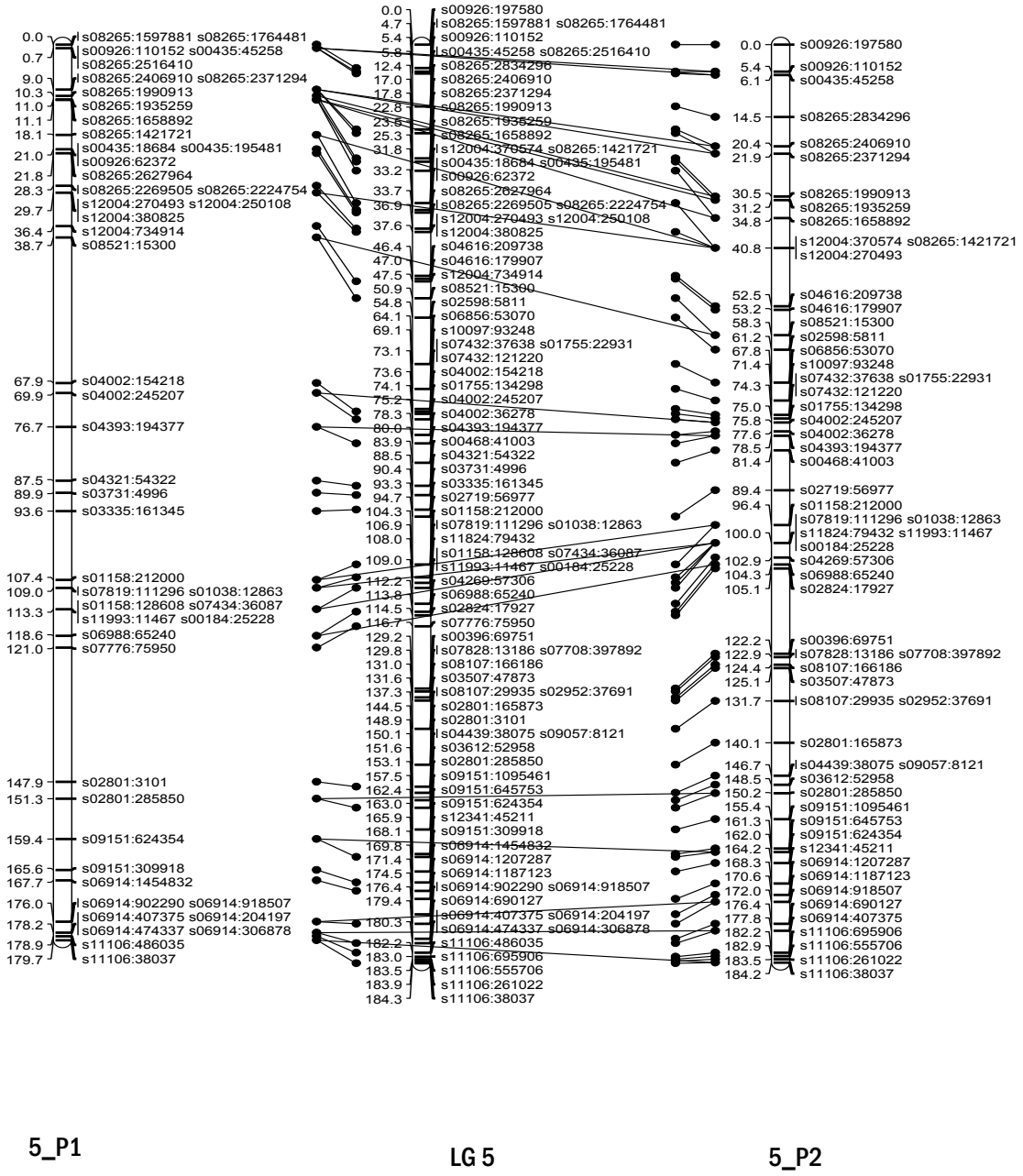
Appendix II continued



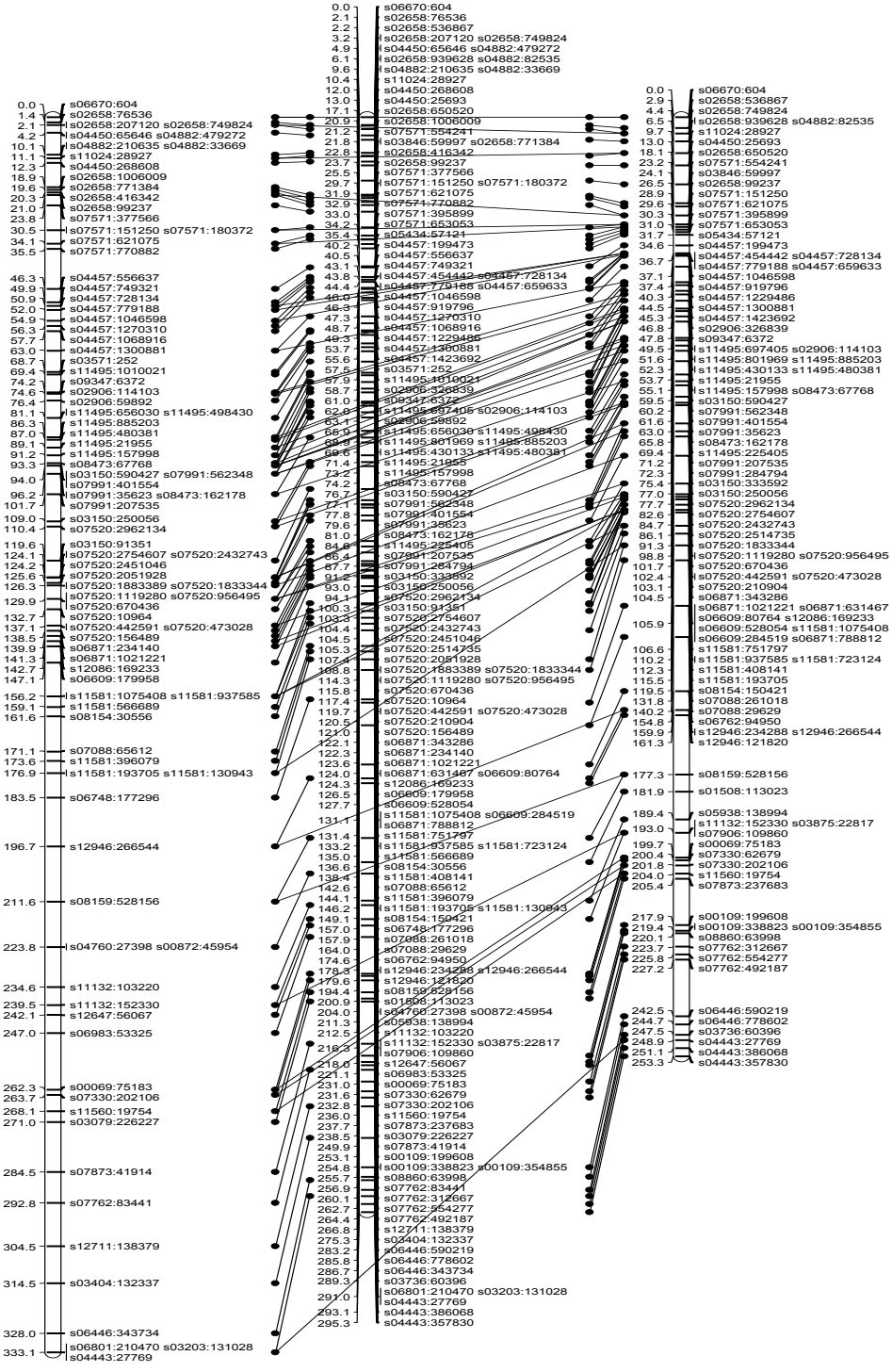
Appendix II continued



