# GENETIC ANALYSIS OF

# PLANT MORPHOLOGY IN BAMBARA GROUNDNUT

## [*Vigna subterranea*  (L.) Verdc.]

**Nariman Salih Ahmad**

BSc Agronomy, MSc Plant breeding

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

October 2012

School of Biosciences

## Abstract

Bambara groundnut (*Vigna subterranea*  (L.) Verdc.) is an important underutilised legume crop, grown mainly by female subsistence farmers in Africa under traditional low input agricultural systems. Bambara groundnut is known as being of high nutritional value, as an atmospheric nitrogen fixer and to possess high levels of drought, pest and disease tolerance. Bambara groundnut is a predominantly self-pollinated crop and is grown as locally adapted landraces. These are expected to exist as non-identical inbred lines and are generally low yielding. Strategies involving genetic analysis of this species could provide important data for breeding programmes that could enhance food security in Africa.

A set of 124 SSR primers designed from different library sources were tested to screen  a 'narrow' genetic cross ($F_3$) and a 'wide' genetic cross ($F_2$) . The former is a cross between domesticated landraces (DipC and Tiga necaru) while the latter is a cross between a domesticated landrace and a wild ancestor (DipC and VSSP11).

Residual heterozygosity in the $F_3$ 'narrow' cross was confirmed to be around 25% based on 33 polymorphic SSR primers, consistent with an $F_3$ population. A 'narrow' cross linkage map was constructed for the first time in bambara groundnut using 269 polymorphic markers (236 DArT and 33 SSR). The map consisted of 238 markers in 21 linkage groups of two or more linked markers, totalling 608.1cM and covering a predicted 54% of the bambara groundnut genome, although the high marker-marker linkage (at 89%) suggests a more comprehensive coverage. QTL analysis was carried out for 73 bulked lines of an $F_3$ population and plants were evaluated for traits in a controlled glasshouse suite and a field trial in Indonesia. Data from single plant analysis of the $F_2$ generation of this cross grown in a controlled environment

glasshouse was also used. Most of the QTLs detected were clustered on linkage groups 1, 4 and 12. Major QTLs for *internode length* and *biomass dry weight* were detected on LG4 and LG1, respectively, for the FutureCrop glasshouse and field datasets. The highest LOD score of 9.7 was detected for *peduncle length* and was located within the confidence interval for a QTL for *internode length* locus. Marker locus bgPabg-596774 was detected to be associated with QTL for six traits; *node no./plant*, *pod no/plant, pod weight, seed no./plant, seed yield and biomass dry weight,* on LG1 within one LOD score of confidential interval, potentially suggesting pleiotropic effects of a more limited number (or even one) gene(s).

One hundred and fifty-nine additional markers (136 DArT and 23 SSR) were used to improve the existing partial 'wide' map (141 AFLP, 1 SSR) constructed in an $F_2$ population of 98 plants. A total of 194 markers were assigned to 20 linkage groups spanning a total of 901 cM. The linkage map derived from the 'wide' cross (DipC x VSSP11) had an expected genome coverage of 79.6%. An attempt to combine both maps through 32 common markers allowed a common QTL for *days to emergence* to be detected in both populations in close association with the common DArT markers 601384 and 601748.

The main segregating traits were found to be *plant spread*, *internode length*, *growth habit*, *peduncle length*, *pod weight, seed yield* and *biomass dry weight*. Detecting the same QTL positions for a number of traits, suggested that common underlying genes might be responsible. The QTL-DNA marker associations developed in this study could be used practically for MAS in a future breeding program of this crop.

## Publication

Ahmad, N *[1]; S. Basu[1]; E. Sri Redjeki[1], E. Murchie[1], S. Azam-Ali[2], A. Kilian[3] and S. Mayes[1] (2012). Developing genetic mapping and marker-assisted breeding techniques in Bambara groundnut (*Vigna subterranea* (L) Verdc.). 2nd International Symposium on Underutilised Plant Species. ISHS, Acta Horticulturae (In press).

[1]Plant and Crop Sciences, School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK.

[2]University of Malaysia Nottingham Campus, Jalan Broga,43500 Semenyih, Selangor Darul Ehsan, Malaysia.

[3]DArT Pty Ltd., 1 Wilf Crane Cr., Yarralumla, ACT 2600, Australia

*Corresponding author: Sbxna3@nottingham.ac.uk

# Acknowledgements

# Table of contents

## List of Tables

## List of Figures

## List of Appendices

# List of abbreviations

μL      Microlitre (0.000001 Litres)

μM     Micromole, equal to $10^{-6}$ mole

ABI     Applied bio system

AFLP Amplified Fragment Length Polymorphism

BC      Back cross

BDW   Biomass dry weight

bp       Base pair

CAPS   Cleaved amplified polymorphic sequences

CFFRC Crops for the Future Research Centre

CID     Carbon isotope discrimination

CIM    Composite interval mapping

CISR    Conserved intron spanning region

cM      CentiMorgan

COS    Conserved orthologous sequences

CP      Cross Pollinator

CRD    Complete randomised design

DArT   Diversity Arrays Technology

DE      Days to emergence

DF      Days to flowering

DNA   Deoxyribonucleic Acid

dNTP   Deoxynucleotide Triphosphates (usually mix of dATP/dTTP/dCTP/dGTP)

DPL    Pod length of double seeded

DPN    double seeded pods/plant

DPW    Pod width of double seeded

EDTA   Ethylenediaminetetraacetic acid

$E_i$    Expected frequency

EP     Eye pattern around hilum

EST-SSR     Expressed Sequence Tag-SRR

FN     Flower no./plant

FTh    Flower bud thrip

GFU    Global Facilitation unit for Underutilized species

GH     Growth habit

GMM    Genic molecular marker

HSW    100-seed weight

ICUC   International Centre for Underutilized Crops

IITA   International Institute of Tropical Agriculture

IL     Internode length

IM     Interval mapping

ISSR   Inter-simple sequence repeat

KW     Kruskal-Wallis analysis

LA     Leaf area

LG     Linkage group

LOD    Logarithm of odds

LPA    Linear polyacrylamide

MAS Marker-assisted selection

MIM Multiple interval mapping

MISA Microsatellite identification tool

mM Millimole, equal to $10^{-3}$ mole

ng Nanogram (One billionth ($10^{-9}$) of a gram)

NN Node no./stem

$O_i$ Observed frequency

PCR Polymerase chain reaction

PEL Peduncle length

PLE Pod length

pmol Picomole, equal to $10^{-12}$ moles

PN Pod no./plant

PS Plant spread

PT Permutation test

PVE Phenotypic variation explained

PWD Pod width

PWE Pod weight

QTL Quantitative trait loci

RAPD Random Amplification of Polymorphic DNA

RFLP Restriction fragment length polymorphism

RIL Recombinant inbreed line

SDW Sterile distilled water

SEL Seed length

SEN    Seed no./plant

SEW    Seed width

SH%    Shelling percentage

SLA    Specific leaf area

SLS     Sample loading solutions

SNP    Single Nucleotide Polymorphism

SMA     Single marker analysis

SS     Size standard

SSR    Simple sequence repeat

STN    Stem no./plant

STR    Short tandem repeat

SWT    Seed weight

TBE    Tris/Borate/EDTA

TCRU Tropical Crops Research Unit

TLL    Terminal leaflet length

TLW    Terminal leaflet width

tRNA   Transfer ribonucleic acid

UV     Ultraviolet

# Chapter 1.  INTRODUCTION

## 1.1 Legumes: an overview

The legume family is one of the largest families of the angiosperms. Plants in the leguminosae family (alternative name, Fabaceae) form three identified sub-families; Papilionoideae, Caesalpinioideae and Mimosoideae. The first sub-family, with approximately 70% of the Leguminosae species, is the largest subfamily and includes the important edible legume crops and major model legume species (Cannon *et al.*, 2009; Doyle and Luckow, 2003). The Fabaceae family is the third largest amongst flowering plants, with 18,000 species and 650 genera (Polhill *et al.*, 1981).

