SNP-based genetic linkage mapping and whole-genome diversity assessment in cassava (*Manihot esculenta* Crantz) accessions from Africa and Latin America

by

Sharon Melissa Reynolds

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

July 2017

Under the supervision of Prof. Alexander A. Myburg

and co-supervision of Dr Morag E. Ferguson

Declaration

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree of MSc (Genetics) to the University of Pretoria, contains my own independent work and has not been submitted for any degree at any other university.

Sharon Melissa Reynolds

July 2017

Thesis Summary

SNP-based genetic linkage mapping and whole-genome diversity assessment in cassava (*Manihot esculenta* Crantz) accessions from Africa and Latin America

Sharon Melissa Reynolds

Supervised by **Prof. A.A. Myburg**

Co-supervised by **Dr M.E. Ferguson** (International Institute for Tropical Agriculture, Nairobi, Kenya)

Submitted in partial fulfilment of the requirements for the degree ***Magister Scientiae***

Department of Genetics

University of Pretoria

Summary

Cassava is a staple crop grown in the tropics and supplies food to over half a billion people. It is recognised as a valuable crop for food security in the face of climate change, since it grows in resource poor areas and displays tolerance to drought conditions. Furthermore, the high starch content of the storage roots makes cassava a suitable feedstock for biofuel production. Considerable advancements have been made towards the development of cassava genomic resources, such as the release of the first draft genome sequence in 2012, leading to the identification of new DNA markers and the use of genotyping-by-sequencing technology.

The aim of this MSc was to i) construct a genetic linkage map using single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) markers implemented in a drought-tolerant mapping

population (MCOL1734 x VEN77), ii) assess genetic diversity and genetic structure within African and Latin American cassava landraces and breeding varieties using SNP markers, iii) to assess genome-wide patterns of diversity, and iv) to select 128 SNP markers for routine genotyping of cassava germplasm.

Genetic diversity and structure was assessed within genetic material from African and Latin American germplasm collections, research institutes and breeding programmes. Three potentially heterotic genetic groupings were identified predominantly representing: a) Latin American germplasm, b) West and East African improved varieties, and c) East African landraces. Interestingly, one West African landrace displayed higher than average relatedness to landraces from East Africa, suggesting that it was widely distributed as a favourable variety and hybridised to local landraces in East Africa. Furthermore, evidence of mislabelling and redundancy was observed, highlighting the need for improved management of germplasm collections through the use of molecular markers.

A genetic linkage map, representing 23 linkage groups, was constructed using 443 SNP and 73 SSR markers. Alignment to the *Manihot esculenta v6.1* genome assembly revealed regions of the genome with poor marker coverage, resulting in some chromosomes being split into several small linkage groups. Minor allele frequency for all mapped SNP markers was plotted to assess the genome-wide distribution of diversity for each genetic group. This revealed one region of the genome with low heterozygosity across all three groups and one region displaying divergence in the East African landraces, likely as a result of local adaptation and selection by farmers. Using this data, 128 highly informative SNP markers, distributed throughout the genome, were selected as candidates for routine genotyping in cassava.

This study contributes to our understanding of genetic diversity and structure in African and Latin American cassava landraces and improved varieties. Furthermore, two new molecular breeding tools were developed: a) a genetic linkage map, constructed using a drought-tolerant mapping population, will provide future opportunities to identify markers associated with this important trait for marker-assisted breeding, and b) a set of informative SNP markers that will provide a cost-effective tool for routine genotyping for improved management of germplasm collections and breeding applications, such as the identification of genetically distant accessions for controlled crosses, pedigree reconstruction and maintenance of identity of accessions.

Preface

The development and uptake of cassava genetic resources and molecular breeding strategies in cassava has long lagged behind that of other staple crops for various reasons. However, in the past decade significant advances have been made to develop genomic resources and molecular breeding tools for cassava improvement. In 2012, the first draft cassava genome sequence was released by Prochnick *et al.* (2012, Tropical Plant Biology **5**: 88) which provided the opportunity to develop molecular marker resources and make use of genotyping-by-sequencing technology. In recent years, single nucleotide polymorphism (SNP) markers have become the DNA marker technology of choice in many crop species, including cassava. Several studies have reported the development of SNP marker resources and their applications in genetic mapping studies to identify markers associated with traits of interest, particularly for resistance to cassava mosaic disease (CMD). This is of specific importance since CMD has long been a major threat to cassava production in Africa and has been the focus of several breeding efforts aimed at making use of resistance conferred by wild relatives and using marker-assisted breeding to introgress resistance into favourable genotypes displaying susceptibility to CMD.

Cassava genetic diversity in Africa has been affected by many factors such as multiple introductions, disease pressure and other challenges to cassava cultivation, breeding efforts to overcome these challenges, propagation practices, and sharing of germplasm among farmers from different regions. In addition, introgression and hybridization with wild relatives, through natural hybridization in the field and in formal breeding programmes, in African and Latin American germplasm collections, has influenced genetic diversity in cassava. Furthermore, evidence of a population bottleneck in African germplasm has been reported, however, the lack of diverse, resistant genotypes has limited the introduction of additional diversity from Latin America into Africa. Other studies have shown that the diversity in certain regions is poorly represented in germplasm collections, and is therefore at risk of being lost. It is therefore essential to characterise

cassava genetic diversity to exploit it for breeding purposes and to maintain diverse germplasm in non-redundant germplasm collections at multiple sites for future breeding requirements.

The aims of this MSc study were to add to our understanding of the genetic structure and diversity in cassava landraces and improved varieties maintained in germplasm collections and breeding programmes in Latin America and Africa, to develop a SNP-based genetic linkage map for future use in QTL mapping for drought tolerance, and to combine this information to assess genetic diversity at a whole-genome level. Furthermore, using the data generated in this study, I was able to select a subset of SNP markers for use in routine genotyping applications to improve cassava breeding, and as a tool to better manage cassava germplasm collections.

Chapter 1 of this thesis provides a review of current literature and focusses on molecular markers and their application in cassava improvement. I discuss the challenges experienced in cassava cultivation, breeding efforts to address these challenges and the effect on genetic diversity in Africa. In addition, I focus on cassava genetic resources, and particularly the use of molecular markers to understand cassava biology, and as a tool for molecular breeding for cassava improvement and conservation of diversity for future breeding requirements. Furthermore, I highlight the factors limiting the uptake of molecular breeding strategies in cassava and the research efforts currently underway to address these issues.

In **Chapter 2**, I report the development of a SNP-based genetic linkage map using a drought-tolerant mapping population, which, when combined with phenotypic information, could be used in the future for QTL mapping. Furthermore, using the same set of SNP markers, I describe the genetic structure and diversity within cassava accessions from breeding programmes and germplasm collections in Latin America (Brazil and Colombia) and Africa (Ghana, Nigeria, Uganda

and South Africa). Genetic diversity parameters for genetic groups of accessions identified in this study were combined with the mapped position of the markers to assess the distribution of diversity throughout the genome, with the view to identify regions of the genome displaying divergence between the genetic groups, as a result of selection and adaptation. Using this information, I was able to select a subset of 128 SNP markers, distributed throughout the genome, as candidates for routine genotyping applications, providing a cost-effective alternative to genotyping with large sets of SNPs or GBS technology for cassava molecular breeding. The outputs and applications of the study are further discussed in the section entitled "Concluding Remarks".

This MSc study was undertaken as a part-time degree in the Department of Genetics, at the University of Pretoria, from February 2011 to January 2017. The study was supervised by Prof. A.A. Myburg (University of Pretoria) and co-supervised by Dr Morag Ferguson (International Institute of Tropical Agriculture (IITA), Nairobi, Kenya). Plant material for the mapping population was cultivated by researchers at the International Centre for Tropical Agriculture (CIAT), in Colombia. All SNP genotyping was performed at the University of Pretoria and all SSR genotyping of the mapping population used in this study was performed at International Institute of Tropical Agriculture (IITA) by Dr Rosemary Mutegi-Murori as part of her PhD study. Chapter 2 has been prepared in the format of a manuscript for submission to a peer-reviewed journal (*Molecular Breeding*), and as such, there may be some redundancy between Chapter 1 and the introduction to Chapter 2.

The following conference proceedings have emanated from this MSc work:

International:

MELISSA REYNOLDS, Daleen van Dyk, Mingcheng Luo, Emmanuel Okogbenin, Luis Augusto Becerra Lopez-Lavalle, Morag Ferguson, Alexander Myburg, Pablo Rabinowicz. 2011. Preliminary genetic linkage mapping of Cassava (*Manihot esculenta*) using SNP markers. Generation Challenge Program General Research Meeting, Hyderabad, India, 21-25 September 2011. (Poster presentation)

National (South Africa):

REYNOLDS SM, Van Dyk MM, Luo M, Okogbenin E, Becerra Lopez-Lavalle LA, Ferguson M, Myburg AA and Rabinowicz P. 2012. SNP genetic linkage maps of cassava (*Manihot esculenta*). South African Association of Botanists Symposium, Pretoria, 15-18 January 2012. (Poster presentation)

REYNOLDS SM, Van Dyk MM, Luo M, Okogbenin E, Becerra Lopez-Lavalle LA, Ferguson M, Myburg AA and Rabinowicz P. 2012. Genetic linkage mapping and whole-genome SNP diversity analysis of cassava (*Manihot esculenta*). South African Genetics & Bioinformatics Society Conference, "The Data-mining Revolution", The Conservatoire, University of Stellenbosch, Stellenbosch, 10-12 September 2012. (Oral presentation)

Acknowledgements

I would like to acknowledge the following people for their involvement in the study and their support throughout the duration of my MSc:

- Dr Pablo Rabinowicz (Principal Investigator) for conceiving the study and co-ordinating all aspects of the study including the development of the SNP genotyping kits and sampling of all material included in this study.

- Dr Luís Augusto Becerra López-Lavalle (CIAT) for overseeing the development of the mapping population at CIAT and ensuring that samples were safely delivered to South Africa.

- Dr Emmanuel Okogbenin (IITA, Nigeria) for co-ordinating sample selection of diverse cassava accessions from Africa, and for your valuable input in providing information and assistance in interpreting the data.

- All those involved in selecting cassava germplasm for the diversity assessment. I am certain that there were many people involved that I was not aware of and therefore have not listed below, but I am thankful to each one for your contributions. Those that I am aware of include:

    - Dr Alfredo Alves & Dr Alberto Vilarinhos (Brazil)

    - Dr Luís Augusto Becerra López-Lavalle (Colombia)

    - Dr Elizabeth Parkes, Prof Safo Kantanka, Dr Joseph Adjabeng & Dr Joe Manu (Ghana)

    - Dr Emmanuel Okogbenin, Dr Chiedozie Egesi & Princess Onyin (Nigeria)

    - Dr Joseph Ndunguru & Dr Geoffrey Mkamilo (Tanzania)

    - Dr Jedidah Danson, Dr Lynelle van Emmenes & Nokuthula Myeza (South Africa)

- Dr Anthony Pariyo (Uganda)

- Dr Rosemary Mutegi-Murori for sharing the SSR genotypes generated for the mapping population as part of your PhD study

- Ryan Vogt and Bradley Theron for your assistance in setting up the BeadXpress genotyping facility. Particularly, to Bradley for always being available when I needed help, even at night and over weekends, for practically living at UP at one point, and for always making us laugh despite the difficulties in establishing this platform. You made a troublesome machine enjoyable and we always appreciated your visits!