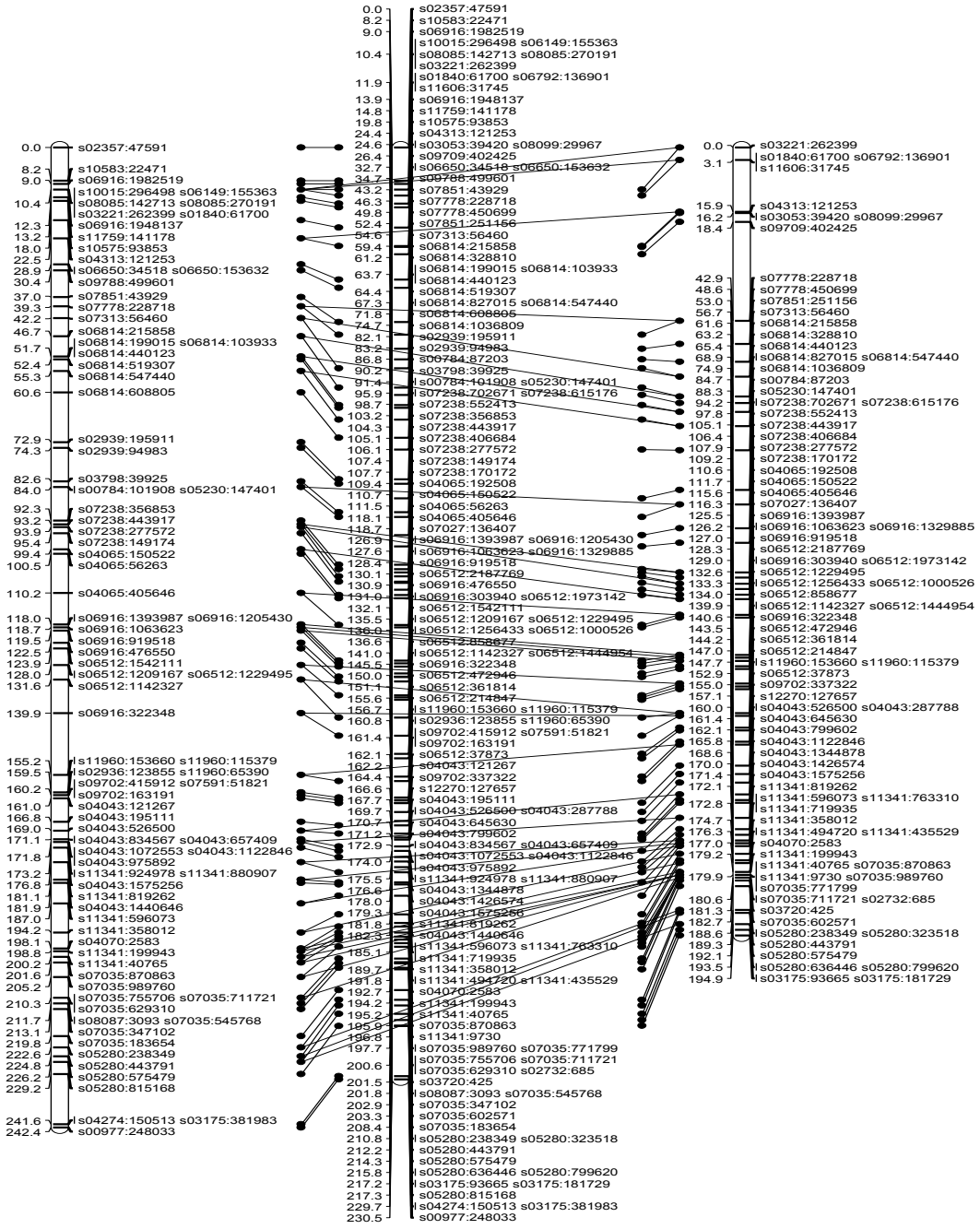
Appendix II continued



Appendix II continued

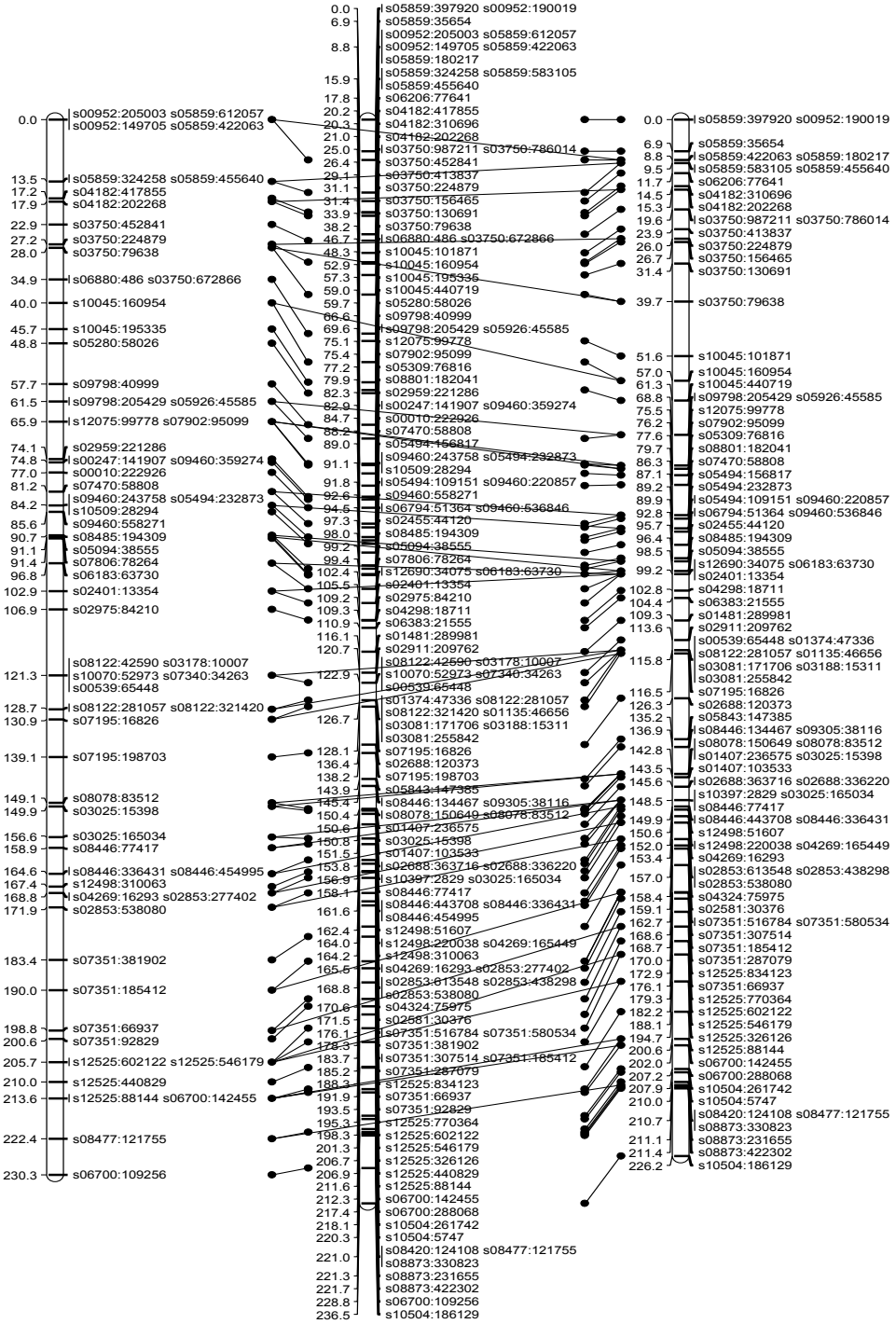