The wide distribution of leguminosae throughout the world and their high protein content gives them great potential as food sources. Their further cultivation and increased utilization would positively contribute toward improving food security. In Europe  consumption of legume seeds is 2.5 kg/capita per annum while in other parts of the world higher consumption rates have been recorded (5.4 – 14.4 kg/capita) (Schuster-Gajzágó, 2009). Focusing on accessible plant protein sources like legumes is advisable in developing countries to improve the nutritional status of the low-income groups and restrict malnutrition (Iqbal *et al.*, 2006). Although, legume crops rank second in importance to cereals, legumes are up to 2-3 times richer in protein than cereal grains. Thus they are an important component of a balanced diet and are a cheap source of protein in many African countries where animal protein is costly and beyond the purchasing power of the low income groups in these areas (National Academy of Sciences, 1979).

The low nutrient and degraded status of many soils is a particular problem for small landholders, as many cannot afford to use artificial fertilizers. Legume yields are less dependent upon an external supply of nitrogen fertilizer (although most do respond to additional nitrogen) and the ability of most of them to grow in poor soils has favored their cultivation (Sandal *et al.*, 2002). They are an important part of subsistence agriculture as in addition to providing a protein-rich food they also improve the soil structure (Sato *et al.*, 2010). Faba beans are among the best legume crops to increase and restore organic matter in the soil, in developing countries and arid zones (Solh and Saxena, 2011).

Proteins are major components of legume seeds as seeds can contain 200-250g protein/kg. Legume seeds are rich in lysine which cereals are deficient in. While they contain a lower percentage of methionine and cysteine compared to the cereals, this combination can be balanced by integrating grain legumes with cereals in the diet (Schuster-Gajzágó, 2009). In an investigation of nutritional value of some alternative crops in organic agriculture, grain legumes were reported as a source rich in protein, carbohydrates and minerals (Table 1-1). The grain seed of faba bean, chickpea and lentil contains around 21g protein/100g, which is higher than the content of pseudo cereals. The highest energy value was registered for faba bean, with 396.58 kcal/100g seed (Toader *et al.*, 2011). These pseudo cereals themselves were found to be in higher contents for carbohydrate, protein and oil compared to wheat (Souci *et al.*, 2000).

More than 30 species of grain legumes are grown across the tropics, in sub-Saharan Africa and South Asia and they are an important component of sustainable agriculture (Abate *et al.*, 2012).

**Table 1-1: Nutritional values of alternative crops (Moara Domneasca Experimental Field, Bucharest; (Toader *et al.*, 2011)**

| Alternative crops | Species | Carbohydrate (g/100g) | Proteins (g/100g) | Lipids (g/100g) | Minerals (g/100g) | Energy value (kcal/100g) |
|---|---|---|---|---|---|---|
| Pseudo cereals | *Amaranthus* spp. | 66.87 | 16.47 | 4.91 | 2.61 | 389.97 |
| | Quinoa | 64.32 | 16.71 | 5.80 | 2.89 | 389.06 |
| | Buckwheat | 65.50 | 16.03 | 3.53 | 2.31 | 351.05 |
| Grain legumes | Faba bean | 63.90 | 21.50 | 4.40 | 5.85 | 396.58 |
| | Chickpea | 56.20 | 21.23 | 4.31 | 3.41 | 360.95 |
| | Lentils | 33.29 | 22.18 | 3.03 | 4.00 | 259.60 |
| Oil crops | Safflower | 26.41 | 12.60 | 28.38 | 3.60 | 426.73 |
| | Camelina | 36.27 | 20.43 | 31.75 | 4.28 | 532.02 |
| | Oil flax | 27.73 | 22.56 | 34.10 | 5.25 | 528.56 |

The recent millennium report of ecosystem assessment indicates that desertification threatens over 41% of the world's land area, mostly in the dry areas of Middle East, North Africa and sub-Saharan Africa. Water scarcity and available water quality are potentially serious threats to food security and health in dry areas countries of developing world which characterized by marginal production with scarce resources and very high population growth rates of 2.5% (Erskine, 2003; Solh and Saxena, 2011).

## 1.2 Bambara groundnut

Bambara groundnut (*Vigna subterranea* (L.) Verdc.) (Figure 1-1) is an indigenous, underutilized African legume crop from semi-arid Africa, grown mainly by subsistence farmers for food. It has potential to assist with providing food security in the dry areas of Africa and also in other parts of the world. Bambara groundnut gives reasonable yields with little rainfall and is a favored food crop for many local people, being nutritionally comparable to other legumes, such as soybean, in the essential amino acids of lysine, methionine and cysteine (Brough and Azam-Ali, 1992). It is also grown more broadly, having shown the ability to adapt to a broad range of

ecological conditions and can be found in South East Asia. In Indonesia it is a minor crop, while the average rainfall can exceed 2300 mm per year across Java (Aldrian and Djamil, 2008).



**Figure 1-1: Bambara groundnut [*Vigna subterranea* (L.) Verdc.]. (a) a botanical sketch, (b) flower, (c) fruits, (d) seed, (Maesen and Somaatmadja, 1989), (e) freshly harvested plant.**

A phylogenetic tree is presented here (Figure 1-2) showing that bambara groundnut belongs to the Phaseoloid-Millettioid clade, which diverged some 45–50 million years ago from the Hologalegina clade. The Phaseoloid-Millettioid clade contains most legume crops, such as pea, alfalfa (*Medicago spp.*), chickpea, soybean and common bean (Choi *et al.*, 2004b).
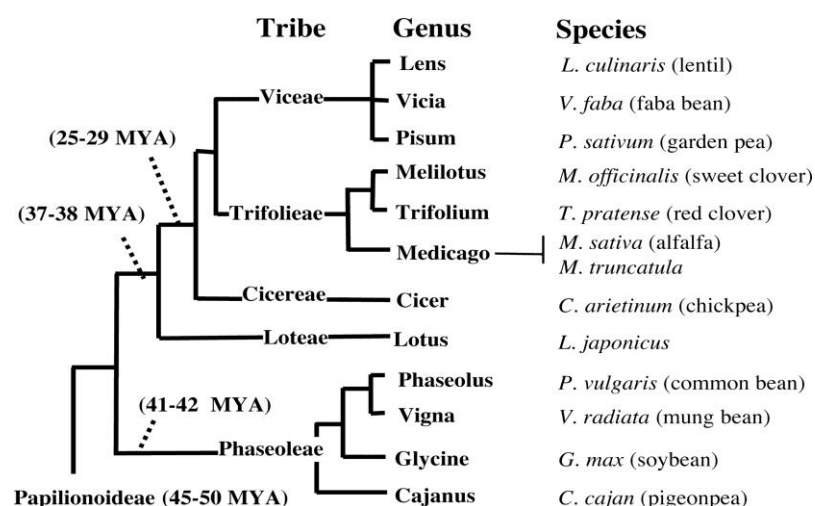
**Figure 1-2: Taxonomic relationships between crop legumes. Penalised Likelihood analysis (Doyle and Luckow, 2003) was used to estimate the species divergence. Most crop legumes occur either within the Galegoid clade, including tribes Viceae, Trifolieae, Cicereae and Loteae, or within the phaseoloid clade (tribe Phaseoleae). MYA, million years ago (Choi *et al.*, 2004b).**

## 1.3 The importance of Bambara groundnut

Bambara groundnut ranks as the third most important food legume crop in semi-arid Africa in terms of production and consumption after groundnut (*Arachis hypogaea* L.) and cowpea (*Vigna unguiculata* L. Walp) (Aremu *et al.*, 2006; Doku and Karikari, 1970; Howell *et al.*, 1994; Linnemann, 1992; Sellschop, 1962) because of the shortage of sufficient rainfall for major legume cultivation. Drought tolerance is the most important trait that this crop possessed as it can be grown and gives acceptable yields in  marginal areas where the cultivation of other legumes is not practicable (Vietmeyer, 1979). In addition it has undergone a prolonged period of exposure to pest and disease pressure under low input agriculture (Azam-Ali *et al.*, 2001). It is a rich source of protein and along with other local sources of protein could help to alleviate nutritional problems in areas where staple foods are predominantly carbohydrate sources (Massawe *et al.*, 2005; Okpuzor *et al.*, 2010). It has been concluded that bambara groundnut seed is a useful ingredient for different food