- Dr Morag Ferguson for your supervision, assistance and valuable insight in every aspect of the study. You are truly deserving of your excellent reputation as a humble scientist and leader, and I consider you to be an amazing example of what a woman in science should be. You have always been patient, understanding and encouraging, and most of all, so enjoyable to work with. I really appreciate everything that you have put into my masters; for reviewing my work in record time in spite of all your responsibilities, for always being ready and willing to help even when you were busy, for providing clarity and leadership when I felt like I was going around in circles, and for making me see the fun in what I was doing.

- Prof. Zander Myburg for your supervision and leadership throughout this study. You have always inspired me to keep going, to strive for excellence and to remember why I love science. I am truly grateful for all your patience, kindness, encouragement, support and grace throughout the duration of my studies and in all the time I have worked for you. You are really a one in a million boss, you have believed in me when I didn't think I could be a success, you have encouraged me to grow, and you have trusted me with much. Most of all, I am grateful for your gracious leadership, for your gentle correction when I have needed it, and for your understanding when I needed space to deal with personal issues. The years that I have taken to complete this degree have undoubtedly been the most difficult

of my life, and I am absolutely certain that I would not have been able to grow through them and find my feet had I not been in such a supportive group. You are truly an amazing and inspirational scientist, boss and friend. I am proud to be associated with FMG and I have learned so much from you. Thank you for being such a wonderful example to us always, for quietly and gladly carrying the heavy burden of a research group as big as FMG, and for daily inspiring us to fall in love with science again and to always do everything with excellence.

- Every single member of the Forest Molecular Genetics group, past and present, that has taken the time to encourage me and support me during my studies. It has been the culture of excellence, the love for science and the enjoyment of a group like ours that has kept me inspired during this time, and I have valued every single word of encouragement. I wish to specifically acknowledge the following people for their contributions and support:

    - Dr Daleen van Dyk for advice and guidance in the early stages of the project and specifically for teaching me how to use JoinMap and helping with the genetic linkage map construction. I learnt a lot from you during the time that you managed our team, and I am grateful that you equipped me to take over the team, even though it was daunting to attempt to fill your shoes.

    - My office mates, past and present, for putting up with me through all the struggles and for supporting me through all the challenges. I have enjoyed the laughter, the encouragement and the many discussions about science and what it means to be a successful scientist. Each one of you have inspired me and I am grateful to have had the pleasure of your company in our little haven.

    - To the leadership of FMG that have not been involved in my project, each of you have inspired me and set a fantastic example of excellence and passion for science, and I am very grateful for your continued encouragement and belief in me, and also for your friendship.

- To Dr Nanette Christie for your advice and assistance with what I consider to be "very scary statistics". Somehow, you have always brought a calmness in the face of the nightmare that is statistics and you have made it simple and manageable. Thank you for your kindness and patience in explaining everything in great detail.

- To Desre Pinard for your assistance with figures. Thank you for giving your time and always being willing to help.

- To Elna Cowley for being the ultimate cheerleader! You have been a constant source of support and encouragement for which I am truly grateful, thank you for believing that I really can do this and never letting me give up.

- To the DNA marker team, some of my favourite people in the world! You have consistently been supportive and patient when my head was filled with the fog of thesis writing, or when I just couldn't figure out something and was a moody cloud of frustration. Thank you for providing a happy and supportive space for me and for being my cheerleaders. I have to specifically thank Patience Motaung and Christy Marais for your continued and unwavering support and encouragement even when you had to carry an extra load because I was just not able to do so. You have become very precious friends to me and I am grateful daily for your presence and support. I am very blessed to have such a wonderful team of people to manage, you make the burden worth it. I hope to have many more "secret-team-building-cupcake-sessions" with you!

- To the students who have worked alongside the DNA marker team (2.0 (Stephan Henning), Stephanus (Stephan Engelbrecht) and Juju (Junior Tii-Kuzu)), you have challenged and inspired me to be a better scientist and to read more! Many times you have come to me and asked me questions that I could not answer, or showed me something in your projects that has made me think more critically about my

own work. My thesis is better because of your work and I have really enjoyed being able to learn alongside you and contribute to your growth as scientists.

- Minique De Castro and Nicky Creux, who were an inspiration in the early days of my degree, and who, to my great surprise, actually thought I had what it took to do this degree! You have both inspired me to love science and set such an amazing example of teamwork within FMG.

- Prof. Paulette Bloomer for advice, input and encouragement with this study. Thank you for your leadership of the Department of Genetics and your continued belief in the success and abilities of everyone that you lead. Thank you for creating a great environment for young scientists to grow and make mistakes.

- To every member of the genetics department who has encouraged me, inspired me, taught me, corrected me and made me into a better scientist. I feel very blessed to be able to grow and learn in an environment like ours.

- To all my friends who listened endlessly and patiently to my troubles, who encouraged me and for some reason thought I was extremely intelligent! This has been the most difficult season of my life and I know I would not have made it through without your continued support. I wish to specifically honour the following friends for all that you have been to me:

  - Ronishree Mangwanda and Katrin Fitza for your support and friendship. You both know the struggles and doubts that accompany us throughout the journey of a degree and you have continually encouraged me to keep going and to finish this and not give up. Your friendship means so much to me and I value you both so much. Thank you for including me in your group and for patiently understanding when I was too busy to socialise outside of work.

- Zama Siqalaba, for paddling the "part-time masters" boat with me. You are one of the few friends who truly understands the demands of a full-time job and a part-time masters. You have always provided the comic relief and made me laugh at the absurdity of my situation many times over. Your undemanding friendship means the world to me and I always enjoy hearing my phone ring and seeing that it's you. Even in the darkest valleys of my life you have made me laugh and I love that about you.

- Kerryn Hugo, my oldest and nerdiest friend. You are really no longer a friend, but a most treasured sister. You have always been understanding when I am too busy and you have never complained about the demands that a part-time degree has put on our friendship. Thank you for pretending to be interested when I talk about work, and for being the best nerdy friend I could wish for! Your delight over all things nerdy always makes me smile! You have been a consistent, generous and loving friend since preschool and I don't know what I would do without you.

- Belinda Adigun, my sweetest friend and sister. You have always been patient and understanding with me even when it cost you greatly. Your consistent, generous friendship means the world to me. You have always been an amazing example of a forgiving, gentle, kind and gracious friend and you challenge me every time we talk. Thank you for never judging me and for persisting with me when I haven't had it in me to be a good friend to you. You have consistently loved me and encouraged me and gone way beyond all expectations to make me feel valued. I am truly blessed to have you in my life.

- Marja O'Neill, you have been the friend that has been consistently present and I can honestly say that you have carried me through the last few years. I am grateful beyond measure for your friendship, your brilliant sense of humour, your strength, your very soft and compassionate heart, your patience and empathy, your consistency, your faith and encouragement, your rational approach to life, the tough

truth you have spoken to me, your gentle and considerate correction, and countless other things. I consider it to be a great honour to work with such a treasured friend. At work you have always inspired me with your strength, professionalism, unrivalled efficiency and brilliant brain! You have given me the space to struggle through my process when I have needed it, you have corrected me firmly when necessary, and you have celebrated my victories with me. Of all my friends, I know that you have had to absorb the most, listen the most and advise the most, and I appreciate the way you have done so patiently and consistently, even when you have had your own burdens to bear. You have always been there for me and I know that I would not have found my feet after losing my father if it had not been for you. Thank you for taking the time to really listen and understand me, and then for giving me the space and freedom to be exactly who I am without judgement and with patience. You really are a precious gift from God and I am very thankful for a friend like you.

- To my family for your continued support and encouragement my whole life, but particularly over the last few years. I am blessed to be surrounded by such loving people who really believe that I can do anything in the world. Most particularly, I am grateful for the amazing parents that I was blessed with. I am thankful to my late father, Dennis, for inspiring me to love maths and science, and for spending hours teaching me and encouraging me to be all that I can be. You were the only person in this world that would have voluntarily read every word of my thesis and googled all the big words to make sure that you properly understood every single detail. Thank you for always being so proud of me, for encouraging me to do the very best that I can, for celebrating every achievement no matter how small, for moving heaven and earth to help me, for teaching me as much as you could with real passion and enjoyment, for taking such an interest in everything that I did, for supporting me constantly and without question, and for instilling an obsession with excellence and a passion for science from a very young age. I am thankful to my mother, Judy, for never letting me go

one minute of my life without feeling like the most special person alive. You have loved me patiently and consistently my whole life and I am proud to call you my best friend. You have always thought that I am so incredibly intelligent and capable. You taught me to think, to observe and to question everything I saw around me, and you instilled a love for reading and for understanding the world around me. But most importantly, you have been a constant example of love, empathy and compassion, and you have taught me how to give that to others. You have always been kind, loving and patient with me, and you have always believed the best of me and been my greatest ally. I am eternally grateful for the privilege of being your daughter and best friend, and I love you with my whole heart.

- To my Heavenly Father for choosing each of these precious people and placing them in my life at the exact time that I have needed them, and for the grace and strength to keep going and finish well.

This thesis is dedicated to my amazing parents, for loving me completely every minute of my life and for making me the person that I am today.

Table of Contents

List of Tables

List of Figures

List of Supplementary Data

List of Electronic Supplementary Files

Electronic Supplementary File S2.1 List of germplasm.

Electronic Supplementary File S2.2 List of SNP markers used to genotype cassava accessions.

Electronic Supplementary File S2.3 List of SSR markers used to genotype cassava accessions.

Electronic Supplementary File S2.4 Diversity parameters for all successful SNP markers.

Electronic Supplementary File S2.5 Duplicates identified within cassava accessions from Latin America and Africa.

Electronic Supplementary File S2.6 Mapped markers that are monomorphic in one or more groups or where the minor allele differs between groups identified in the PCA.

Electronic Supplementary File S2.7 A subset of 128 informative SNP markers for routine genotyping of cassava accessions.

Chapter 1

Literature review:

Molecular markers and their applications for management of genetic resources and crop improvement in cassava (*Manihot esculenta* Crantz)

1.1 Introduction

Cassava (*Manihot esculenta* Crantz) is a perennial plant of the family *Euphorbiaceae* (Balagopalan *et al.* 1988) and is the third most important crop grown in the tropics of Africa, Asia and Latin America as a staple for more than half a billion people (FAO 2014). In 2014, over 270 million metric tonnes (Mt) of cassava roots were produced worldwide with Nigeria being the top producer at 54 million Mt (FAO 2014). Cassava is tolerant to drought and poor soil conditions making it an ideal crop for cultivation on land that is unsuitable for other crops. Furthermore, it has the potential to adapt to altered environmental conditions resulting from climate change (Jarvis *et al.* 2012). As a result, cassava has long been recognised as an important crop for both food security and biofuel production due to the high starch content of the roots (Jansson *et al.* 2009; Zhou and Thomson 2009).