Appendix II continued

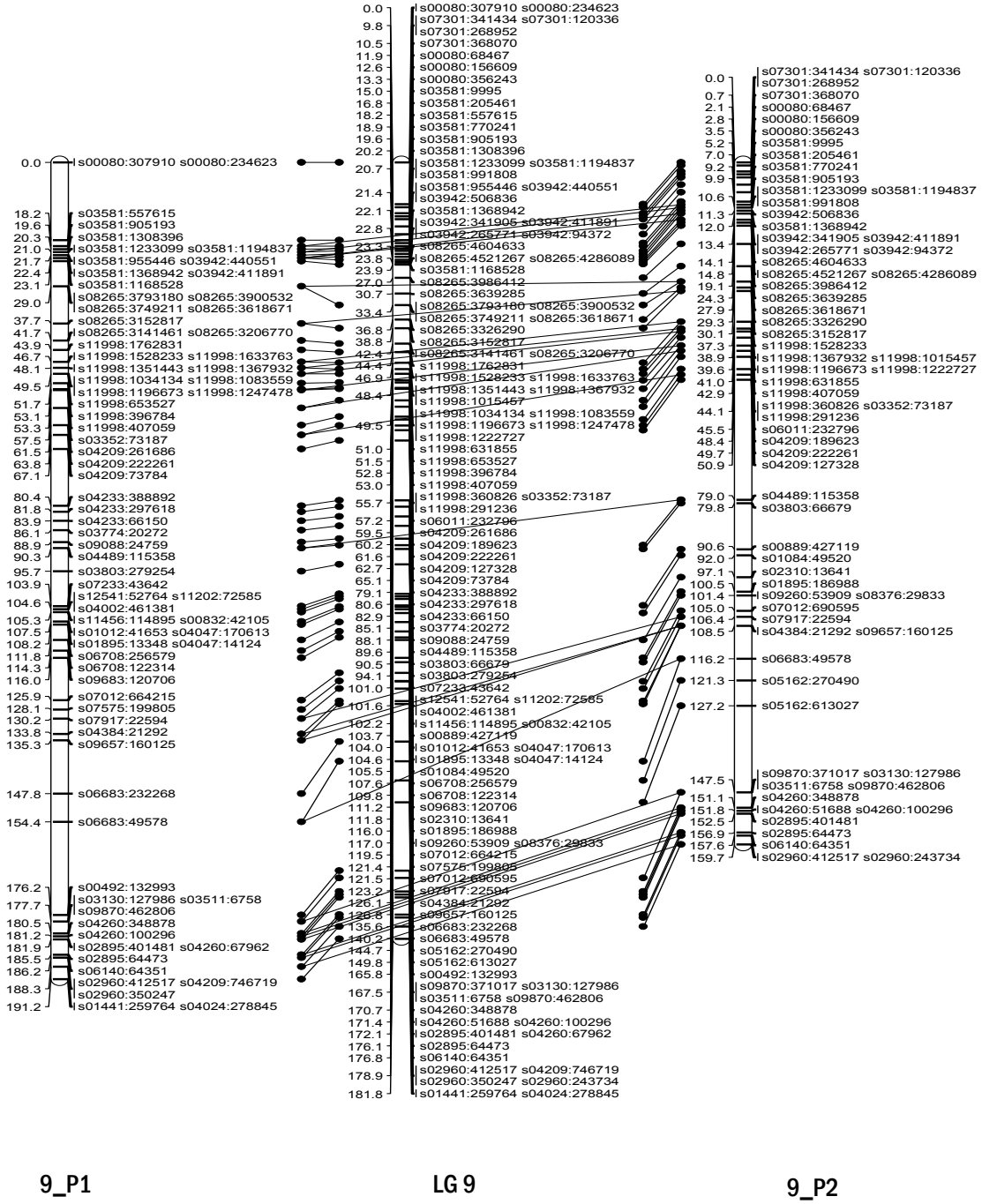


7_P1 LG 7 7_ P2

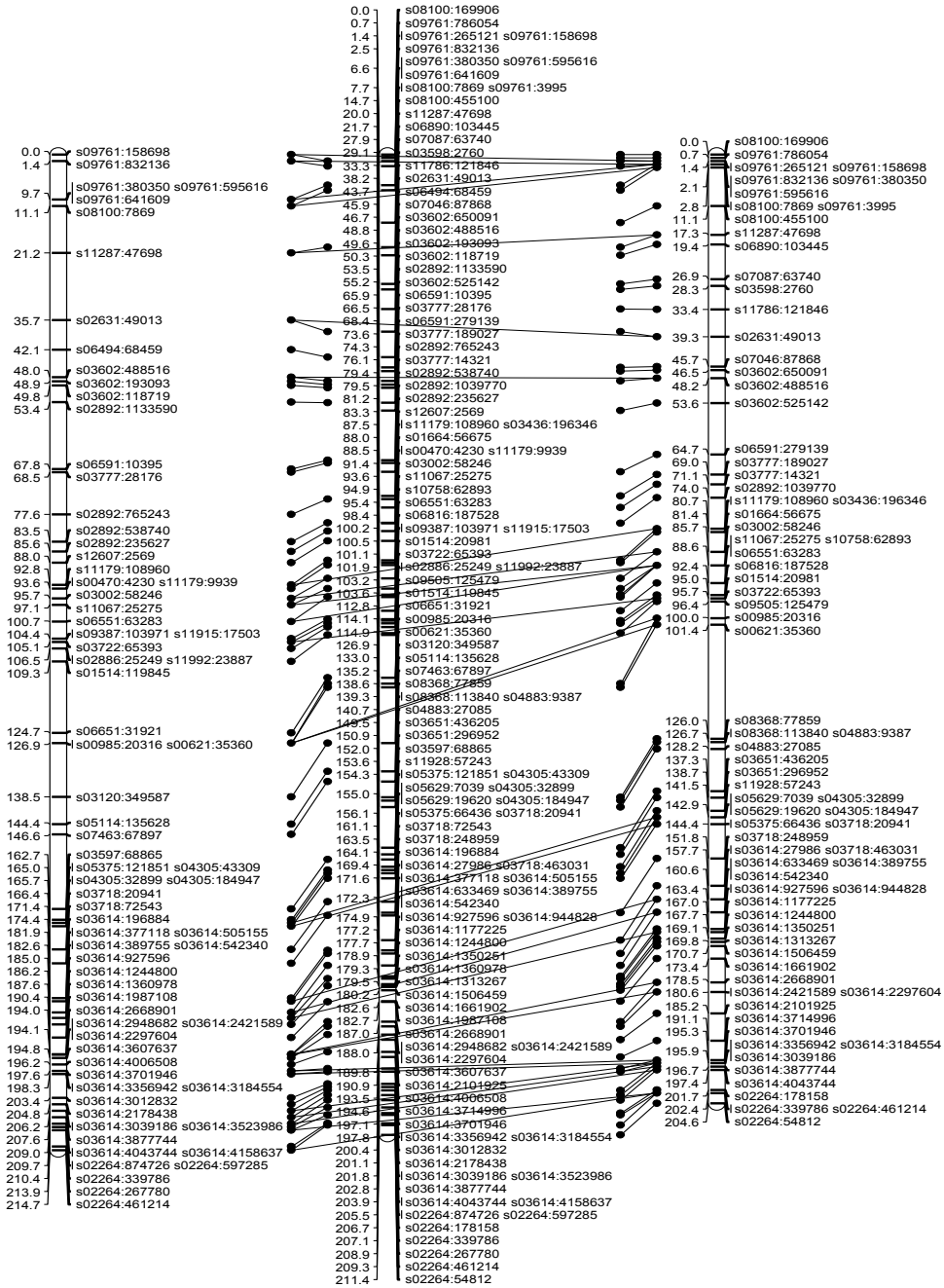