5

products of beverages, infants, children milk food and different drinks (Eltayeb *et al.*, 2011). Bambara groundnut seed makes a 'complete food', as on average the seed contains sufficient protein (19%), carbohydrate (63%) and fat (6.5%) for a nutritionally balanced diet (Amarteifio and Moholo, 1998; Brough and Azam-Ali, 1992; Ijarotimi and Esho, 2009). Mineral content was also estimated for 100g seed, giving; iron 59 mg, potassium 1240 mg, phosphorus 296 mg, sodium 3.7mg and calcium 78 mg (Amarteifio and Moholo, 1998). In addition it has high protein quality with a good balance of essential amino acids, compared to most of other grain legumes, with relatively high lysine (6.8%) and methionine (1.3%) (Brough *et al.*, 1993; Ellah and Singh, 2008; Okpuzor *et al.*, 2010) which are often only available at low levels in legumes. In a cream testa bambara groundnut a methionine content of 2.84% (of total crude protein) was reported (Olaleke *et al.*, 2006). Also the seed pods and other by-products have been recommended for feeding all kinds of livestock. Some anecdotal medicinal uses of bambara groundnut seed and leaves mixed with other crops in North Eastern Nigeria have also been reported (Atiku, 2000; Directorate plant production, 2011). Symbiosis of bambara groundnut with *Rhizobium* bacteria to fix atmospheric $N_2$ enhances the value of this crop for crop rotation and intercropping, as it contributes to the supply of soil nitrogen for other crops (Karikari *et al.*, 1999). Additionally, naturally-occurring $NO_3^-$ ion tolerant symbioses in bambara groundnut have been identified. These compare well to tolerance of artificial nitrate in other legumes, where there is a strong inhibitory effect on symbiosis. This potentially allows Nitrogen fertilization in intercropping systems without inhibiting $N_2$ fixation in the associated legumes (Dakora, 1998).

The annual world production of bambara groundnut is estimated to be around 330,000 tons, with West Africa (Nigeria, Niger, Burkina Faso, Chad, Cote d'Ivoire) producing

nearly half of this amount, and the demand for bambara currently is reported to be greater than the quantities produced (Brink and Belay, 2006; Directorate plant production, 2011).

Despite its importance as a part of the diet of much of sub-Saharan Africa, bambara groundnut has not as yet received significant research interest. There are no established varieties of bambara groundnut and the crop is still cultivated mainly from local landraces, rather than as genotype-specific varieties (Massawe *et al.*, 2005). Bambara groundnut is characterized by higher genetic diversity in the wild ancestor (Pasquet *et al.*, 1999) compared to domesticated landraces. Many landraces also consist of multiple genotypes, the mixes of which potentially increase the tolerance of the landrace to biotic and abiotic stresses and make them better able to survive and adapt to marginal areas under harsh conditions (Massawe *et al.*, 2005). Both wild material and genotypes within landraces could represent an important potential source of beneficial genes for bambara groundnut breeding programs. Landraces could have the advantage of yield stability through genetic and phenotypic heterogeneity, but this is also likely to be one reason for limited yield compared to pure line crops. To improve this crop through breeding and enable the distribution of this crop to currently non-producing areas, detailed genotypic information and modern processing methods for domesticated and wild landraces are required. The ability of the breeder to exploit molecular technology to generate information at the gene level using different kinds of molecular markers (based on mapping and marker-trait analysis to locate important qualitative and quantitative trait loci (QTLs)) offers the opportunity to gain a better understanding of bambara groundnut genetics. This could potentially contribute to increased farmer income, better food security and also improve the quality of local diets.

**1.4 Literature review**

**1.4.1  Underutilized crops**

Underutilized crop species are often highly adapted to marginal environments and are inherited as part of the local culture and production systems. Many such crops are considered to be adapted to uncertain environments and niches which cannot be filled by other major crop species (Bhag, 2007). They have often been selected both directly and indirectly to resist harsh conditions under low agricultural input, both in terms of nutrients and prophylactic chemicals.   Plant scientists have emphasized the availability of these species because of their significant adaptability to less favored areas and harsh environments (IPGRI, 2002; Padulosi and Hoeschle-Zeledon, 2004; Padulosi *et al.*, 2002). Underutilized plant species are those grown primarily in their centre of origin or centers of diversity that have been maintained by cultural preferences and traditional practices. They play an important role in food security, nutrition, and income generation for local subsistence farmers, as well as often having important cultural functions (International Atomic Energy Agency, 2004). Some underutilized crops were once widely grown but are today neglected by research and conservation for a variety of agronomic, genetic, economic and cultural reasons and the understanding of their utilization is restricted often to only indigenous knowledge (Magbagbeola *et al.*, 2010; Padulosi *et al.*, 2002). Using nutritionally-rich underutilized species which are adapted to low input agriculture systems could have immediate consequences for food security and the income of local people (IPGRI, 2002; Naylor *et al.*, 2004; Oniang'o *et al.*, 2006). These facts suggest that the conservation of, and more research on, these underutilized species is required by current and subsequent generations (Dansi *et al.*, 2012).

## 1.4.2  Important considerations for underutilized crops

Humans have vigorously practiced plant and animal breeding for more than 12,000 years by selection within the crops they grow and animals they rear. The development of crop agriculture, in particular, has encouraged the development of permanent settlements and has been key in the development of civilization (Moschini, 2008). They have selected and developed those crop and breeds with traits of interest appropriate to the regions where they were first domesticated and have subsequently altered these species by artificial selection as their range has changed geographically and climatically, through the process of selective adaptation. Current climate change, urban expansion, deforestation, and other human activities have accelerated genetic erosion and are increasing the rates at which species are becoming threatened and vanishing. While the development of high yielding varieties in the  Green Revolution allowed the rapidly increasing populations to be fed, but they have arguably also increased the rate that local crops are displaced from many farmland areas, especially in developing countries (Thies, 2000). At present, human beings rely on a limited number of food and non-food crops to meet the requirements of their staple diets and associated needs. Although, in the past more than 7,000 species of plants have been cultivated for human consumption,  95% of human food energy currently comes from only 30 crops, four of which (rice, wheat, maize and potato) provide  more than 60% of our energy  intake (FAO, 2011b). So a small number of species occupy the majority of the agricultural land area for cultivation.

Although some underutilized species are globally distributed, they remain inadequately characterized and neglected by research and conservation (Eyzaguirre *et al.*, 1999). They are barely represented in *ex situ* gene banks and have no supply system for their germplasm. The plant genetic resources conserved today for food and

agriculture exceeds 6 million accessions. However, underutilized crops have poor representation at around 20% of the total collection of crop germplasm with  on average only eight accessions per species  (Padulosi *et al.*, 2002) (Figure 1-3)
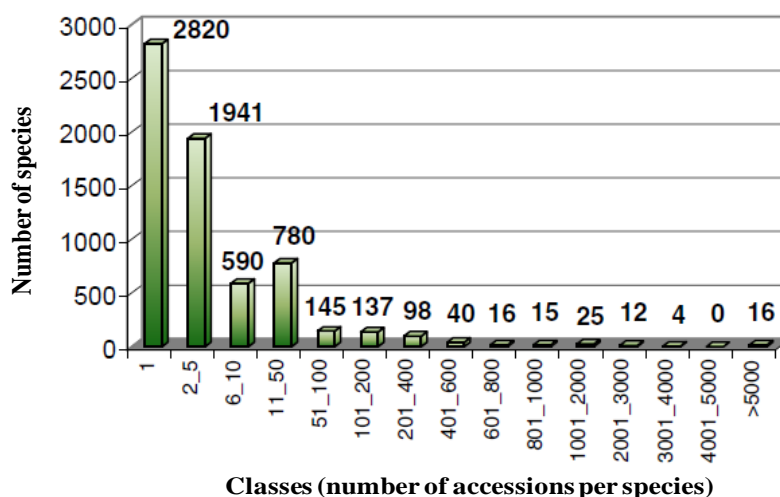


**Figure 1-3: Minor crop representation in *ex situ* gene banks (Padulosi *et al.*, 2002).**

Much of the underutilized species germplasm is preserved by farmers in their field through their use of them, which makes loss of germplasm more of a risk than for species where there are extensive genebanks (IPGRI, 1998).