Cassava is a monoecious, protogynous plant (Onwueme 1978) which may grow up to 3-4 m high and consists of a large tuberous root system, a strong stem and large palmate leaves (Silvestre 1989). The disparity in flowering times result in outcrossed, highly heterozygous seedlings. However, germination and seedling survival rates are low and often result in small, weak plants with comparatively low root yields (Onwueme 1978). Consequently, vegetative propagation from stem cuttings of high yielding varieties is the preferred method of cultivation among farmers (Silvestre 1989; Alves 2002). The roots, which are high in starch, are predominantly used for human consumption, animal feeds and as a source of starch for industrial processing (Balagopalan *et al.* 1988). Cassava leaves, which have a higher protein content, are consumed in some regions as a nutrient and mineral rich food source (Silvestre 1989).

Approximately 98 *Manihot* species have been described, with around 40 different species growing in Brazil, the centre of diversity for cassava (Nassar *et al.* 2008). A recent study by Bredeson *et al.* (2016) which included a *Manihot* accession from Brazil believed to be *M. pseudoglaziovii* Pax. &

Hoffm., found that this "species" was actually an interspecific hybrid (*M. esculenta* x *M. glaziovii*). This raises the possibility that some *Manihot* species may in fact represent interspecific hybrids and not separate species. Many studies have been conducted to ascertain the origin of modern cassava varieties. The most widely accepted of these postulates a single progenitor, *M. esculenta* ssp. *flabellifolia* (Olsen and Schaal 1999). Natural hybridization and breeding efforts have resulted in introgression from *M. glaziovii*, and at least one other unknown *Manihot* species, into modern cassava varieties (Bredeson *et al.* 2016). Following its domestication in Latin America, cassava was introduced to West Africa in the 16th century by Portuguese traders as a food source for slaves, and then later to East Africa and Asia (Jones 1959). From these multiple introductions, cassava cultivation spread to many other African countries. In Africa, the only known purely wild relative of cassava is the Ceara rubber tree, *M. glaziovii*, which has been used in breeding programmes to confer disease resistance in *M. esculenta* x *M. glaziovii* hybrids (Hillocks and Jennings 2003; Halsey *et al.* 2008).

Cassava production is constrained by several different abiotic and biotic factors which are presently the focus of much research. Although cassava is tolerant to abiotic stresses such as drought and poor soil conditions (Alves 2002), these factors have the ability to reduce crop yield, particularly in rural farming communities where mitigation measures cannot be easily implemented and farmers are dependent on cassava as a source of food and income. Despite this, a recent study which modelled the effects of climate change on cassava production in Africa (Jarvis *et al.* 2012) found that cassava is well suited to adapt to predicted changes in temperature which may negatively impact other staple crops. However, improved tolerance to drought is vital to secure cassava as a staple food source within Africa. To address this, the Generation Challenge Programme (GCP) was established in 2003 by the Consultative Group on International Agricultural Research (CGIAR). The primary aim of the GCP was to assess and use the genetic diversity in existing germplasm collections to develop improved varieties of selected economically important crops, focusing specifically on breeding drought-tolerant cultivars which would then be made available to farmers

in developing countries and breeders in National Programmes. The collaboration involved over two hundred network partners from both the developed and developing world allowing scientists in resource poor countries access to genetic diversity harboured in germplasm collections and molecular breeding resources that would otherwise be out of their reach, to develop improved cultivars (http://generationcp.org).

While abiotic factors are a concern for cassava production in the long-term, biotic factors such as cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), both transmitted by whitefly (*Bemisia tabaci*) (Calvert and Thresh 2002), present more immediate threats due to their ability to cause substantial crop losses (Varma and Malathi 2003; Legg *et al.* 2011). The first report of CMD was recorded in East Africa in the 18th century, from which it later spread into Central and West Africa (Legg and Fauquet 2004). This considerable threat to cassava production prompted the establishment of a research initiative in the 1930s at Amani in Tanzania, aimed at breeding CMD resistant cassava varieties through hybridization with a resistant wild relative, *M. glaziovii* (Legg and Fauquet 2004). Several of these hybrids were later adopted into a breeding programme at the International Institute of Tropical Agriculture (IITA) in Nigeria and crossed with West African selections resulting in the CMD resistant Tropical *Manihot* Series (TMS) (Legg and Fauquet 2004). TMS varieties were subsequently distributed to Ghana, and then to Uganda, in response to the threat of a more virulent strain of the East African CMD (Nweke 2009).

Initially, less effort was concentrated on breeding for CBSD resistance due to the limited distribution of the disease. However, several wild relatives conferring resistance were identified at the research station in Amani (Hillocks and Jennings 2003). Breeding efforts to back cross *M. esculenta* to two of these wild relatives, *M. glaziovii* and *M. melanobasis*, resulted in resistant hybrids with favourable characteristics, which were later distributed throughout Africa (Hillocks and Jennings 2003). Over the last decade, reports of CBSD in previously unaffected areas of Central

Africa (Legg *et al.* 2011) have raised concerns about the effects on cassava production in Africa, particularly if it were to spread to West Africa.

Traditionally, conventional breeding approaches based on selection of favourable traits have been applied to produce new varieties able to withstand local biotic and abiotic challenges. However, while valuable to the rural subsistence farmer, this approach is time consuming, costly and requires large field trials (Okogbenin *et al.* 2007; do Carmo *et al.* 2015). With the introduction of DNA marker technologies, many studies have been conducted to understand the biology and genes underlying important traits, providing a foundation for molecular breeding approaches such as marker-assisted selection (MAS) (Akano *et al.* 2002; Lokko *et al.* 2005; Mohan *et al.* 2013). MAS has been valuable in several important crops such as maize (Bouchez *et al.* 2002), wheat (Zhou *et al.* 2003), barley (Jefferies *et al.* 2003) and rice (Hittalmani *et al.* 2000) to develop improved cultivars. This approach has been successfully applied to cassava for the introgression of CMD resistance, conferred by the *CMD2* gene which was discovered in the Tropical *Manihot esculenta* (TME) landraces from Nigeria, into Latin American breeding material using trait-linked DNA markers (Okogbenin *et al.* 2007). More recently, however, approaches such as genetic modification of widely planted varieties for disease resistance (Taylor *et al.* 2012), genomic selection (de Oliveira *et al.* 2012) and the development of larger sets of DNA markers (Ferguson *et al.* 2012b) have been adopted to address key breeding objectives in cassava.

The focus of this review will be to discuss the current status of cassava breeding and genomic resources available to the cassava community. Specific focus will be given to the molecular marker resources available for cassava, genetic linkage mapping, the discovery of quantitative trait loci (QTLs) and their application to marker-assisted breeding (MAB), and genetic diversity studies in cassava. These focus areas are critical for understanding cassava biology and for the improvement of this economically important crop.

1.2 Molecular markers in cassava

Cassava has long been referred to as an "orphan crop" since comparatively little research had been conducted on this important crop, resulting in a paucity of knowledge surrounding the biology underlying key traits in cassava (Ceballos *et al.* 2012). In the past decade, significant advances have been made towards expanding our understanding of cassava biology, such as the release of the draft cassava genome sequence in 2012 (Prochnik *et al.* 2012), and the more recent release of an improved genome assembly, providing a means to study the cassava genome at a chromosomal level (Bredeson *et al.* 2016). Tools such as these provide a valuable resource for further genomics studies to broaden our understanding of the genetics underlying key traits in cassava. However, most critical is the implementation of this knowledge in breeding programmes to produce improved varieties.

Initial studies in cassava predominantly made use of Amplified Fragment Length Polymorphisms (AFLP), Restriction Fragment Length Polymorphisms (RLFP), Random Amplified Polymorphic DNA (RAPD) and isozyme markers (Fregene *et al.* 1997; Fregene *et al.* 2000a; Asante and Offei 2003). However, in the past decade, Simple Sequence Repeat (SSR) markers have been favoured due to their high polymorphic information content and their multiplexing capabilities (de Bang *et al.* 2011). Several sets of SSR markers have been developed for cassava and applied to a variety of studies such as diversity assessment (Fregene *et al.* 2003; Moyib *et al.* 2007; Hurtado *et al.* 2008; de Bang *et al.* 2011), understanding the relationships between domesticated cassava and its wild relatives (Olsen and Schaal 2001), assessing genetic differentiation and substructure within breeding programmes and landraces from different regions (Raji *et al.* 2009; Asare *et al.* 2013), genetic linkage mapping (Mba *et al.* 2001; Okogbenin *et al.* 2006; Kunkeaw *et al.* 2011; Sraphet *et al.* 2011), and identification of QTLs underlying economically important traits (Lokko *et al.* 2005; Lopez *et al.* 2007; Okogbenin *et al.* 2008).

Over the years, SSR markers have been developed and implemented by a number of independent research groups resulting in the same loci being given different names, complicating marker comparison between laboratories (Ferguson *et al.* 2012a). To address this, a recent effort by Ferguson *et al.* (2012a) aimed to identify a non-redundant set of SSR markers. Primers for 3367 SSRs markers were evaluated and a set of 2146 unique SSR markers were identified and made available to the cassava community. Studies such as this are important since it has been noted by Ayling *et al.* (2012) that in order to translate the advances in cassava genomics into improved varieties through the many breeding initiatives, standardised sets of molecular markers and the accompanying technology platforms should be made publically available.

While SSRs have proven to be valuable DNA markers, studies in animals and plant crops have demonstrated that significant gains from molecular breeding approaches rely on high-throughput genotyping using large sets of markers, such as Single Nucleotide Polymorphism (SNP) and Diversity Arrays Technology (DArT) markers that provide dense, genome-wide coverage (Jaccoud *et al.* 2001; Rafalski 2002). In 2005, Xia *et al.* reported the first cassava DArT array, which was developed using approximately 1000 diverse cassava clones. The array was implemented on a set of 31 accessions representing four *Manihot* species to assess the value of these markers in discriminating between the four species. Using these markers, accessions were accurately assigned to the correct species using an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram, and the DArT array was made available to the cassava community as a resource for high-throughput genotyping of cassava. However, subsequent investigation at the International Centre for Tropical Agriculture (CIAT) found SSRs to be more effective than DArT markers in characterizing diversity in cassava germplasm (Hurtado *et al.* 2008). It was noted that the relatively poor performance of the DArT array, when compared to the SSR markers, may be attributed to the fact that the array was designed to represent several *Manihot* species, whereas, an array designed to specifically target cassava may yield more promising results.

More recently, SNP markers have been favoured since they are highly abundant throughout the genome and are well suited to high-throughput, automated genotyping platforms (Liu *et al.* 2005). These characteristics allow for genotyping of large sets of markers, thus compensating for the fact that SNPs are biallelic and therefore less informative individually than SSR markers (Rafalski 2002). SNP frequencies vary between the coding and non-coding regions of the genome and between different species (Kahl *et al.* 2005) with an average of one SNP per 100 to 300 bases in most genomes (Gupta *et al.* 2001). In maize, SNP frequencies have been reported to range from one SNP in 31 bp in the non-coding region to one SNP in 124 bp in the coding region (Ching *et al.* 2002), while in the coding regions of rye (Varshney *et al.* 2007), sugar beet (Schneider *et al.* 2001) and apple (Chagné *et al.* 2008) one SNP was found every 58, 130 and 149 bp, respectively.