Appendix II continued



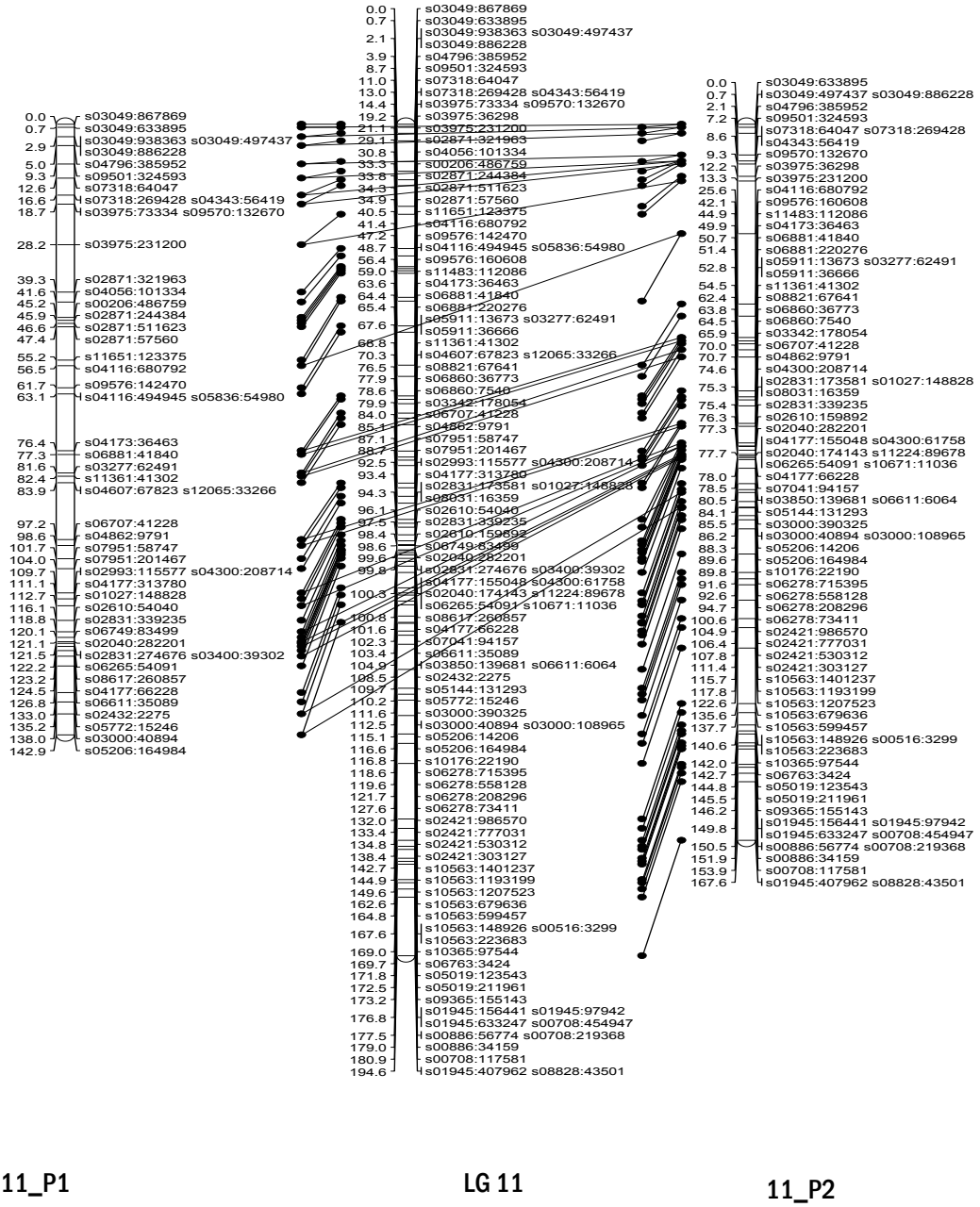
Appendix II continued



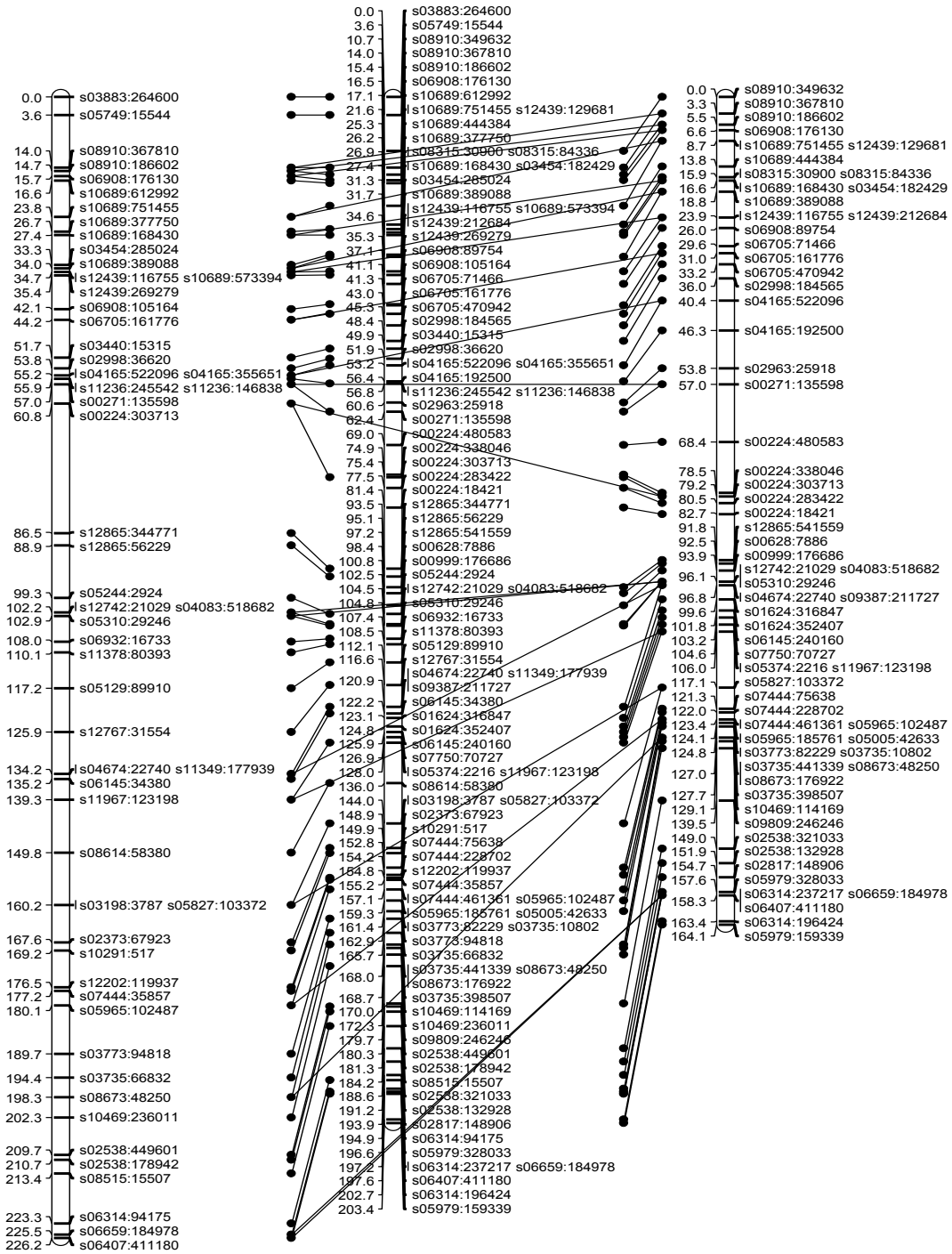