Food production to sustain a steadily increasing population is becoming one of the most important issues facing Humankind, with world population expected to exceed 9 billion by 2050 (Conforti, 2011).  Global agriculture production has to grow substantially, by between 70% and by 100% in developing countries over the next four decades to feed this population (FAO, 2011a). Developing more sustainable practices to produce the food required for this number of people is a big challenge, especially with the potential impact of climate change and the increased demand for water resources for human use on agricultural production. It has been reported that climate change has decreased agricultural production in low-latitude regions, where

most developing countries are located (Easterling *et al.*, 2007). In addition, population growth rates are highest in these areas, especially in the regions of the greatest poverty and food shortages, making an existing problem potentially far worse (Davies, 2003).

Research has focused primarily on the current major crops in recent decades, as they have dominated agriculture and human consumption. Among the commercialized major crops for human consumption three crops (maize, wheat, and rice) supply half of the world's daily protein and calories (Bharucha and Pretty, 2010). The displacement of indigenous and underutilized crops by introduced major staples has also lead to a serious erosion of germplasm for many local landraces (Brookfield, 2002). Thus, relying only on a small number of crops is potentially dangerous given the circumstances of increasing population pressure and climate change (Taylor *et al.*, 2009). Any attempt to improve global food security has to include staple crops, as our reliance on them is so great at the moment. However, we also have no choice but to focus on increased sustainability of crop production in the future and one means of realizing this may be through growing crops which require fewer inputs or are more resistance to stress. (Mayes *et al.*, 2011). Growing indigenous crops could help resource poor farmers obtain a satisfactory (and more stable) income. In many cases, women in these areas are often responsible for growing additional crops, while the main staple crops are grown by men. Although, germplasm has been selected by local farmers in marginal areas, food shortages have often not been alleviated, because their farming practice is limited by the resources available and often associated depleted, degraded or simply poor soils. From this argument it is obvious that a focus on breeding for neglected crop germplasm, especially in more hostile environments, by focusing on minor crops which are resistant to drought, pests and disease and which

are of high nutritional value is required to help to tackle the increasing demands for food due to population pressure (Holmens, 1998).

It has been proposed that improving the genetics of existing underutilized crop varieties, exploring post-harvest technology and better management, particularly of soils, in the regions of Africa and Latin America could bring underutilized species back into cultivation to fill some of the yield gaps and enhance food security. As they are generally more adapted to the extreme soil and climatic conditions predominantly found in Africa than the major crops and they are extremely important source for food production in low income, food deficit countries (Foley *et al.*, 2011; Mohan and Suprasanna, 2011).

### 1.4.3 Strategic approaches for underutilized species

Many successful projects for underutilized crops have focused on a specific and important single end-use, however many underutilized crops are multipurpose and their end-uses vary from place to place. For example, the legume *Lathyrus* is used largely for fodder in Turkey, and as a food legume in South Asia, while it is used as animal feed and as a green manure in Australia (Williams and Haq, 2002). Identifying a specific and important single end-use of a product which could substitute for a major crop product from minor crops such as *Triticale*, *Amaranthus*, buckwheat and sesame has proven successful in development of strategic approaches to increase uptake of these crops. This process was applied through decades of research and the development of substantial germplasm collections, followed by selection, breeding and multilocational trials and commercialization (Ashri, 1994; Williams and Brenner, 1995).

Some organizations are engaged in strategic frameworks for research and development activities to enhance the use of underutilized species, aiming to improve the living condition of people worldwide. The International Centre for Underutilized Crops (ICUC) was established in 1992 as a global research, development and training organization. It provided expertise and acted as a knowledge center for tropical, sub-tropical and temperate plant development. ICUC has conducted research programmes in partnership with international and local research groups in over 30 countries to address increasing the use of underutilized crops for food, medicinal and industrial products, and for environmental conservation (ICUC, 2006).

The Global Facilitation Unit for Underutilized Species (GFU) was established in 2002 as a global partnership program focusing on networking and knowledge-sharing globally and to encourage incorporation of underutilized plant species into international and national research. It focused on the analysis of existing policies and legal frameworks that affect the conservation and use of underutilized species rather than being directly involved in the implementation of projects on particular plant species.

In 2008 a new international organization was formed under the name of 'Crops for the Future.' This organization has evolved from a union of the International Centre for Underutilized Crops (ICUC) and the Global Facilitation Unit for Underutilized Species (GFU). It supports, collects, synthesizes and promotes knowledge on neglected and underutilized species for the benefit of the poor and the environment (Jaenicke *et al.*, 2009). An associated research centre (CFFRC) was established recently as a unique public-private partnership between a national government and an international research-led university (http://www.nottingham.edu.my/CFFRC/index.aspx). It will operate as the research

arm of the global Crops for the Future (CFF) body hosted in Malaysia by Bioversity International and University of Nottingham Malaysia Campus (UNMC) and is a new approach of a limited company without shareholders, but guaranteed by the Government of Malaysia and the University of Nottingham in Malaysia. The Government of Malaysia has provided money for facilities and for running costs for seven years (end-2017), by which time CFFRC is expected to be self-financing, through competitive grant funding and direct interaction with industry. CFFRC has a remit to implement research and development of underutilized plants for food and non-food purposes worldwide (http://www.nottingham.ac.uk/news/pressreleases/2011/june/newcropsforthefutureres earchcentre.aspx). The strategy of CFFRC is to focus on different minor and underutilized plant species, as an attempt to promote diversification of agriculture. This will hopefully improve nutritional food security, but also provides some resilience if climate change or other factors such as disease and pests caused major crops to fail long-term in parts of the world (Mayes *et al.*, 2011). Underutilized crops could also help to provide purchasing power for poor and subsistence farmers as the recent spikes in food commodity prices show that even if food is physically available, the poor often do not have the purchasing power to obtain it (Padulosi *et al.*, 2011).

### 1.4.4   Origin and distribution of bambara groundnut

'Bambara groundnut' (botanically known as *Vigna subterranea* (L) Verdc. ;$2n=2x=22$; (Heller *et al.*, 1995) belongs to the family Leguminosae, subfamily Papilionoideae. It is related to cowpea (*Vigna unguiculata* L.) and is believed to have originated from Central Africa, before the introduction of groundnut (*Arachis hypogaea* L.) from South America (Goli, 1997). It comprises two botanical forms; the

wild sub-species (*V. subterranea* var. *spontanea*) found in Cameroon, Sudan and some parts of Nigeria, and the cultivated type (*V. subterranea* var. *subterranea*) which is found widely distributed in Sub-Saharan Africa (Basu *et al.*, 2007b). Recently, a study of genetic diversity provided evidence for Cameroon/Nigeria as the putative centre of origin for bambara groundnut, using both phenotypic descriptors and DArT markers (Olukolu *et al.*, 2012). In an investigation by Doku and Karikari (1971) it was proposed that cultivated bambara groundnut had developed gradually from var. *spontanea*. The suggested developments involved a shift in growth habit from open to bunch type, a change from outbreeding to inbreeding and a reduction in pod shell thickness. The germination of cultivated forms is rapid and uniform while in the wild forms it is erratic and germination takes longer (15-30 days) (Basu *et al.*, 2007b). Other differences were also observed between the two types; domesticated landraces were found to have larger seeds and do not wrinkle upon drying, compared to the wild type (Basu *et al.*, 2007b; Pasquet, 2003). It has been demonstrated that wild bambara groundnut (*spontanea*) is likely to be the true progenitor of domesticated bambara groundnut (*subterranea*) using both morphological and isozyme data (Pasquet *et al.*, 1999).