Lopez *et al.* (2005) conducted a study using six different cassava cultivars and estimated the SNP frequency to be one SNP in 62 bp in the coding regions of accessions included in the study. In a later study involving only nine genes in 74 diverse cassava cultivars, Kawuki *et al.* (2009) reported the SNP frequency to be one SNP in 121 bp. Several recent studies have reported the discovery of large sets of SNP markers for cassava (Ferguson *et al.* 2012b; Sakurai *et al.* 2013) in addition to using genotyping-by-sequencing (GBS) technology (Rabbi *et al.* 2014a; Rabbi *et al.* 2014b; International Cassava Genetic Map Consortium 2015) for genomic studies, such as genetic linkage mapping and QTL discovery. These resources provide a basis for the dissection of key traits and can be applied to molecular breeding approaches to develop improved varieties.

1.3 Other genomic resources

Several efforts have been made towards developing additional cassava genomics resources over the past decade. Earlier studies reported the development of EST collections (Anderson *et al.* 2004; Lokko *et al.* 2007; Li *et al.* 2010), however, in 2009 a collaborative initiative was established between the RIKEN Institute (Japan), Mahidol University (Thailand) and the International Centre

for Tropical Agriculture (CIAT, Colombia) (Utsumi *et al.* 2012), with the aim of establishing a functional genomics platform for the worldwide cassava community. This was to be achieved by developing cassava genomics resources such as a collection of full-length cDNAs and an Agilent oligoarray consisting of over 30 000 expressed genes, creating a database (Cassava online archive: http://cassava.psc.riken.jp/) to make these resources available to the wider community, and establishing a transformation system for cassava based on an elite cultivar from Asia (Utsumi *et al.* 2012). These objectives will provide valuable resources for cassava improvement. However, perhaps the most notable advancements in cassava genomics in the past decade was the release of the draft cassava genome assembly in 2012 by Prochnick *et al.* (2012), and more recently an improved version (Bredeson *et al.* 2016), which will serve as a valuable reference tools for all further genomics studies, such as the discovery of additional molecular markers reported by Sakurai *et al*. in 2013. Additional work is, however, required to further improve and annotate the genome assembly.

1.4 Applications of molecular markers in cassava

1.4.1 Factors affecting the uptake of molecular breeding in cassava

Molecular markers have been widely applied to the improvement of many economically important crops and animal breeds (van Marle-Koster *et al.* 2013; Win *et al.* 2013; Liu *et al.* 2015). In cassava, molecular markers have been used to identify regions of the genome associated with key traits such as growth (Boonchanawiwat *et al.* 2011), disease tolerance (Jorge *et al.* 2001) and post-harvest deterioration (Cortés *et al.* 2002). Studies such as these provide the basis for MAB which aims to accelerate genetic improvement and address future threats to cassava production. Furthermore, molecular markers can be used for molecular breeding strategies such as genomic selection, where prediction of the future performance of an individual allows for selection of superior plants at the seedling stage, thus reducing the need for lengthy, costly trials (de Oliveira *et al.* 2012). Molecular markers also provide an essential tool for the management of germplasm collections and for understanding and exploiting diversity for breeding purposes (Ceballos *et al.* 2015).

These applications demonstrate the value of molecular marker technologies to complement traditional breeding approaches and have led to significant gains in many crop species, however, the adoption of molecular breeding strategies in cassava breeding programmes has been limited for several reasons. In outbred crops such as maize that have been largely purged of inbreeding depression, inbred lines are frequently used to identify marker-trait associations and as breeding resources to maximise heterosis (Reif *et al.* 2005). Recombinant inbred lines (RILs) are a valuable resource for genetic mapping studies since replicated trials of the same genotypes can be phenotyped for many traits across several sites (Simon *et al.* 2008), increasing the accuracy of phenotypes (breeding values) for detection and validation of marker-trait associations. However, cassava is a highly heterozygous, outcrossing crop reported to suffer greatly from inbreeding depression (Okogbenin *et al.* 2006), which has limited the use of inbreeding strategies for cassava improvement in the past. Nevertheless, in recent years, studies addressing these limitations have been successful, making the use of inbred lines a feasible option for cassava improvement into the future (Ceballos *et al.* 2015).

Another limiting factor in cassava molecular breeding is the development of large mapping populations. Flowering time in cassava is highly variable and depends on the genotype and optimal environmental conditions (Halsey *et al.* 2008). Asynchronous flowering limits the generation of crosses by controlled pollination for breeding and the development of discovery populations (Ceballos *et al.* 2004). Furthermore, flowers typically develop at branching points along the stems. Tall, non-branching phenotypes are preferred by farmers, and as a result, many of the preferred cultivars may be poor flowerers with low seed set, limiting the number of seedlings that can be generated for genetic mapping purposes (Ceballos *et al.* 2004). An initiative called the Next Generation Cassava Breeding Project (NextGen Cassava), has been established at Cornell University (Ithaca, NY) to address several issues related to cassava improvement, one of which is to identify methods of inducing flowering which will allow poor flowering cultivars to be used for breeding purposes (http://www.nextgencassava.org/).

Until recently, identification of marker-trait associations, a prerequisite for MAB, has been hindered by the lack of genomic resources (Ceballos *et al.* 2015). Furthermore, rather than representing multiple diverse genetic backgrounds, many of the studies conducted in cassava to identify QTLs have made use of mapping populations derived from only two crosses (Fregene *et al.* 1997; Okogbenin *et al.* 2006; Kunkeaw *et al.* 2011; Sraphet *et al.* 2011; Whankaew *et al.* 2011; Soto *et al.* 2015). Moreover, large sets of markers such as SNP markers were not available for high-throughput, genome-wide genotyping; a prerequisite for applications such as genome-wide association studies (GWAS) and genomic selection (GS) (Varshney *et al.* 2012). In recent years efforts have been focussed on generating larger sets of markers such as SNPs which are favoured for molecular breeding applications (Rafalski 2002). In addition, several recent studies have made use of genotyping-by-sequencing (GBS) technology, which is becoming an increasingly popular approach (Rabbi *et al.* 2014a; Rabbi *et al.* 2014b; International Cassava Genetic Map Consortium 2015). A key focus of the NextGen Cassava project is to make use of these new genomic resources to develop and validate tools for genomic selection, and to empower breeders to make use of the available resources to further cassava improvement (http://www.nextgencassava.org/).

1.4.2 Genetic linkage mapping and QTL discovery in cassava

Over the years several studies have been conducted to construct genetic linkage maps for cassava (Table 1.1), however, many of the published maps have been based on only two biparental mapping populations. The first population was produced by crossing TMS 30572 (female parent) and CM 2177-2 (male parent) (Fregene *et al.* 1997; Okogbenin *et al.* 2006; Soto *et al.* 2015). TMS 30572 is an elite clone from the IITA in Nigeria that displays high tolerance to CMD derived from ancestral hybridization to *M. glaziovii*, and tolerance to cassava bacterial blight (CBB) (Fregene *et al.* 1997). CM 2177-2 is an improved line from CIAT that is favoured for its cooking quality, high photosynthetic rates and tolerance to CBB and cassava mealy bug (Fregene *et al.* 1997). The first framework genetic map constructed using this cross was published in 1997 by Fregene *et al.* and

consisted predominantly of RFLP and RAPD markers. This was later saturated further with SSR markers by Mba *et al*. (2001) and Lopez *et al.* (2007).

In 2006, Okogbenin *et al.* published a linkage map consisting of SSR markers using an F_2 inbred population (K150) derived by selfing a seedling from the TMS 30572 x CM2177-2 F_1 cross. More recently this F_1 population was genotyped using the GBS approach by Soto *et al.* (2015) and a map constructed using 2141 SNP markers incorporated into 18 linkage groups, corresponding to the haploid number of cassava chromosomes. This map was integrated with the physical map to anchor scaffolds representing approximately half of the cassava genome (45.6%). High colinearity was observed between this map and two maps published in 2014 by Rabbi *et al*., and QTLs identified for CMD resistance were mapped to the same position as reported by Rabbi *et al.* (2014b). The map was further used as a tool to map immunity-related proteins. This mapping population has formed the basis of several other QTL mapping studies for traits such as early yield (Okogbenin *et al.* 2008), early root bulking (Okogbenin and Fregene 2002), productivity and plant architecture (Okogbenin and Fregene 2003) and resistance to cassava bacterial blight (CBB) (Jorge *et al.* 2001; Lopez *et al.* 2007).

The second widely used mapping population consists of 100 F_1 progeny derived from a cross between two Thai cultivars; Huay Bong 60 and Hanatee (Kunkeaw *et al.* 2011; Sraphet *et al.* 2011; Whankaew *et al.* 2011). Huay Bong 60 has several traits of importance for industrial processing such as high root yield, starch and cyanide content, while Hanatee has lower cyanide and root starch content but excellent cooking quality (Sraphet *et al.* 2011). Three linkage maps were published in 2011 using this mapping population (Kunkeaw *et al.* 2011; Sraphet *et al.* 2011; Whankaew *et al.* 2011). These maps consisted exclusively of EST-SSR and SSR markers, many of which were added to the first genetic map (Fregene *et al.* 1997) by Mba *et al.* (2001), to allow for estimation of co-linearity between maps. Although these maps consisted of more markers than the previously

published maps and all had intermarker intervals below 6 cM, none provided sufficient coverage to resolve the markers into 18 linkage groups. The concurrent release of the cassava genome sequence (Prochnik *et al.* 2012) provided the first opportunity for anchoring the genetic map developed by Sraphet *et al*. (2011) to the draft of the cassava genome sequence, revealing under-represented regions of the genome which could be targeted for further marker discovery. This mapping population was also later used for QTL studies for traits such as plant and first branch height (Boonchanawiwat *et al.* 2011) and starch pasting viscosity (Thanyasiriwat *et al.* 2014).

Two other linkage maps using predominantly AFLP markers, but also incorporating a small number of SSR markers, were published in 2010 (Chen *et al.* 2010; Kunkeaw *et al.* 2010). These maps were based on different F_1 pedigrees derived from popular Chinese and Thai cultivars. The first, published by Kunkeaw *et al.* (2010) was based on an F_1 hybrid mapping population (Rayong 90 x Rayong 5) consisting of only 58 progeny and made use of AFLP markers and the SSR markers published previously by Mba *et al.* (2001), arranged on 33 linkage groups. The excess of linkage groups and the small population size limited the application of this map to further studies. The second map published in 2010 (Chen *et al.* 2010) made use of an F_1 mapping population arising from a cross between an elite cultivar, SC6 (female parent), which is planted widely in southern China and is known for its favourable traits, and Mianbao (male parent) which is a landrace exhibiting tolerance to abiotic stresses present in Hainan Island. This map represented the first published genetic map for cassava with sufficient genome-wide marker coverage, resolution and linkage to construct a map with 18 linkage groups; equal to the haploid number of cassava chromosomes, and with an intermarker interval below 5 cM. A variety of markers were incorporated into this map, including AFLP, Sequence-related Amplified Polymorphisms (SRAP) and SSR markers, some of which were incorporated into the first genetic map (Fregene *et al.* 1997) by Lopez *et al.* (2007), to allow for assessment of co-linearity between the two maps.