Appendix II continued



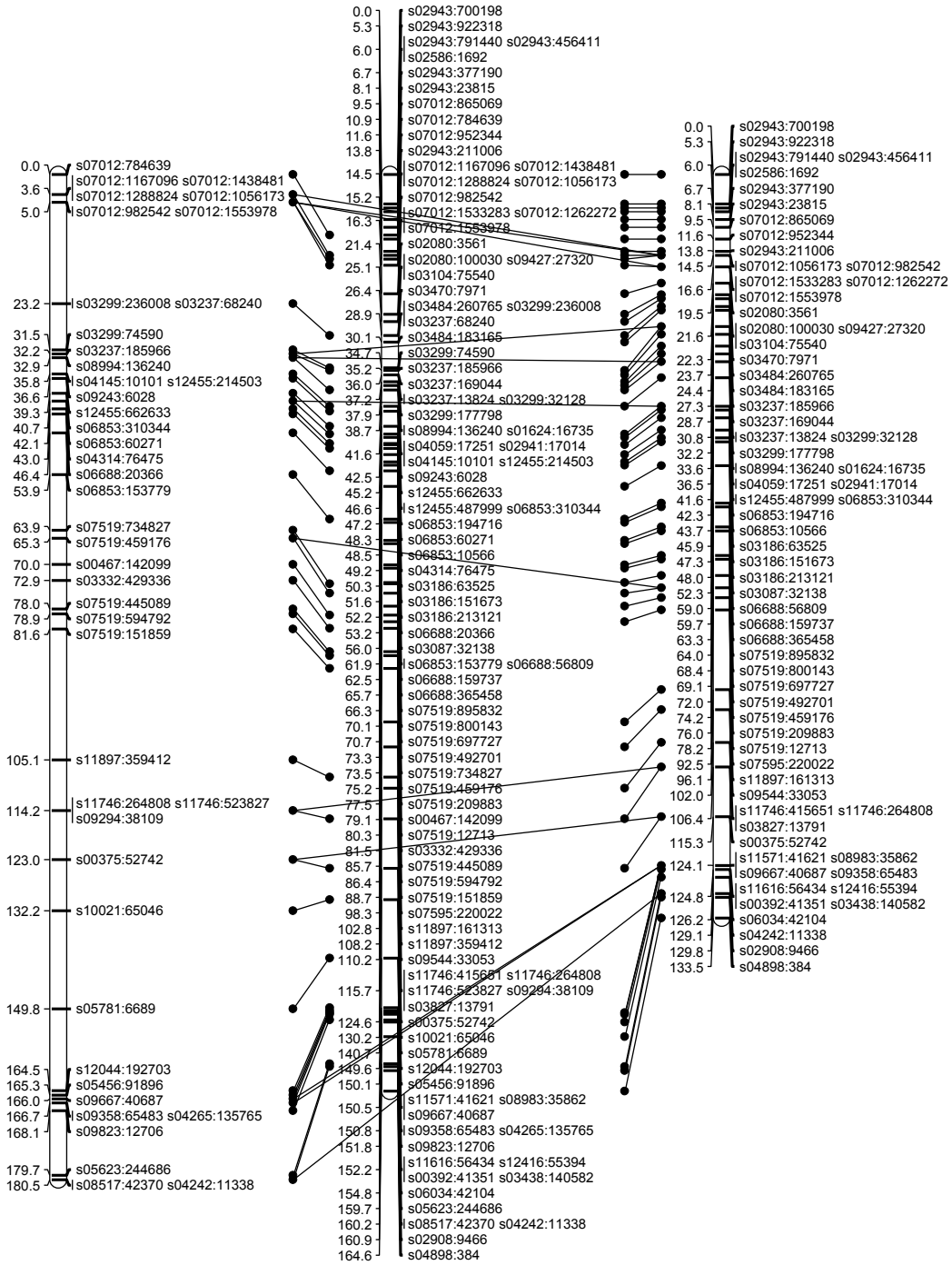
Appendix II continued