It is believed that bambara groundnut was brought first to East Africa and Madagascar, then later it was introduced by slaves at the time of the Slave Trade and distributed to South and South East Asia (Hanelt, 2001). Somta *et al.*, (2011b) using molecular markers proposed that Thai bambara groundnut landraces originate from both West Africa (Nigeria) and East Africa, implying that it was introduced to Thailand more than once. Bambara groundnut is widely cultivated throughout tropical Africa, India, Sri Lanka, Indonesia and Malaysia (Goli, 1997). It is also reported to be

cultivated in other parts of the world, including South and Central America and parts of northern Australia (Linnemann and Azam-Ali, 1993; Suwanprasert *et al.*, 2006).

Evaluation and collection of bambara groundnut germplasm was carried out initially in the 19th century (Anonymous, 1947). The International Institute of Tropical Agriculture (IITA) in Nigeria has carried out numerous collection expeditions in 26 countries around the world and now there are 2030 germplasm accessions in their collections (Stadler, 2009).

### 1.4.5 Morphological and physiological characteristics of bambara groundnut

### 1.4.5.1 Morphological characteristic

Bambara groundnut is a herbaceous, intermediate, annual plant and believed to be mainly self-pollinating (Heller *et al.*, 1997). The morphological structure of the crop largely matches that of the groundnut (*Arachis hypogaea*), in that the pale yellow flower stalk bends downwards after fertilization bearing its pods below the ground (Uguru and Ezeh, 1997).

It has two main contrasting growth habits; the branched form and the bunched habit, with a reproductive cycle of usually 90 to 150 days, depending on environment and landraces (Berchie *et al.*, 2010; Goli, 1997). The tap root is well developed with many profuse geotropic lateral roots of around 20 cm long on the lower part (Akpalu, 2010). Nodules formed on the roots fix atmospheric nitrogen through symbiosis with *Rhizobium* bacteria, which makes them useful for crop rotation and intercropping (Karikari *et al.*, 1999; Linnemann and Azam-Ali, 1993).

Bambara groundnut is believed to be autogamous and floral reproduction starts 30 to 35 days after sowing and may continue until the end of the plant's life (Directorate

Plant Production, 2009; Swanevelder, 1998). Flowers are normally carried in pairs on short peduncles by a pedicle which arises from the axis formed by the petioles and the stem (Doku, 1968). Flowers produced on the same peduncle do not open synchronously, although they will open within a 24 hours interval. Delayed flower opening may be caused by low temperatures and cloudy skies (Massawe *et al.*, 2003). It has been reported that fertilization in bambara groundnut takes place on the same day as anthesis (Linnemann and Craufurd, 1994). After fertilization, the flower stem elongates. During this time, the peduncle elongates to bring the ovaries to the soil level and the pedicels penetrate the soil surface after fertilization to form the pods (Heller *et al.*, 1995). The sepal enlarges and the fruit develops above or just below the soil surface. Pod development lasts up to 30 days after fertilization and the seed develops over a further 10 days (Swanevelder, 1998).

The pod is small, round or slightly oval shaped and wrinkled. Generally a single seed is produced in the pod, although two seeds per pod have been reported (Pasquet and Fotso, 1997). Seeds are mature when the parenchymatous layers surrounding the embryo have disappeared and the pods become a light brown (Toungos *et al.*, 2009). The seeds are round, smooth and very hard when dried, with highly variable testa colors, including cream, brown, red and blotched (Stephens, 2003) (Figure 1-4).
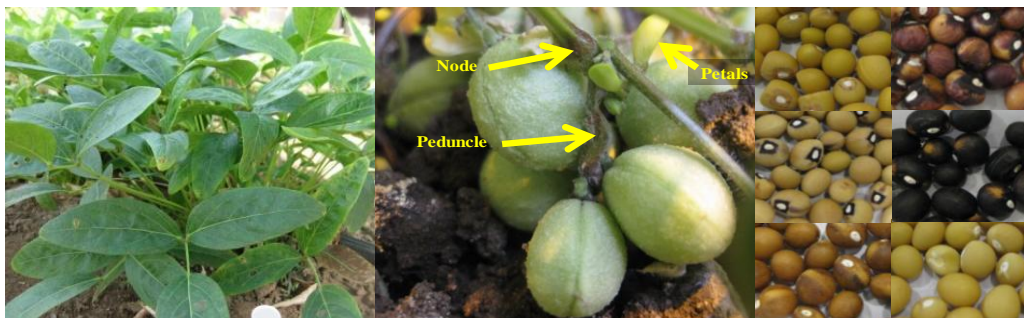


**Figure 1-4: Bambara groundnut; plant, flowers, pods and seeds with different testa colors.**

**1.4.5.2 Physiological characteristic**

The growth and development of bambara groundnut is affected by the major abiotic factors of moisture, temperature and photoperiod. Bambara groundnut requires warm temperatures with an average day temperature of 20 to 28°C being ideal for the crop. Germination in bambara groundnut usually takes seven to 15 days under favorable temperature conditions (28.5$^{o}$C-32.5°C) for bambara groundnut; while under lower temperatures, the germination and seedling emergence can be delayed up to 31 days with some seeds remaining dormant indefinitely (Linnemann and Azam-Ali, 1993; Swanevelder, 1998). Development in bambara groundnut appears to be indeterminate with flowers often small and yellow color arising alternatively with leaves at each node (Dimakatso, 2006).

A number of physiological studies have been undertaken in bambara groundnut. They confirmed that different landraces have different responses to photoperiod, sowing date (Sesay *et al.*, 2008), moisture deficit (Collinson *et al.*, 1997; Mwale *et al.*, 2007), growth rate (Massawe *et al.*, 2003) and drought tolerance. These recorded variations in the response of bambara groundnut provides a good background to improve the performance and develop varieties for bambara groundnut in future breeding programs.

*Photoperiod*

The life cycle of the bambara groundnut crop varies from 3-6 months, depending on the landrace and climatic conditions. Photoperiod affects the days to maturity and it has been found that the podding is retarded by long photoperiods in some landraces (Harris and Azam-Ali, 1993) and different landraces have different photoperiod requirements for pod production. In the photothermal response of the onset of flowering and the onset of podding for three Nigerian landraces it was found that the

flowering of Ankpa 2 and Yola was controlled by temperature only, while Ankpa 4 was sensitive to temperature and photoperiod. Whereas the onset of podding was affected by temperature and photoperiod in all three landraces and the most photoperiod-sensitive was Ankpa 4, followed by Yola and Ankpa 2. (Linnemann and Craufurd, 1994). In another study it was also determined that photoperiod influences the flowering and podding in Ankpa 4, with no pods being produced under 14 and 16 hour photoperiods, while in Tiga necaru only the onset of podding is affected by photoperiod (Linnemann *et al.*, 1995). A greenhouse experiment was conducted under 16-17 hours photoperiod and $20^{o}C$ - $25^{o}C$ temperature to identify photoperiod neutral lines in bambara groundnut. Three landraces (TZA-1498, TZA-1505 and TZA-2114) collected from Tanzania close to equator have been reported as photoperiod-neutral lines having clear pod setting whereas all other landraces remained in the vegetative stage (Jorgensen *et al.*, 2009). It has been generally reported that floral initiation is not photoperiod dependent but this is now being questioned, although the current data suggests that flowering is more likely affected by temperature, while the reproductive development and yield generation in bambara groundnut is affected by photoperiod after flowering (P. Kendabie, personal communication).

### *Drought tolerance*

With the potential risk of drought associated with climate change, drought tolerance is likely to become even more important in African agriculture. Bambara groundnut as a drought tolerant plant has the potential as a crop to provide improved food security in the dry areas of Africa (Berchie *et al.*, 2012). The plant is most suited for hot dry areas, is highly adaptable and tolerates harsh conditions (Karunaratne *et al.*, 2011). The strong root system with a compact tap root has been suggested to enhance the resistance of this plant to drought (Begemann, 1986). Bambara groundnut is

apparently able to maintain turgor through a combination of osmotic adjustment, reduction in leaf area index and effective stomatal regulation of water loss (Collinson *et al.*, 1997). Although the mechanisms that allow bambara groundnut to still produce some yield during severe droughts are poorly understood, it has been stated that this mechanism may be linked with its relatively high root/shoot biomass ratio and small leaf area, which restricts transpirational loss of water (Collinson *et al.*, 1996).