The release of the draft cassava genome sequence (Prochnik *et al.* 2012) provided an opportunity for the discovery of large sets of SNP markers (Ferguson *et al.* 2012b; Sakurai *et al.* 2013) and more recently, for GBS technology to be implemented for cassava genomics studies (Rabbi *et al.* 2014a; Soto *et al.* 2015). The last few years have seen a number of cassava genetic maps being published using more advanced genotyping technologies implemented in new and diverse mapping populations. Rabbi *et al.* (2012) published a high-density SNP linkage map consisting of SNP and SSR markers, using an F_1 mapping population derived by crossing two East African cultivars, Namikonga and Albert. Namikonga, the female parent, is known for its resistance to cassava brown streak disease (CBSD), while the male parent, Albert, is tolerant to CMD and has high flowering potential, but is susceptible to CBSD. The inclusion of SSR markers previously incorporated into the map published by Whankaew *et al.* (2011) provided an opportunity for comparison between the maps and revealed co-linearity between the two, lending higher confidence in the accuracy of the map. Scaffold positions were also identified for approximately 94% of the markers from the linkage maps of Rabbi *et al.* (2012), and in many cases, markers found on the same linkage groups were observed to map to the same scaffolds, demonstrating the value of these maps for both validating and improving the genome assembly.

A further two linkage maps were published by Rabbi *et al.* in 2014 using two different mapping populations. The first consisted of F_1 progeny derived from a cross between two varieties from the IITA in Nigeria; the female parent, TMS IBA961089A, resulting from the cross between TME9 (a Nigerian landrace) and a Nigerian improved variety from the IITA (IITA TMS IBA30572), is known for its resistance to CMD. The male parent was a landrace, TMEB117, which is highly susceptible to CMD (Rabbi *et al.* 2014a). The second study conducted by Rabbi *et al.* in 2014 made use of an F_1 mapping population arising from a cross between IITA-TMS-011412 and IITA-TMS-4(2)1425. The female parent is an improved Nigerian variety (IITA) and is known for its high levels of pro-vitamin A and resistance to CMD. The male parent, which is highly susceptible to CMD, is also an improved line (IITA, Nigeria) resulting from a cross between a Nigerian landrace (TME109

– local name Oyarugbafunfun) and the CMD resistant *M. esculenta* x *M. glaziovii* hybrid (58308) involved in the development of the tropical *Manihot* series (TMS lines) (Rabbi *et al.* 2014b).

Both genetic maps consisted exclusively of SNP markers (772 and 6756 markers, respectively) incorporated into 19 linkage groups, with small intermarker distances (mean of 2.0 and 0.52 cM, respectively). However, in both cases the haploid chromosome number was exceeded by one linkage group, indicating a paucity of markers in certain regions of the genome resulting in insufficient linkage between smaller linkage groups representing different parts of the same chromosome. These maps were further used for QTL discovery for traits such as CMD resistance (Rabbi *et al.* 2014a; Rabbi *et al.* 2014b) and root carotenoid content (Rabbi *et al.* 2014a). All three mapping populations used by Rabbi *et al*. represented diverse germplasm that segregated for key traits such as response to CMD and CBSD, both of which represent significant threats to cassava cultivation in Africa, making these maps ideal resources for QTL discovery.

All but one of the previously published genetic maps for cassava have been based on single F_1 pedigrees. Recently, the International Cassava Genetic Map Consortium (ICGMC) published a consensus genetic map constructed using several F_1 mapping populations and GBS technology (International Cassava Genetic Map Consortium 2015). This linkage map was based on 10 different mapping populations segregating for important traits such as root rot, CBSD, CMD and CGM resistance, and starch and carotenoid content of the roots. Nine of the mapping populations represented F_1 crosses consisting of between 154 and 303 progeny, while the remaining cross represented a selfed line comprising 149 S_1 progeny. The map consisted of 22 403 markers arranged on 18 chromosomes and was used to anchor 71.9% of the genome sequence including 90.7% of the predicted protein-coding genes. This map is the most comprehensive map available for cassava and provides a valuable resource for molecular breeding and future studies to identify markers associated with key traits.

1.4.3 Marker-assisted breeding

Considerable effort has been given to the introgression of genes conferring resistance to CMD into susceptible germplasm (Okogbenin *et al.* 2007; Okogbenin *et al.* 2012). CMD is a potentially devastating disease which is responsible for significant crop losses in Africa, but is not present in Latin America (Legg and Fauquet 2004). Studies conducted to assess genetic diversity have reported higher levels of diversity in Latin American germplasm, when compared to African germplasm, suggesting a population bottleneck (Kawuki *et al.* 2013). CMD and subsequent breeding efforts to combat the effects of CMD, have contributed further to the narrowing of the genetic base within African germplasm (Okogbenin *et al.* 2007). While African breeding programmes display moderate diversity, they do not adequately represent the diversity held in Latin American germplasm collections (Okogbenin *et al.* 2007). African breeding programmes would, therefore, benefit from the introduction of diverse germplasm from Latin American in order to broaden the genetic base and provide further opportunities for the development of improved varieties. However, a study by Okogbenin *et al*. (2007) reported high susceptibility to CMD in elite cultivars from Latin America. This is of great concern should CMD be accidentally introduced into Latin America, as it has the potential to cause widespread crop losses and significantly reduce the availability of diverse germplasm, thus limiting the potential for cassava improvement.

Several studies have been conducted to identify the genes responsible for CMD resistance and two main sources of CMD resistance have been identified. The first originates from the Nigerian landraces designated the TMS series which were generated as part of a breeding effort at the IITA in Nigeria (Hahn *et al.* 1980). CMD resistance in the TMS series originates from *M. esculenta* x *M. glaziovii* hybrids generated at Amani in Tanzania (Rabbi *et al.* 2014b). The resistance derived from the wild relative *M. glaziovii* is believed to be polygenic and has proved challenging to incorporate into elite germplasm (Hahn *et al.* 1980; Okogbenin *et al.* 2007). The second source of CMD resistance was discovered in Nigerian landraces, designated the TME series (Akano *et al.* 2002). These landraces have been shown to be closely related and resistance is believed to be monogenic

(Rabbi *et al.* 2014b; Bredeson *et al.* 2016). The gene responsible for CMD resistance in these lines, *CMD2*, was identified by Akano *et al.* (2002) along with an associated SSR marker, SSRY28. This marker was subsequently used, with additional markers (Fregene *et al.* 2006) by Okogbenin *et al.* (2007) to introgress CMD resistance into elite Latin American cultivars.

Okogbenin *et al.* (2007) used several approaches to introgress CMD resistance into elite cultivars from CIAT. Seedlings produced by hybridising elite cultivars from CIAT to elite cultivars from IITA, with CMD resistance from *M. glaziovii*, revealed incomplete transfer of CMD resistance. This was attributed to the fact that CMD resistance from *M. glaziovii* hybrids is believed to be polygenic (Okogbenin *et al.* 2007). Further crosses were performed using CIAT elite cultivars hybridized to TME3, a Nigerian landrace with CMD resistance conferred by the single gene, *CMD2* (Akano *et al.* 2002). MAS using eight associated markers (Fregene *et al.* 2006) was then performed to select seedlings with the *CMD2* resistance gene. This approach proved to be more successful in identifying CMD resistant hybrids with approximately 67% of the selected genotypes scoring either 1 (no symptoms) or 2 (mildly symptomatic but with green, healthy leaves) on the disease severity index (SI), indicating resistance to CMD. This study illustrates the importance of molecular markers in understanding the genetics underlying key traits and their value in cassava improvement. However, disease resistance from a single gene may not be able to withstand evolutionary changes in the pathogen, leaving cassava vulnerable to infections by new strains of the virus. Therefore, in order to ensure robust, durable disease resistance, efforts are underway to identify additional sources of disease resistance (Okogbenin *et al.* 2012; do Carmo *et al.* 2015).

There have been numerous reports of QTL discovery and subsequent MAB successes in many crop plants and animal breeding programmes resulting in improvement in key traits. In cassava, over the past 20 years, at least 12 different genetic linkage mapping studies have been reported (Table 1) and numerous QTL studies have been successfully conducted to identify trait-linked markers for

important traits such as plant architecture, productivity and yield, and disease resistance. However, the application of these studies has been extremely limited. Only one study (Okogbenin *et al.* 2007) has reported the application of these markers for MAB for CMD resistance, with moderate success, in a region of the world not yet affected by CMD. In spite of all the research efforts aimed at developing genetic marker resources and identifying QTLs, very little impact has been made through MAB on cassava improvement, suggesting that alternative approaches are required to make significant gains in cassava breeding programmes.

MAB has been shown to be well suited to introgressing monogenic traits such as resistance conferred by the *CMD2* gene (Ribaut and Hoisington 1998). However, alternative approaches such as genome-wide association studies (GWAS) and genomic selection (GS) may be more successfully applied to understand marker-trait associations and to select superior performing seedlings for more complex, polygenic traits. Moreover, GS can be used to improve several traits simultaneously without prior information of marker-trait associations making it ideal for cassava improvement, which has not benefitted extensively from molecular breeding approaches in the past and where limited large effect QTLs have been identified and verified in alternate genetic backgrounds (de Oliveira *et al.* 2012; Ceballos *et al.* 2015). The recent advances in genomic resources for cassava, and the reduction in genotyping costs makes this an increasingly attractive alternative to MAS for cassava improvement.

Molecular markers can also be applied to identify potentially heterotic groups and track varieties (Rabbi *et al.* 2015). Cassava is predominantly vegetatively propagated from stem cuttings and frequent transfer of material occurs between farmers from different regions (Elias *et al.* 2001). Additionally, within Africa, where the threat of CMD is high, CMD resistant varieties have been widely distributed throughout many countries (Nweke 2009). A recent study by Rabbi *et al.* (2015) used GBS to track released varieties and landraces in Ghana and reported that farmers

predominantly planted several varieties in their fields which were allowed to hybridise with other varieties and landraces, resulting in admixture. Furthermore, varieties are often given local names with up to 33 different names being recorded for a single variety, making tracking of varieties resulting from different breeding programmes challenging without the use of molecular markers. The frequent transfer of material between regions and admixture has also resulted in moderate to poor substructure within African cassava germplasm (Kawuki *et al.* 2013; Rabbi *et al.* 2015). However, molecular markers can be applied to identify heterotic groupings and genetically distant individuals to generate new crosses in order to exploit the effects of heterosis, minimise inbreeding depression and develop improved varieties (Ferguson *et al.* 2012a).