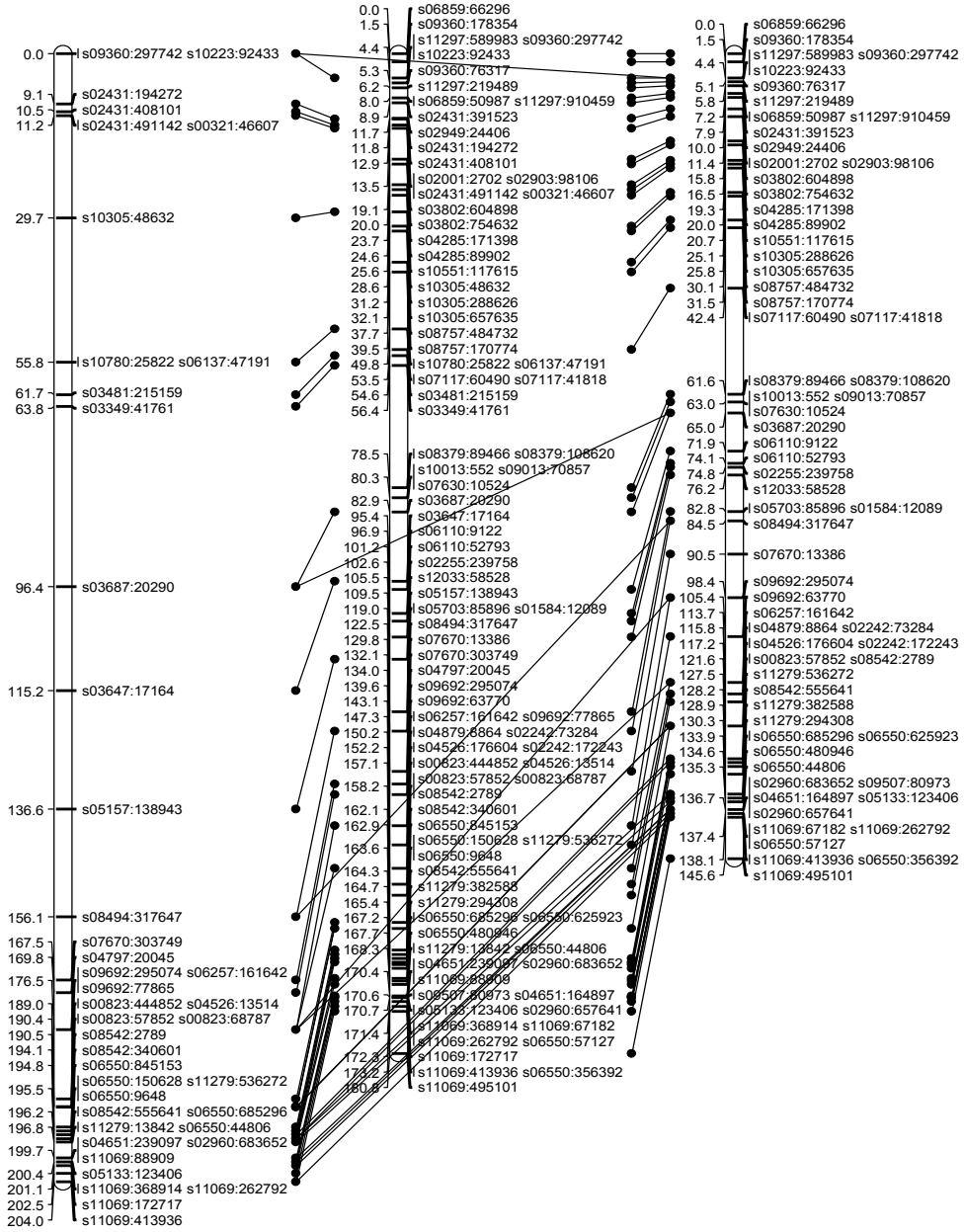
Appendix II continued



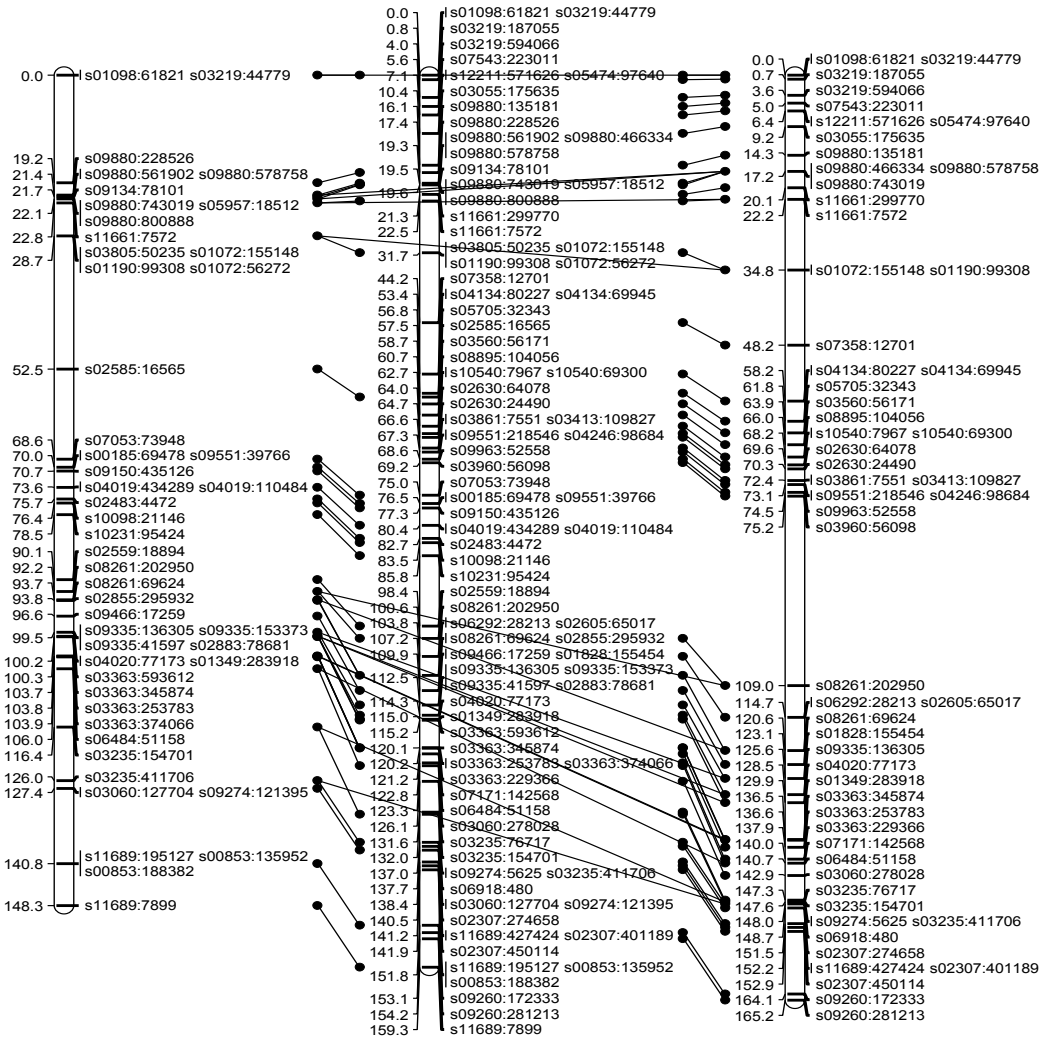
Appendix II continued