Drought response was identified to be different among bambara groundnut landraces. Uniswa Red from Swaziland defined as a drought avoiding landrace, a 'water-saver' according to the three adaptation terms defined by Turner (1979). It had the most 'optimistic' growth with the largest leaf area reduction and the earliest stomata closure which is important for survival during intermittent drought. While S19-3 from the low rainfall area of Namibia is defined as a 'water-spender' with a late closure of stomata, with a more 'conservative' growth pattern, with compressed phenology combined with reduction in leaf area and stomata closure. The drought responses of S19-3 are fit well with the short cropping season of Namibia with annual terminal drought problems (Jorgensen *et al.*, 2010).

Berchie *et al.* (2012) evaluated the performance of five bambara groundnut landraces (Black eye, Burkina, NAV 4, NAV Red and Tom) to drought and heat stress, at Tono-Navrongo, upper east region of Ghana. They stated the existence of variation among landraces with respect to drought tolerance as Burkina proved the most drought and heat tolerant, exhibiting the greatest root dry weight and leaf area at 120 DAS, and the longest leaf area duration. The authors believe that this tolerance to drought may be a result of adaptation to environmental conditions in the material.

### 1.4.5.3 Management and agronomy

Different plant spacings are used in bambara groundnut cultivation and the recommended spacing is broadly 10 to 15cm in rows of 45 to 90cm apart (for rows cultivation) and a spacing of $20 \times 20$cm (250,000 plants ha$^{-1}$) in flat seedbeds, with semi-bunch types giving the highest yields in Côte d'Ivoire (Directorate plant production, 2011; Kouassi and Zoro, 2010).

Bambara groundnut crops can grown in poor soils low in nutrients and requires soil of pH 5-6.5, preferring well-drained sandy loam and the average rainfall requirement of this crop ranges between 600-700mm during the growing season (Swanevelder, 1998). It can be cultivated up to 1600m above sea level (Chittaranjan, 2007).

Yields of bambara groundnut vary widely. A yield of 400-1400 kg ha$^{-1}$ unshelled pods was reported in Zimbabwe (Heller *et al.*, 1995). In Swaziland yields of 2600 kg ha$^{-1}$ was recorded in the field (Sesay *et al.*, 2008), and over 3000 kg ha$^{-1}$ have been obtained in South Africa (Swanevelder, 1998). While in Côte d'Ivoire seed yield was recorded as high as 4000 kg ha$^{-1}$ (Kouassi and Zoro, 2010) with similar yields in Ghana (J. Berchie, personal communication).

### 1.4.6 Genetic diversity of bambara groundnut

The evaluation of available genetic diversity is a pre-requisite for genetic improvement in crop plants, especially in underutilized crops such as bambara groundnut (Olukolu *et al.*, 2012). Investigation of genetic diversity in both wild and domesticated species is equally important. Wild populations are known to be a potential source of useful genes and traits which could be introduced into the domesticated gene pool; in particular, genes responsible for adaptation to stressful environments such as those providing a particular resistance to a pathogen or to arid

conditions (Cattan-Toupance *et al.*, 1998). Wild populations in centers of diversity or domestication constitute the initial gene pool of crops species. Crop failures and dispersal of germplasm within the centre of origin or limited introduction or isolated locations ('Founder Effects') could lead to reduced genetic diversity in particular breeding populations, which could have long-term negative consequences for production (Trethowan and Mujeeb-Kazi, 2008). By focusing on commercial and elite germplasm the breeder may further reduce the genetic diversity of the domesticated gene pools (Rauf *et al.*, 2010; Yi *et al.*, 2008). Studies of genetic diversity can help to guide the exploitation of wild relatives in a breeding program to retrace or enhance gene flow between wild and domesticated populations which may increase the genetic diversity in domesticated gene pools (Gepts and Papa, 2002).

Estimating the genetic diversity of crop species can be achieved using different marker methods, including; morphological, trait/agronomic, biochemical and molecular. The latter has several advantages over conventional phenotypic markers, as they can be used efficiently regardless of the developmental stage of the plant under investigation (Mondini *et al.*, 2009).

Genetic diversity was analyzed within 100 single plant accessions of bambara groundnut from a wide range of locations in Tanzania. They were assessed with 49 polymorphic bands of 11 informative AFLP primers. Cluster analysis revealed that bambara groundnut has two major groups in line with their putative geographic origins (Ntundu *et al.*, 2004). The results of this study agreed with a previous study that used RAPD markers on 25 African accessions from the collection in IITA (Ibadan and Nigeria) showing two main groups of accessions corresponding to their geographic distribution (Amadou *et al.*, 2001).

High genetic identity between wild and domesticated accessions was detected in an isozyme diversity study of bambara groundnut, Ht= 0.087 with 14 polymorphic loci and Ht= 0.052 with only 7 loci for the wild and domesticated, respectively. The study suggests that wild bambara groundnut is likely to be the true progenitor of domesticated bambara groundnut. Beside the high value of intra-population diversity in both wild and domesticated accessions, the study also suggests that self pollination is the major mode of sexual reproduction for both accession types (Pasquet *et al.*, 1999). Two hundred and forty single plant accessions of bambara groundnut were assessed using 22 SSR markers. Higher gene and allelic diversity were obtained in the West African and Cameroon/Nigeria regions than others (east African, central African, and Southeast Asian) with 6.68 and 6.18 alleles per locus, and 0.601 and 0.571, respectively (Somta *et al.*, 2011b).

An extensive and diverse range of germplasm was investigated to study genetic diversity of bambara groundnut in the study by Olukolu *et al.* (2012). Morphological and quantitative descriptors, alongside DArT markers that represent wide genome coverage, were used and a high genetic diversity was observed for the Cameroon/Nigeria region relative to other regions. This supports the hypothesis that this region is the centre of diversity for bambara groundnut (Olukolu *et al.*, 2012).

The available literature reveals a number of studies of genetic diversity in bambara groundnut in the wild and domestication material. They offer a reasonable start to understanding the genetic basis of the domestication event(s) in this crop, potentially enabling parents with a wide genetic base to be identified for developing mapping populations and subsequent QTL analysis.

## 1.5 Molecular markers

Evaluation of genetic variation within and between populations and the interaction of genes with the environment has been studied using different types of markers over the last century. The first phenotypic markers were mapped in fruit fly in 1913 (Sturtevant, 1913). In plants, in 1923 a limited number of morphological markers were tested and identified an association between seed size and seed-coat pigmentation patterns in the common bean, *Phaseolus vulgaris* (Sax, 1923). However, such phenotypic markers can often not be used as reliable indicators of the traits of interest due to an often complex interaction of the genotype and the environment. Moreover, in some species these markers are expressed at a very low frequency which makes their effective utilization in plant breeding difficult (Farooq and Azam, 2002). Marker development moved forward with the discovery and development of protein and isozyme markers which resolved some of these problems (Markert and Moller, 1959). Protein and isozyme markers can be affected by the environment and their use is constrained by the tissue and developmental stage of the species under study (Avise, 1993). In addition, they often have limited polymorphism to assess genetic diversity within a species. A different method is required to detect different isozyme systems, so a range of application techniques are required to generate data for a number of isozyme markers (Fairbanks and Andersen, 1995). In a study based on DNA markers to evaluate genetic polymorphism within cultivated tomatoes (*L. esculentum* and *L. pennellii*), none of the isozyme markers revealed polymorphism compared to 63% that detected polymorphism using RAPD markers between these varieties (Foolad *et al.*, 1993).