1.4.4 Genetic diversity in cassava

Cassava was domesticated from the wild *Manihot* species, *M. esculenta* sub spp *flabellifolia*, within the last 10,000 years (Olsen and Schaal 1999; Olsen and Schaal 2001). Previous studies have shown greater genetic diversity in the wild relatives of cassava than in modern cassava cultivars indicating a recent population bottleneck (Elias *et al.* 2004). These wild relatives and the various cassava germplasm collections around the world may prove to be valuable sources of genetic diversity for breeding programmes (Duputie *et al.* 2007; Akinbo *et al.* 2014). In order to exploit this genetic material, molecular markers are required to assess the diversity within germplasm collections and wild relatives, enabling the introduction of additional variation and unique alleles into breeding programmes (Fregene *et al.* 2000b; Glaszmann *et al.* 2010). This is a key application of molecular markers since the genetic improvement of cassava has been restricted to some degree by lack of knowledge regarding the genetic diversity within these germplasm collections worldwide (Asante and Offei 2003).

Brazil, which has over 40 wild *Manihot* species, is the main centre of diversity for cassava, due to its close proximity to the point of origin along the Amazon basin (Olsen and Schaal 1999). The

introduction of cassava to many different countries from Brazil could have resulted in founder effects and a certain degree of relatedness among cassava populations in different regions (Fregene *et al.* 2003). Genetic diversity in Africa may be further reduced by CMD pressure, and subsequent breeding efforts which have been centred around two main sources of resistance (Okogbenin *et al.* 2007). Furthermore, traditional breeding approaches and the widespread planting of disease resistant cultivars, in addition to the vegetative propagation of cassava, may contribute to decreased genetic variation in modern varieties (Lokko *et al.* 2006; Nweke 2009; van de Wouw *et al.* 2010). Any reduction in diversity has the potential to diminish the crop's ability to withstand environmental stresses such as new strains of diseases and climate change, therefore it is important to retain genetic diversity in order to combat these stresses and to provide genetic resources to breed for future threats (van de Wouw *et al.* 2010).

Various diversity studies have been conducted on cassava germplasm from different collections held worldwide. In several of these studies, high levels of diversity were recorded (Zacarias *et al.* 2004; Lokko *et al.* 2006; Siqueira *et al.* 2010), despite the many factors that may contribute to a reduction in diversity. A study by Fregene *et al*. (2003) to assess diversity in germplasm from African and Latin America observed high levels of genetic diversity in all countries, specifically in the germplasm from Brazil and Colombia, as expected. A later study by Hurtado *et al*. (2008) observed high levels of diversity within germplasm from Tanzania, suggesting that it is a valuable source of diversity in Africa. These high levels of diversity may be attributed to the fact that cassava is an outcrossing species and despite being primarily vegetatively propagated, farmers often select and plant seedlings to increase diversity and safeguard against crop loss from abiotic and biotic stresses (Fregene *et al.* 2003). This results in many different landraces being present in each propagated field (Fregene *et al.* 2003) ensuring the maintenance of genetic diversity. A study by Raji *et al.* (2009) conducted using a set of elite African cultivars and landraces observed higher diversity within landraces than within cultivars, suggesting that landraces may be a valuable source of genetic variation because they are subjected to natural selection and outcrossing and are often

exchanged among farmers from different regions (Siqueira *et al.* 2010). Therefore, traditional farming practices, the use of landraces, and the transfer of plant material between regions may be valuable in maintaining and generating diversity.

More recent studies have shown a reduction in diversity in Africa and evidence of a population bottleneck. A study by Kawuki *et al*. (2013) was conducted using 1401 germplasm accessions from southern, eastern and central (SEC) Africa consisting of improved cultivars and landraces from seven countries. This study, which made use of 26 SSR markers, reported low to moderate genetic diversity and evidence of a population bottleneck in cultivars and landraces in SEC Africa. Similar results were reported by Ferguson *et al*. (2012b), where germplasm from Africa and Latin America were genotyped using a large set of 1190 SNP markers. Lower levels of diversity were observed in African accessions when compared to Latin American germplasm showing evidence of a population bottleneck. Furthermore, Kizito *et al*. (2005) reported a loss of rare alleles in areas with high disease pressure and crop loss due to CMD. Founder effects resulting from the introduction of cassava to Africa, crop loss due to diseases such as CMD and CBSD and subsequent breeding efforts, and the widespread planting of resistant genotypes may all contribute to the reduction of diversity observed in African germplasm. The reduced diversity in African germplasm provides an opportunity for improvement through breeding with diverse germplasm and the introduction of novel alleles. Genetic diversity held in germplasm collections in Latin America may be valuable for expanding the genetic base of cassava in Africa, provided susceptibility to CMD and CBSD can be overcome (Okogbenin *et al.* 2007).

Diversity within a species can be assessed using germplasm collections which ideally represent the genetic resources of the crop. Studies to assess and quantify diversity within breeding programmes and landraces in different regions are key steps towards driving conservation of genetic diversity present in different regions. Kawuki *et al*. (2013) noted that diverse germplasm in SEC Africa has

not been well represented in germplasm collections. Since this area is subjected to high disease pressure, specifically from CMD and CBSD, it is vital that germplasm from SEC Africa is conserved in case it is lost due to a new disease strain. This highlights the need for more comprehensive studies to assess the genetic diversity in cassava and the effects of breeding to direct conservation strategies. Furthermore, these collections provide a means to identify unique variants in cultivars that are adapted to specific environments. Various factors, such as those mentioned previously, limit quantification and conservation of diversity in cassava germplasm collections which in turn limit the value of the collections in breeding programmes (Glaszmann *et al.* 2010).

Evidence of divergence between germplasm from Latin America and Africa has been reported in several studies (Fregene *et al.* 2000a; Fregene *et al.* 2003; Kawuki *et al.* 2009; Ferguson *et al.* 2012b). African cassava germplasm originates from multiple introductions to different regions, associated with the slave trade (Jones 1959), and has since been subjected to natural and artificial selection, disease pressure and local environmental stresses (Elias *et al.* 2001; Legg *et al.* 2011). These factors all contribute to the genetic differentiation of African cassava germplasm, however, within Africa, poor to moderate substructure has been observed. Studies by Kizito *et al.* (2005) and Lokko *et al.* (2006), using 35 and 18 SSR markers, respectively, reported little or weak differentiation between West and East African germplasm. However, a recent study by Ferguson *et al.* (2012b), using 1190 SNP markers, observed moderate differentiation between genetic material from West Africa and SEC Africa. This may also be, in part, due to the greater resolution provided by the larger number of informative markers used in this study.

The genetic structure within African germplasm may be affected by several factors. Firstly, the frequent transfer of genetic material among farmers has ensured constant gene flow between regions (Kawuki *et al.* 2013). Secondly, due to the severity of the threat of CMD in Africa, CMD resistant material resulting from the two main sources mentioned previously, has been widely planted in

Africa (Nweke 2009). Furthermore, one source of CMD resistance, the Nigerian landraces designated the TME series, have been shown to be genetically similar (Fregene *et al.* 2000a; Rabbi *et al.* 2014b; Bredeson *et al.* 2016). Additionally, many improved cultivars which have been distributed throughout Africa were developed at the IITA in Nigeria from germplasm originating from West Africa and Latin America (Nweke 2009). As a result, the effects of local adaptation and subsequent genetic differentiation has been diminished by cultivation practices and by breeding programmes aimed at the development and distribution of improved cultivars to secure the crop against the threat of CMD.

1.5 Conclusions

Cassava genetic advancement through the translation of molecular resources into breeding applications has long fallen behind that of other important crops. In recent years, however, significant advances have been made toward the development of cassava genomic resources, such as the recent efforts in resequencing the cassava genome (Bredeson *et al.* 2016), the implementation of GBS technologies (Rabbi *et al.* 2014a; Rabbi *et al.* 2014b; Rabbi *et al.* 2015), and the discovery of large sets of DNA markers (Ferguson *et al.* 2012b; Sakurai *et al.* 2013). These resources pave the way for further dissection of traits such as disease resistance, drought-tolerance, growth and yield, through QTL mapping and GWAS, towards MAB and other molecular breeding approaches such as GS. However, approaches such as these require large populations often planted over several sites to gain sufficient statistical power. This still remains problematic in cassava breeding, due to low flowering rates, particularly in elite cultivars, and is currently a major limiting factor for the development and uptake of molecular breeding strategies in cassava (Ceballos *et al.* 2004).

The availability of the draft genome assembly has created the opportunity for GBS, thus circumventing the need for larger sets of SNP markers such as those available for maize, wheat and rice, to detect marker-trait associations. However, GBS requires substantial bioinformatics resources and expertise, and the resulting set of markers may not be consistent across study populations. This may pose a problem for applications such as GS where models rely on the same set of markers being genotyped in training and study populations. There is, therefore, still a need for fixed genotyping arrays consisting of tens of thousands of genome-wide markers for cassava breeding.

Germplasm collections with diverse genetic material provide a valuable resource for crop improvement, it is therefore important to characterise the diversity in germplasm collections using DNA markers, both for current breeding purposes and to conserve the genetic diversity as a resource for the future. Factors such as breeding, disease prevalence and propagation practices may affect diversity and population substructure, and a clear understanding of these effects is essential. The recent studies by Rabbi *et al.* (2015) and Bredeson *et al.* (2016) demonstrate the value of genome-wide sets of DNA markers to map out the footprint of various breeding programmes and to fully understand the genetic makeup of propagated cassava in each region. However, a more comprehensive picture is required to effectively guide breeding decisions, identify heterotic groupings to exploit in breeding programmes, and direct conservation efforts.

While low to moderate genetic differentiation has been reported in germplasm from various countries in Africa (Kizito *et al.* 2005; Lokko *et al.* 2006), genome-wide scans may highlight regions of the genome that have been subjected to selection either through breeding or exposure to factors such as high disease pressure, or adaptation to local environments, or admixture as a result of hybridization to wild relatives. The recent study by Bredeson *et al.* (2016) illustrates the value of chromosomal-scale studies to reveal the lasting effects of breeding programmes and the

identification of unexpected interspecific hybridization and introgression. Furthermore, studies such as these, using thousands of genome-wide markers, provide the opportunity to estimate genomic relatedness values between individuals, rather than depending on recorded pedigree information where available, for the purpose of improving molecular breeding strategies.

Genotyping of whole germplasm collections is important to properly establish and manage these collections by ensuring that the available genetic diversity is properly captured, reducing redundancy and ensuring the correct labelling of accessions. This is particularly vital in the case of disease resistant accessions that may be deployed in areas of high disease pressure, or incorporated into breeding programmes. Large sets of DNA markers may be costly for large-scale genotyping of accessions in germplasm collections. While markers such as SSRs are available for routine genotyping, they are not suitable for comparison between laboratories and are time-consuming to implement. A small set of informative SNP markers is therefore ideal for low-cost, routine genotyping applications and can be used on a variety of platforms allowing comparisons between laboratories.

The aim of this study is to add to the body of knowledge regarding genetic diversity held in germplasm collections in Africa and Latin America and to develop DNA marker-based tools for the cassava community. A SNP-based genetic linkage map was constructed using a drought-tolerant mapping population, which can be used in the future for QTL mapping of this important trait, as phenotypic data is made available. The same set of SNP markers was used to characterize genetic diversity in 215 cassava accessions from breeding programmes and germplasm collections in Africa and Latin America. This data was combined with the mapped position of each marker, based on alignment to the physical map of cassava using the draft genome assembly (*Manihot esculenta* v6.1), to assess genome-wide patterns of diversity within three genetic groupings of germplasm identified in this study. Using the combined data, a subset of 128 highly informative SNP markers

was selected for routine genotyping applications such as pedigree reconstruction, identity verification and the identification of heterotic groupings and diverse germplasm for breeding purposes.