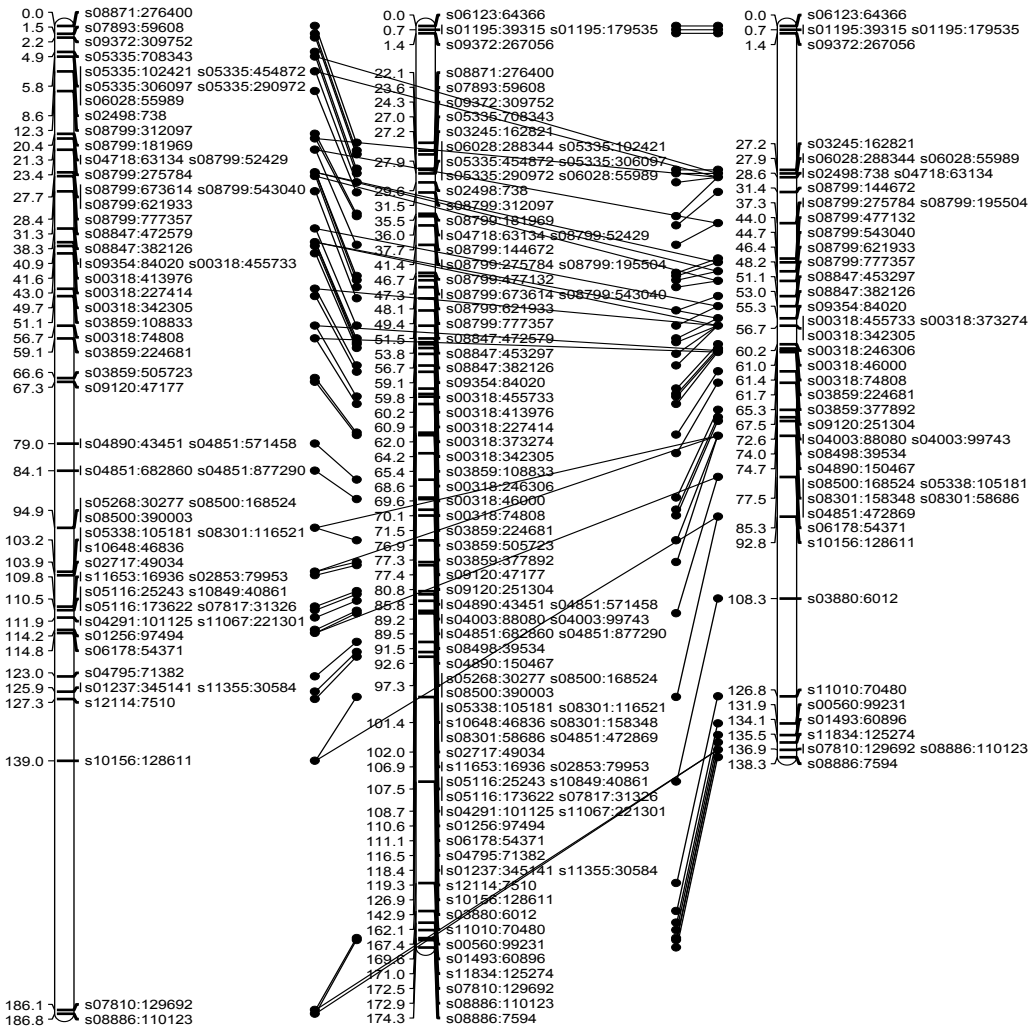
Appendix II continued



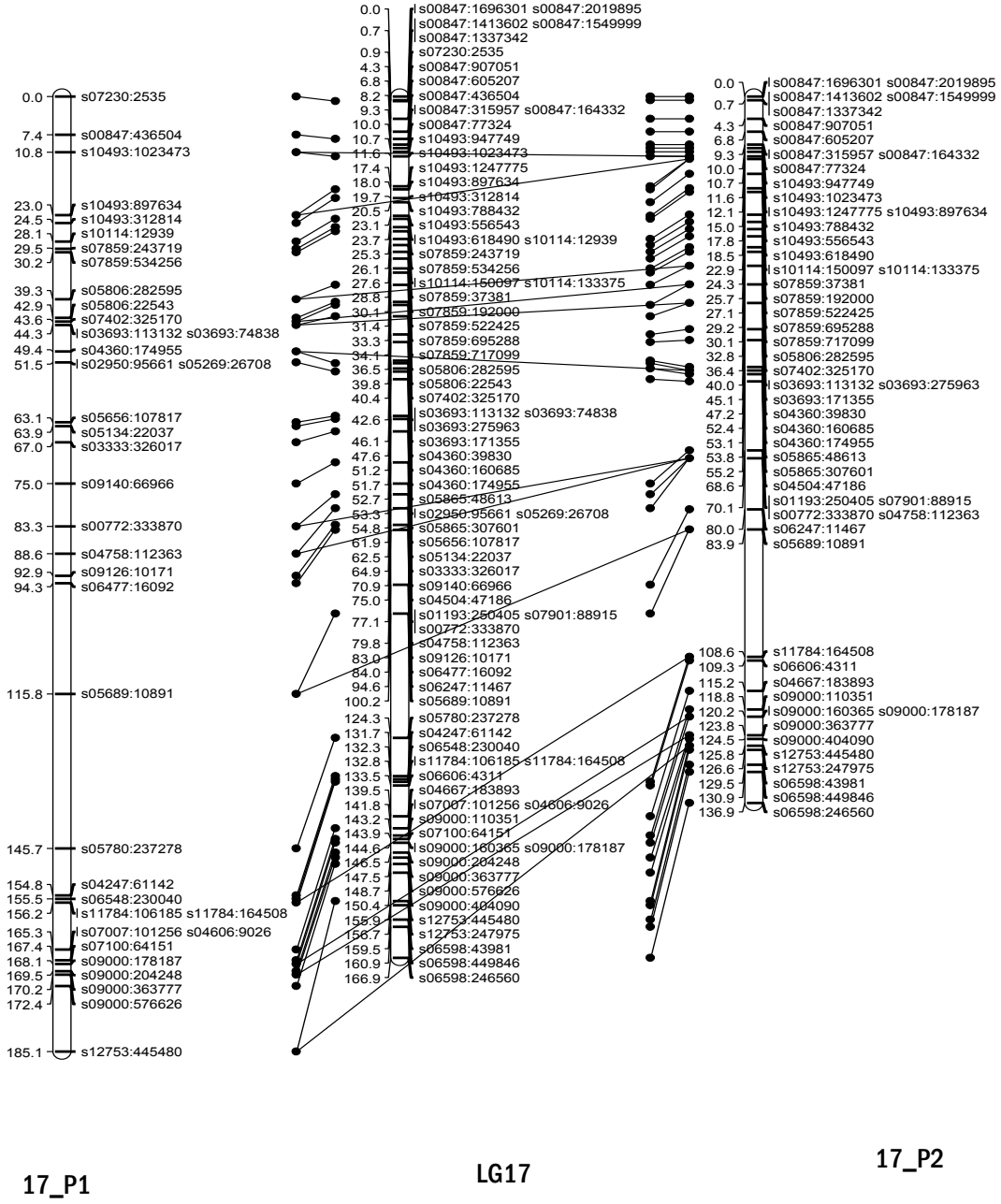
Appendix II continued