Breeding a new variety with conventional methods takes many years especially when there are effects of trait pleiotropism and or when there is a multifactorial basis to

morphological traits. Hence breeders are interested to try new techniques to make this process more efficient. Developments in molecular marker technology offers such a possibility by adopting a wide range of novel approaches which have altered the way plant breeding is being undertaken, allowing the breeder to use them potentially in estimating the genetic diversity and the level of heterozygosity among plants and animals (Dani *et al.*, 2008; Kumar *et al.*, 2008) as a first step to determining the best parents and best strategies for breeding.

DNA marker systems were introduced to genetic analysis in the 1980's, primarily based on the development of the Southern blot technique to assess polymorphism in complex genomes. DNA-based molecular markers have acted as versatile tools in various fields, such as taxonomy, physiology, embryology, genetic engineering (Joshi *et al.*, 1999). DNA molecular markers are used to identify a particular DNA sequence. DNA markers have become a popular means for identification of plant and animal species; they are generally stable and detectable in any part of the genome. Additionally, they are not usually affected by the age, physiological condition of the cell or environmental factors and are generally held to have no pleiotropic or epistatic effects (Mondini *et al.*, 2009).

Currently, extensive use of molecular markers derived from different technical approaches allows the segregation patterns of different alleles to be scored easily and construction of genetic maps from them. Construction of linkage maps is one of the main uses of DNA markers in research on crop species (Collard *et al.*, 2005). Such genetic maps serve several purposes, including detecting association between the genes and traits studied in QTL analysis, with the aim to use the markers to tag those traits, allowing the application of marker assisted selection of these target traits in subsequent breeding programs (Semagn *et al.*, 2006b).

DNA-based markers have been established in many agricultural crops and the availability of reliable molecular markers is of great importance for plant breeding as molecular markers linked to desirable traits have been used to accelerate plant breeding programs (Ribaut and Hoisington, 1998). The ideal molecular marker technique should generate hundreds of molecular markers that cover the entire genome in a single, simple and reliable experiment (Luikart *et al.*, 2003). DNA markers are divided based on the method of their detection into three classes, hybridization-based; polymerase chain reaction (PCR) based and DNA sequence-based (Gupta *et al.*, 1999; Jones *et al.*, 1997; Joshi *et al.*, 1999; Winter and Kahl, 1995).

DNA markers that discriminate between genotypes are termed polymorphic. Co-dominant markers reveal the allelic state of both alleles for a marker in a diploid, whereas dominant markers do not provide full information on the alleles present. Co-dominant markers may have many different alleles, whereas dominant markers only have two alleles i.e. presence or absence (Collard *et al.*, 2005; Mondini *et al.*, 2009). The polymorphic site of a genetic marker may be a short or even single base DNA sequence difference, such as in a single base-pair change (single nucleotide polymorphism, SNP) (Cotton, 1997) or could be due to greater variation, as with repeat units for minisatellites.

Various types of molecular markers have been used. Some of the more commonly used systems are; RFLPs (Restriction Fragment Length Polymorphism), AFLPs (Amplified Fragment Length Polymorphism) RAPDs (Random Amplification of Polymorphic DNA), VNTRs (Variable Number Tandem Repeat), Microsatellites (or Simple Sequence Repeat; SSR), SNPs (Single Nucleotide Polymorphism), STRs (Short Tandem Repeat), SFP (Single Feature Polymorphism), and DArT (Diversity

Arrays Technology). No single technique fulfills all research needs and it is difficult

to predict the emergence of new standard techniques (Semagn *et al.*, 2006b). Different

aspects of cost-effectiveness, accuracy, sensitivity and reproducibility in addition to

the availability of markers specific to an organism and their limitations should be

taken into account to determine the best suitable technology for a specific genotyping

purpose and approach. The most important features, details on the nature of these

markers, their use, advantages and disadvantages of different applications as adapted

from Zalapa *et al.* (2012) can be found in Table 1-2.

**Table 1-2: Main characteristics of major types of molecular markers.**

| Characteristic | RFLP | SSR | AFLP | RAPD | SNP |
|---|---|---|---|---|---|
| **Locus number detected** | Single locus | Single locus | Multi-loci | Multi-loci | Single locus |
| **Allelism** | Co-dominant | Co-dominant | Dominant | Dominant | Co-dominant |
| **Level of polymorphism** | Good | Excellent | Good | Good | Excellent |
| **Polymorphism at the locus** | 2 to 5 alleles | Multiple alleles | Presence/absence | Presence/absence | Up to 4 alleles |
| **Quantity of DNA needed** | Large | Small | Small | Small | Small |
| **Quality of DNA needed** | V. Good | reasonable | Good | reasonable | reasonable |
| **Reproducibility** | Good | Good | Good | Low | Good |
| **Time** | Long | Fast, once markers are developed | Fast | Fast | Fast, once markers are developed |
| **Cost** | Expensive | Average | Cheap | Cheap | Expensive[a] |
| **Technical difficulty** | High | Low | Medium | Medium | High[a] |

a:   Both cost and technical difficulty are highly dependent on the chosen method of visualization and, hence, on the expected throughput level

## 1.5.1   Restriction Fragment Length Polymorphisms

Restriction Fragment Length Polymorphism or RFLP was once the most widely used

molecular marker type. The technique reveals differences in restriction sites between

individual genotypes which are contained in or flank the region of the hybridization

probe. The differences are caused by evolutionary processes, spontaneous mutations or unequal crossing over (Gonzalez-Chavira *et al.*, 2006). In RFLP analysis the DNA sample is digested using restriction endonucleases which recognize specific DNA sequence motifs, generally 4 or 6 base-pair palindromic sequences. Digested DNA fragments are size fractionated according to their length by gel electrophoresis, then they are transferred to a nylon filter and fixed into place before undergoing the hybridization of a cloned sequence of DNA which has been labeled. The relative positions of the hybridizing fragments are visualized through exposure of the hybridized and washed filter to a light or X-ray sensitive surface. This gives a banding pattern which reflects the position of probe hybridization which can differ between individual genotypes due to fragment length differences due to the restriction cut sites present (Gonzalez-Chavira *et al.*, 2006). High reproducibility, co-dominant inheritance and good transferability between laboratories are considered the major strengths of RFLP markers. However their use requires high quality and quantities of DNA, and it is expensive, time consuming, often involving radioactive/toxic reagents (Mondini *et al.*, 2009). Although now largely obsolete, RFLP analysis was the first DNA profiling technique cheap enough to see widespread application. In addition to genetic fingerprinting, RFLP was an important tool in genome mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing. RFLP markers were used for the first time in 1975 to identify polymorphisms in DNA sequence for genetic mapping of adeno-virus serotypes (Grodzicker *et al.*, 1975), then they were used in human genome mapping (Botstein *et al.*, 1980), and later applied in plant genomes (Helentjaris *et al.*, 1986; Weber and Helentjaris, 1989).

### 1.5.2  Random Amplification of Polymorphic DNA

RAPD (pronounced "rapid") is a PCR-based technology in which random short segments of DNA are amplified (Kumar *et al.*, 2009). This technique was developed due to the limitations in the routine use of RFLP (Garcia *et al.*, 2004). The assay was developed first by Williams *et al.* (1990). RAPDs are DNA fragments amplified by PCR using short single synthetic oligonucleotide primers (8-12 bp) of random sequence. These primers serve as both forward and reverse primer at two different sites on complementary strands, to amplify fragments from multiple genomic sites where the product size is within the range of PCR, simultaneously. The resultant PCR fragments are visualized on agarose gels stained with ethidium bromide, and there is no need for hybridization with labeled probes as in RFLP (Kumar *et al.*, 2009). The main advantages of RAPDs are that they are quick and easy to assay, only relatively low quantities of template DNA are required, and no prior sequence information from the organism under study is required. However, they have the drawback of sensitivity to reaction conditions which can lead to their profile varying between laboratories, even for identical reagents. Additionally, as RAPD markers are not locus-specific and several loci are amplified by each primer, heterozygous and homozygous individuals cannot be distinguished (Kumar *et al.*, 2009; Paraguison *et al.*, 2012). Although RAPDs are less popular due to their limitations, they have been applied in genetic variability analysis and individual-specific genotyping. RAPD has been used to characterize, and trace, the phylogeny of various plant and animal species (Vos *et al.*, 1995; Zabeau and Vos, 1993). RAPDs have been applied in gene mapping studies to fill gaps not covered by other markers (Williams *et al.*, 1990). In chickpea RAPDs have been used for identification and mapping QTLs conferring resistance to *Aschochyta* blight (Santra *et al.*, 2000).