1.6 Tables

Table 1.1 Summary of genetic linkage maps published for cassava.

Reference	Population Type and Cross	Population size	Markers	Linkage Groups	Map Length	Mean Interval
Fregene *et al*. 1997	F_1 hybrid (TMS30572 x CM2177-2)	90	**Female** 132 RFLP 30 RAPD 3 SSR 3 Isoenzymes **Male** 107 RFLP 50 RAPD 1 SSR 1 Isoenzyme	20 24	931.6 cM 1120 cM	7.9 cM -
Okogbenin *et al*. 2006	F_2 Inbred (Selfed F_1 from TMS30572 x CM 2177-2)	268	100 SSR	22	1236.7 cM	17.92 cM
Kunkeaw *et al*. 2010	F_1 hybrid (Rayong 90 x Rayong 5)	58	119 AFLP 18 SSR	33	1095 cM	7.99 cM
Chen *et al.* 2010	F_1 hybrid (SC6 x Mianbao)	210	231 AFLP 41 SSR 48 SRAP 35 EST-SSR	18	1707.9 cM	4.81 cM
Kunkeaw *et al.* 2011	F_1 hybrid (Huay Bong 60 x Hanatee)	100	56 EST-SSR 155 SSR	20	1178 cM	5.6 cM
Sraphet *et al.* 2011	F_1 hybrid (Huay Bong 60 x Hanatee)	100	510 SSR	23	1420.3 cM	4.54 cM
Whankaew *et al.* 2011	F_1 hybrid (Huay Bong 60 x Hanatee)	100	303 SSR	27	1328 cM	5.8 cM
Rabbi *et al.* 2012	F_1 hybrid (Namikonga x Albert)	130	434 SNP 134 SSR	19	1837 cM	3.4 cM

Table 1.1 (cont.)

Reference	Population Type and Cross	Population size	Markers	Linkage Groups	Map Length	Mean Interval
Rabbi *et al.* 2014a	F_1 hybrid (IITA TMS IBA961089A x TMEB117)	182	772 SNP	19	1505 cM	2.0 cM
Rabbi *et al.* 2014b	F1 hybrid (IITA-TMS-011412 x IITA-TMS-4(2)1425)	180	6756 SNP	19	3515 cM	0.52 cM
Soto *et al.* 2015	F1 hybrid (TMS30572 x CM 2177-2)	132	2141 SNP	18	2571 cM	1.26 cM
International Cassava Genetic Map Consortium (ICGMC) 2015	F_1 hybrid (AR40-6 x Albert)	129	22403 SNP	18	2412 cM	0.51 cM
	F_1 hybrid (Kiroba x AR37-80)	132				
	F_1 hybrid (TMS-IBA30001 x TMS-IBA961089A)	177				
	F_1 hybrid (TMS-IBA961089A x TMS-IBA30001)	162				
	F_1 hybrid (Mkombozi x Unknown)	135				
	F_1 hybrid (Nachinyaya x AR37-80)	233				
	F_1 hybrid (NDL06/132 x AR37-80)	244				
	F_1 hybrid (Namikonga x Albert)	256				
	S_1 cross (TMEB419 x TMEB419)	117				
	F_1 hybrid (TMS-IBA4(2)1425 x TMS-IBA011412)	155				

1.7 References

Akano AO, Dixon AGO, Mba C, Barrera E and Fregene M **(2002)**. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theoretical and Applied Genetics **105**: 521.

Akinbo O, Labuschagne M and Fregene M **(2014)**. Introgression of root protein and yield traits from backcross hybrids between cassava and its wild progenitor (*Manihot esculenta* ssp *flabellifolia*). Journal of Plant Breeding and Crop Science **7**: 1.

Alves AAC **(2002)**. Cassava botany and physiology. Cassava: Biology, Production and Utilization. Hillocks RJ, Thresh JM and Bellotti AC, CAB International: 67.

Anderson JV, Delseny M, Fregene MA, Jorge V, Mba C, Lopez C, Restrepo S, Soto M, Piegu B, Verdier V, Cooke R, Tohme J and Horvath DP **(2004)**. An EST resource for cassava and other species of *Euphorbiaceae*. Plant Molecular Biology **56**: 527.

Asante IK and Offei SK **(2003)**. RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. Euphytica **131**: 113.

Asare P, Galyuon I, Sarfo J and Tetteh J **(2013)**. Morphological and molecular based diversity studies of some cassava (*Manihot esculenta* Crantz) germplasm in Ghana. African Journal of Biotechnology **10**(63): 13900.

Ayling S, Ferguson M, Rounsley S and Kulakow P **(2012)**. Information resources for cassava research and breeding. Tropical Plant Biology **5**(1): 140.

Balagopalan C, Padmaja G, Nanda SK and Moorthy SN **(1988)**. Cassava in food, feed and industry. United States, CRC Press.

Boonchanawiwat A, Sraphet S, Boonseng O, Lightfoot DA and Triwitayakorn K **(2011)**. QTL underlying plant and first branch height in cassava (*Manihot esculenta* Crantz). Field Crops Research **121**: 343.

Bouchez A, Hospital F, Causse M, Gallais A and Charcosset A **(2002)**. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics **162**(4): 1945.

Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P, Lebot V, Ndunguru J, Mkamilo G, Bart RS, Setter TL, Gleadow RM, Kulakow P, Ferguson ME, Rounsley S and Rokhsar DS **(2016)**. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nature Biotechnology **34**: 562.

Calvert LA and Thresh JM **(2002)**. Viruses and virus diseases of cassava. Cassava: Biology, Production and Utilization. Hillocks RJ, Thresh JM and Bellotti AC, CAB International**:** 237.

Ceballos H, Iglesias CA, Pérez JC and Dixon AGO **(2004)**. Cassava breeding: opportunities and challenges. Plant Molecular Biology **56**(4): 503.

Ceballos H, Kulakow P and Hershey C **(2012)**. Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. Tropical Plant Biology **5**: 73.

Ceballos H, Kawuki R, Gracen V, Yencho G and Hershey C **(2015)**. Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics **128**: 1647.

Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE and Korban SS **(2008)**. Development of a set of SNP markers present in expressed genes of the apple. Genomics **92**: 353.

Chen X, Xia Z, Fu Y, Lu C and Wang W **(2010)**. Constructing a genetic linkage map using an F_1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). Plant Molecular Biology Reporter **28**: 676.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OSH, Tingey S, Morgante M and Rafalski AJ **(2002)**. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics **3**(19): 1.

Cortés DF, Reilly K, Okogbenin E, Beeching JR, Iglesias C and Tohme J **(2002)**. Mapping wound-response genes involved in post-harvest physiological deterioration (PPD) of cassava (*Manihot esculenta* Crantz). Euphytica **128**(1): 47.

de Bang T, Raji A and Ingelbrecht I **(2011)**. A multiplex microsatellite marker kit for diversity assessment of large cassava (*Manihot esculenta* Crantz) germplasm collections. Plant Molecular Biology Reporter **29**: 655.

de Oliveira EJ, de Resende MDV, da Silva Santos V, Ferreira CF, Oliveira GAF, da Silva MS, de Oliveira LA and Aguilar-Vildoso CI **(2012)**. Genome-wide selection in cassava. Euphytica **187**(2): 263.

do Carmo CD, da Silva MS, Oliveira GAF and de Oliveira EJ **(2015)**. Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz. Scientia Agricola **72**: 520.

Duputie A, David P, Debain C and McKey D **(2007)**. Natural hybridization between a clonally propagated crop, cassava (*Manihot esculenta* Crantz) and a wild relative in French Guiana. Molecular Ecology **16**(14): 3025.

Elias M, Penet L, Vindry P, McKey D, Panaud O and Robert T **(2001)**. Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. Molecular Ecology **10**: 1895.

Elias M, Mühlen GS, McKey D, Roa AC and Tohme J **(2004)**. Genetic diversity of traditional South American landraces of cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. Economic Botany **58**(2): 242.

Food and agriculture organisation of the United Nations **(2014)**. From http://www.fao.org/faostat/en/#compare.

Ferguson M, Rabbi I, Kim D-J, Gedil M, Lopez-Lavalle L and Okogbenin E **(2012a)**. Molecular markers and their application to cassava breeding: past, present and future. Tropical Plant Biology **5**: 95.

Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, Town CD, de Young J, Marri PR, Rabbi IY and de Villiers EP **(2012b)**. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theoretical and Applied Genetics **124**(4): 685.

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). Theoretical and Applied Genetics **95**: 431.

Fregene M, Bernal A, Duque M, Dixon A and J.Tohme **(2000a)**. AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theoretical and Applied Genetics **100**: 678.

Fregene M, Roca W, Okogbenin E, Akano A, Mba C, Chavarriagga C, Tohme J, Ospina B and Ceballos H **(2000b)**. Cassava biotechnology research at CIAT/Columbia. Cassava's potential in Asia in the 21st Century: Present situation and future research and development needs: Proceedings of the sixth Regional workshop, Ho Chi Minh-Vietnam, Centro Internacional de Agricultura Tropical (CIAT).

Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R and Kresovich S **(2003)**. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theoretical and Applied Genetics **107**: 1083.

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.

Generation Challenge Programme. From (http://generationcp.org).

Glaszmann J, Kilian B, Upadhyaya H and Varshney R **(2010)**. Accessing genetic diversity for crop improvement. Current Opinion in Plant Biology **13**: 167.

Gupta PK, Roy JK and Prasad M **(2001)**. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science **80**(4): 524.

Hahn S, Terry E and Leuschner K **(1980)**. Breeding cassava for resistance to cassava mosaic disease. Euphytica **29**: 673.

Halsey ME, Olsen KM, Taylor NJ and Chavarriaga-Aguirre P **(2008)**. Reproductive biology of cassava (*Manihot esculenta* Crantz) and isolation of experimental field trials. Crop Science **48**: 49.

Hillocks RJ and Jennings DL **(2003)**. Cassava brown streak disease: A review of present knowledge and research needs. Internation Journal of Pest Management **49**(3): 225.

Hittalmani S, Parco A, Mew VT, Zeigler SR and Huang N **(2000)**. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theoretical and Applied Genetics **100**(7): 1121.

Hurtado P, Olsen KM, Buitrago C, Ospina C, Marin J, Duque M, de Vicente C, Wongtiem P, Wenzel P, Killian A, Adeleke M and Fregene M **(2008)**. Comparison of simple sequence repeat (SSR) and Diversity Array Technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). Plant Genetic Resources: Characterization and Utilization **6**(3): 208.

International Cassava Genetic Map Consortium **(2015)**. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. G3: Genes Genomes Genetics **5**(1): 133.

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): E25.

Jansson C, Westerbergh A, Zhang J, Hud X and Sun C **(2009)**. Cassava, a potential biofuel crop in (the) People's Republic of China. Applied Energy **86**: S96.