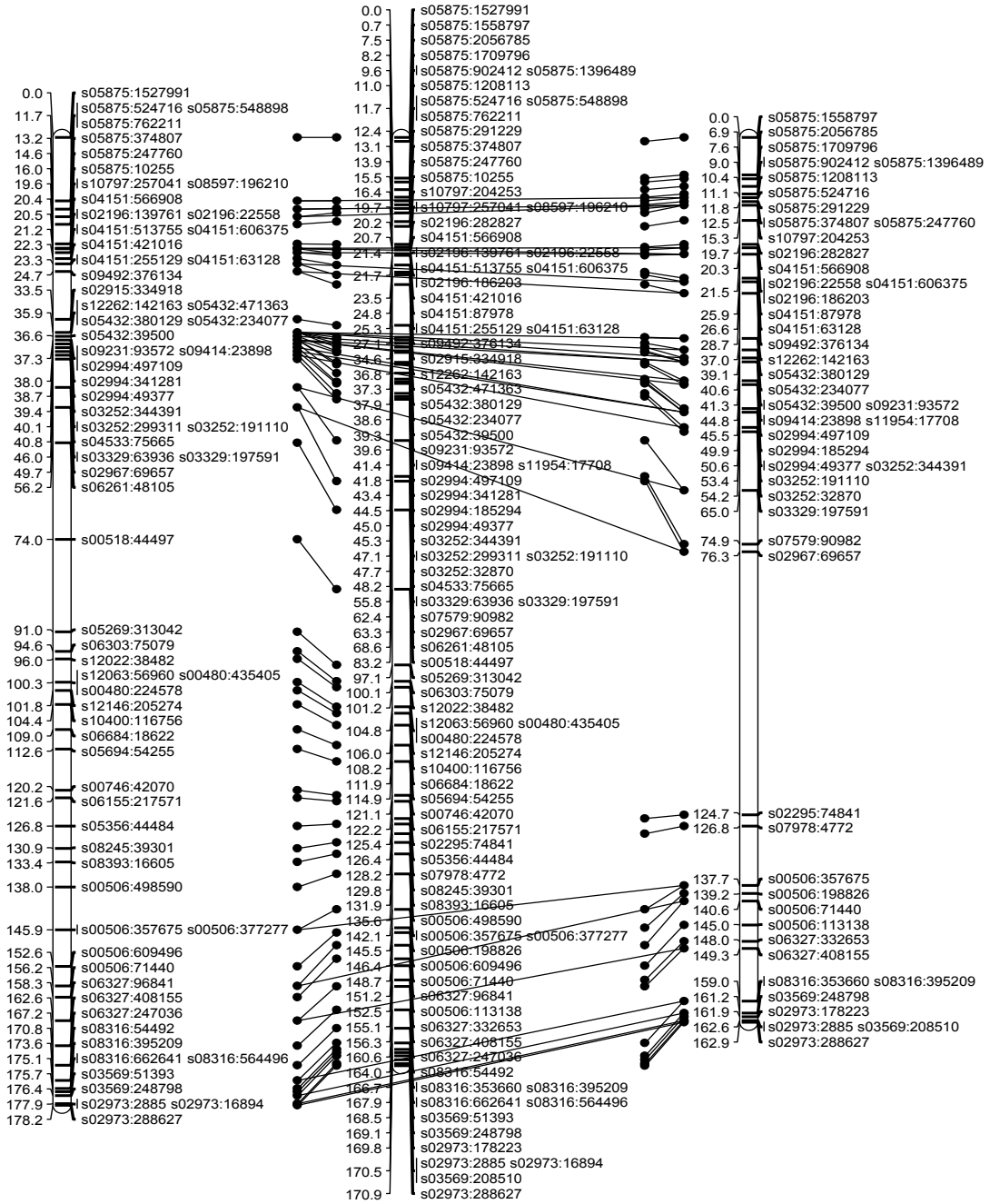
Appendix II continued



Appendix II continued



Appendix II continued



18_P1 LG 18 18_P2