Jarvis A, Ramirez-Villegas J, Campo BVH and Navarro-Racines C **(2012)**. Is cassava the answer to African climate change adaptation? Tropical Plant Biology **5**(1): 9.

Jefferies SP, King BJ, Barr AR, Warner P, Logue SJ and Langridge P **(2003)**. Marker-assisted backcross introgression of the *Yd2* gene conferring resistance to barley yellow dwarf virus in barley. Plant Breeding **122**(1): 52.

Jones WO **(1959)**. Manioc in Africa. Stanford, Standard University Press.

Jorge V, Fregene M, Vélez CM, Duque MC, Tohme J and Verdier V **(2001)**. QTL analysis of field resistance to *Xanthomonas axonopodis* pv. *manihotis* in cassava. Theoretical and Applied Genetics **102**: 564.

Kahl G, Mast A, Tooke N, Shen R and van den Boom D **(2005)**. Single nucleotide polymorphisms: detection techniques and their potential for genotyping and genome mapping. The Handbook of Plant Genome Mapping: Genetic and Physical Mapping. Meksem K and Kahl G. Weinheim, Wiley-VCH.

Kawuki RS, Ferguson M, Labuschagne M, Herselman L and Kim D-J **(2009)**. Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). Molecular Breeding **23**: 669.

Kawuki RS, Herselman L, Labuschagne MT, Nzuki I, Ralimanana I, Bidiaka M, Kanyange MC, Gashaka G, Masumba E, Mkamilo G, Gethi J, Wanjala B, Zacarias A, Madabula F and Ferguson ME **(2013)**. Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa. Plant Genetic Resources: Characterization and Utilization **11**(2): 170.

Kizito EB, Bua A, Fregene M, Egwang T, Gullberg U and Westerbergh A **(2005)**. The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica **146**: 45.

Kunkeaw S, Tangphatsornruang S, Smith DR and Triwitayakorn K **(2010)**. Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. Plant Breeding **129**: 112.

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

Legg J and Fauquet C **(2004)**. Cassava mosaic geminiviruses in Africa. Plant Molecular Biology **56**: 585.

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

Li Y, Pan Y, Sun C, Dong H, Luo X, Wang Z, Tang J and Chen B **(2010)**. An ordered EST catalogue and gene expression profiles of cassava (*Manihot esculenta*) at key growth stages. Plant Molecular Biology **74**(6): 573.

Liu L, Jeffers D, Zhang Y, Ding M, Chen W, Kang MS and Fan X **(2015)**. Introgression of the *crtRB1* gene into quality protein maize inbred lines using molecular markers. Molecular Breeding **35**(8): 154.

Liu N, Chen L, Wang S, Cheongeun O and Zhao H **(2005)**. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genetics **6**: S26.

Lokko Y, Danquah EY, Offei SK, Dixon AGO and Gedil MA **(2005)**. Molecular markers associated with a new source of resistance to the cassava mosaic disease. African Journal of Biotechnology **4**(9): 873.

Lokko Y, Dixon A, Offei S, Danquah E and Fregene M **(2006)**. Assessment of genetic diversity among African cassava *Manihot esculenta* Crantz accessions resistant to the cassava mosaic virus disease using SSR markers. Genetic Resources and Crop Evaluation **53**: 1441.

Lokko Y, Anderson J, Rudd S, Raji A, Horvath D, Mikel M, Kim R, Liu L, Hernandez A, Dixon A and Ingelbrecht I **(2007)**. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. Plant Cell Reports **26**: 1605.

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

Lopez CE, Quesada-Ocampo LM, Bohórquez A, Duque MC, Vargas J, Tohme J and Verdier V **(2007)**. Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. Genome **50**: 1078.

Mba REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J and Fregene M **(2001)**. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics **102**: 21.

Mohan C, Shanmugasundaram P, Maheswaran M, Senthil N, Raghu D and Unnikrishnan M **(2013)**. Mapping new genetic markers associated with CMD resistance in cassava (*Manihot esculenta* Crantz) using simple sequence repeat markers. Journal of Agricultural Science **5**(5): 57.

Moyib OK, Odunola OA and Dixon AGO **(2007)**. SSR markers reveal genetic variation between improved cassava cultivars and landraces within a collection of Nigerian cassava germplasm. African Journal of Biotechnology **6**(23): 2666.'

Nassar NM, Hashimoto DY and Fernandes SD **(2008)**. Wild *Manihot* species: botanical aspects, geographic distribution and economic value. Genetics and Molecular Research **7**(1): 16.

Next Generation Cassava Breeding Project (NextGen Cassava). From http://www.nextgencassava.org/.

Nweke F **(2009)**. Controlling cassava mosaic virus and cassava mealybug in sub-Saharan Africa. 2020 Vision Initiative. Washington DC, IFPRI.

Okogbenin E and Fregene M **(2002)**. Genetic analysis and QTL mapping of early root bulking in an F_1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics **106**: 58.

Okogbenin E and Fregene M **(2003)**. Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics **107**: 1452.

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

Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG, Ospina C, Marín J, Barrera E, Gutiérrez J, Ekanayake I, Iglesias C and Fregene MA **(2007)**. Marker-assisted introgression of resistance into cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. Crop Science **47**: 1895.

Okogbenin E, Marin J and Fregene M **(2008)**. QTL analysis for early yield in a pseudo F_2 population of cassava. African Journal of Biotechnology **7**(2): 131.

Okogbenin E, Egesi C, Olasanmi B, Ogundapo O, Kahya S, Hurtado P, Marin J, Akinbo O, Mba C and Gomez H **(2012)**. Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. Crop Science **52**(6): 2576.

Olsen KM and Schaal BA **(1999)**. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Science of the United States of America **96**: 5586.

Olsen K and Schaal B **(2001)**. Microsatellite variation in cassava *Manihot esculenta*, *Euphorbiaceae* and its wild relatives: further evidence for a southern Amazonian origin of domestication. American Journal of Botany **88**: 131.

Onwueme IC **(1978)**. The tropical tuber crops. New York, John Wiley & Sons.

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS and Rounsley S **(2012)**. The cassava genome: current progress, future directions. Tropical Plant Biology **5**: 88.

Rabbi IY, Kulembeka HP, Masumba E, Marri PR and Ferguson M **(2012)**. An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics **125**: 329.

Rabbi I, Hamblin M, Gedil M, Kulakow P, Ferguson M and Ikpan A **(2014a)**. Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. Crop Science **54**(4): 1384.

Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink J-L and Kulakow PA **(2014b)**. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Research **186**(0): 87.

Rabbi I, Kulakow P, Manu-Aduening J, Dankyi A, Asibuo J, Parkes E, Abdoulaye T, Girma G, Gedil M, Ramu P, Reyes B and Maredia M **(2015)**. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). BMC Genetics **16**(1): 115.

Rafalski A **(2002)**. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology **5**: 94.

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

Reif J, Hallauer A and Melchinger A **(2005)**. Heterosis and heterotic patterns in maize. Maydica **50**: 215.

Ribaut J-M and Hoisington D **(1998)**. Marker-assisted selection: new tools and strategies. Trends in Plant Science **3**(6): 236.

Sakurai T, Mochida K, Yoshida T, Akiyama K, Ishitani M and Seki M **(2013)**. Genome-wide discovery and information resource development of DNA polymorphisms in cassava. PloS One **8**(9): e74056.

Schneider K, Weisshaar B, Borchardt DC and Salamini F **(2001)**. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Molecular Breeding **8**: 63.

Silvestre P **(1989)**. Cassava. London, MacMillan Publishers Ltd.

Simon M, Loudet O, Durand S, Berard A, Brunel D, Sennesal FX, Durand-Tardif M, Pelletier G and Camilleri C **(2008)**. Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. Genetics **178**(4): 2253.

Siqueira MVBM, Pinheiro TT, Borges A, Valle TL, Zatarim M and Veasey EA **(2010)**. Microsatellite polymorphisms in cassava landraces from the Cerrado Biome, Mato Grosso do Sul, Brazil. Biochemical Genetics **48**: 879.

Soto J, Ortiz J, Perlaza-Jimenez L, Vasquez A, Lopez-Lavalle LA, Mathew B, Leon J, Bernal A, Ballvora A and Lopez C **(2015)**. A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. BMC Genomics **16**(1): 190.

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

Taylor NJ, Halsey M, Gaitán-Solís E, Anderson P, Gichuki S, Miano D, Bua A, Alicai T and Fauquet CM **(2012)**. The VIRCA Project: virus resistant cassava for Africa. GM Crops and Food: Biotechnology in Agriculture and the Food Chain **3**: 93.

Thanyasiriwat T, Sraphet S, Whankaew S, Boonseng O, Bao J, Lightfoot D, Tangphatsornruang S and Triwitayakorn K **(2014)**. Quantitative trait loci and candidate genes associated with starch pasting viscosity characteristics in cassava (*Manihot esculenta* Crantz). Plant Biology **16**(1): 197.

Utsumi Y, Sakurai T, Umemura Y, Ayling S, Ishitani M, Narangajavana J, Sojikul P, Triwitayakorn K, Matsui M and Manabe R-i **(2012)**. RIKEN cassava initiative: establishment of a cassava functional genomics platform. Tropical Plant Biology **5**(1): 110.

van de Wouw M, van Hintum T, Kik C, van Treuren R and Visser B **(2010)**. Genetic diversity trends in twentieth century crop cultivars: a meta analysis. Theoretical and Applied Genetics **120**: 1241.

van Marle-Koster E, Visser C and Berry D **(2013)**. A review of genomic selection - Implications for the South African beef and dairy cattle industries. South African Journal of Animal Science **43**(1): 1.

Varma A and Malathi VG **(2003)**. Emerging geminivirus problems: A serious threat to crop production. Annals of Applied Biology **142**: 145.

Varshney RK, Beier U, Khlestkina EK, Kota R, Korzun V, Graner A and Börner A **(2007)**. Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies. Theoretical and Applied Genetics **115**: 1105.

Varshney RK, Paulo MJ, Grando S, van Eeuwijk FA, Keizer LCP, Guo P, Ceccarelli S, Kilian A, Baum M and Graner A **(2012)**. Genome wide association analysis for drought tolerance related traits in barley (*Hordeum vulgare* L.). Field Crops Research **126**: 171.

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

Win KM, Korinsak S, Sirithunya P, Lanceras-Siangliw J, Jamboonsri W, Da T, Patarapuwadol S and Toojinda T **(2013)**. Marker assisted introgression of multiple genes for bacterial blight resistance into aromatic Myanmar rice MK-75. Field Crops Research **154**: 164.

Xia L, Peng K, Yang S, Wenzl P, de Vincente M, Fregene M and Kilian A **(2005)**. DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics **110**: 1092.

Zacarias AM, Botha A-M, Labuschagne MT and Benesi IRM **(2004)**. Characterization and genetic distance analysis of cassava (*Manihot esculenta* Crantz) germplasm from Mozambique using RAPD fingerprinting. Euphytica **138**: 49.

Zhou A and Thomson E **(2009)**. The development of biofuels in Asia. Applied Energy **86**: S11.

Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK and Smith NJ **(2003)**. Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. Plant Breeding **122**(1): 40.