Using RAPDs in bambara groundnut gave high levels of polymorphism in comparison to isozyme markers among landraces investigated by Pasquet *et al.* (1999). They also have been applied in bambara groundnut to study the genetic diversity in some landraces (Amadou *et al.*, 2001; Massawe *et al.*, 2003).

### 1.5.3   Amplified fragment length polymorphism (AFLP)

Amplified Fragment Length Polymorphism PCR (AFLP-PCR) or just AFLP is a PCR-based tool used in DNA fingerprinting. This method is based on the combination of the sequence specificity of digestion of DNA through restriction endonuclease enzymes and the PCR technique. It was developed to overcome the limitation of reproducibility associated with RAPD (Mondini *et al.*, 2009). AFLP uses restriction enzymes to cut genomic DNA, followed by ligation of adaptors to the sticky ends of the restriction fragments. The primer used is extended (by addition of a random base) into the restricted fragments to achieve selective amplification of a subset of those fragments, followed by acrylamide gel analysis of amplified fragments (Kumar *et al.*, 2009; Vos *et al.*, 1995). The amplified fragments are visualized on denaturing polyacrylamide gels in combination with radioactivity (Matthes *et al.*, 1998) or on fluorescent imaging gels systems, such a LiCor or Capillary Electrophoresis systems, such as the Beckmann CEQ 8000. The AFLP technique is highly reliable and reproducible, does not require prior sequence information from the species of interest, gives high levels of polymorphism and allows multiple locus detection. The disadvantages are; it has a complicated methodology, requires both restriction endonuclease and adapter ligation reaction, so the template DNA used is required to be pure and free from inhibitor compounds which interferes with the restriction enzyme or ligation reactions (Gonzalez-Chavira *et al.*, 2006; Kumar *et al.*, 2009). The

technique is also patented, which makes using AFLPs in programme which may lead to commercial products more complex.

AFLP markers have been used to generate a genetic map in maize. 246 markers were mapped to cover 2057 cM (Ajmone Marsan *et al.*, 1999). It was realized that the generation of this marker type was faster than RFLP and markers mapped on the same population could be compared, showing that the AFLP markers show good distribution over the genome and almost all markers in this case (98.1%) followed Mendelian segregation patterns.

Application of AFLP markers was also used in the genetic characterization and mapping of some legume crops. They have been recognized as useful markers to characterize the genetic relationships between wild common bean and lima bean and to deduce the predominant direction of gene flow and the spatial differentiation between domesticated and wild bean (Papa and Gepts, 2003). In chickpea, 89.6% of AFLP bands were found to be polymorphic in 95 accessions that represented the 17 species of *Cicer* (Nguyen *et al.*, 2004). The introgression of the rhizomatous trait from *Trifolium ambiguum* to the important pasture species of white clover (*T. repens*) has been studied and bulked segregant AFLP analysis was used to identify markers linked to the rhizomatous habit (Abberton *et al.*, 2003) to enable this work.

AFLP markers were employed in bambara groundnut to assess genetic diversity and to investigate the genetic relationships among 100 selected landraces from a diverse geographic area of Tanzania, using 49 polymorphic amplification fragments scored bands from 11 markers (Ntundu *et al.*, 2004). In another investigation of genetic diversity in bambara groundnut, AFLP analysis provided sufficient polymorphism to

determine the genetic diversity and to establish genetic relationships among 16 landraces (Massawe *et al.*, 2002).

The AFLP marker technique has also been used to construct an initial genetic map of bambara groundnut. A total of 115 polymorphic AFLP bands were generated through selective amplification using *Eco*RI/*Mse*I primers with a total of 6 base-pairs of additional selection. These primers along with 1 SSR marker were scored in the $F_2$ population and used to construct the first genetic linkage map in this crop derived from the cross between the domesticated and wild landraces (DipC x VSSP11) (Basu, 2005).

### 1.5.4    Simple sequence repeats

Simple sequence repeats (SSR) or short tandem repeats (STR) are comprised of basic short motifs generally between 2 and 6 base-pairs long ( di- to hexa- nucleotide repeat units), they can be found in both coding- and non-coding regions for prokaryotes and eukaryotes (Chambers and MacAvoy, 2000; Ellegren, 2004; Kelkar *et al.*, 2008). They are reliable and popular markers as they possess; co-dominant inheritance, are relatively abundant, are multi-allelic, have extensive genome coverage, are highly reproducible and can often be simply detected (Mondini *et al.*, 2009). SSR markers are tandem repeats interspersed throughout the genome and can be amplified using a primer that flanks these regions (Grist *et al.*, 1993). These markers are often present at high levels of inter- and intra-specific polymorphism, particularly when the tandem repeat number is ten or greater (Bowen and Wheals, 2006). The number of SSR motifs at a locus is variable and it is believed that the basis for generation of high levels of polymorphism is  unequal crossing over and replication slippage, where the

transient dissociation of the replicating DNA strand is followed by misaligned re-association (Levinson and Gutman, 1987; Richards and Sutherland, 1992).

Microsatellites are not only present in the nuclear genome of bi-parental inheritance; there are also chloroplast and mitochondrion genome microsatellites which are usually inherited in a uniparental pattern (Soranzo *et al.*, 1999).

In plants, the presence of SSRs was first demonstrated by hybridization of oligonucleotide probes on phage libraries derived from tropical tree genomes (Condit and Hubbell, 1991). SSRs have become one of the most widely used molecular marker systems in plant genetics and breeding as the ultra-variability of SSRs among related organisms and a high degree of transferability between species make them an excellent markers for genotype identification, analysis of genetic population and diversity, phenotype and genetic mapping and marker assisted selection. It also enables the investigation of comparative genomic analysis (Powell *et al.*, 1996; Tautz, 1989; Varshney *et al.*, 2005a). SSRs are the only molecular markers to provide clues about which alleles are more closely related to each other (Caporale, 2003) with mutation often leading to repeat unit changes in fragment size.

### 1.5.4.1 Inter-Simple Sequence Repeat

ISSR (inter-simple sequence repeat) is a general term for a marker system based on amplification of the genomic region between microsatellite loci (Kebour *et al.*, 2012). The technique is reported by Zietkiewicz *et al.* (1994) and is similar in many ways to RAPDs. DNA fragments of about 100–3000 bp located between adjacent, oppositely-oriented microsatellite regions (Kumar *et al.*, 2009) are amplified. ISSRs are mostly dominant markers, although a few are codominant (Gonzalez-Chavira *et al.*, 2006). They are not proprietary and can be synthesized easily and no sequence information is

required (Joshi *et al.,* 2000). Although the ISSR technique has a reproducibility problem, it is considered by some authors as simple, fast to detect polymorphism, cost-effective and highly discriminative (Shafie *et al.*, 2011) and reasonably reliable. ISSRs are considered useful in gene mapping studies (Zietkiewicz *et al.*, 1994).

### 1.5.4.2 Expressed Sequence Tag-SRR

EST-SSR was identified as a class of microsatellite by Adams *et al.* (1991). Such markers can be obtained by partial sequencing of random cDNA clones (Adams *et al.*, 1991). The expressed genes can be detected within EST libraries (Bouck and Vision, 2007). EST-SSRs have received reasonable attention with increasing numbers of ESTs being deposited in databases for various plants, especially model species (Ner-Gaon *et al.*, 2007). Between 1% and 5% of ESTs from different plant species have been found to contain SSRs suitable for marker development. (Kantety *et al.*, 2002; Scott *et al.*, 2000).

The generation of EST or genic SSR markers is relatively easy and inexpensive, especially in plant species with characterized genes and full-length cDNA clones, such as rice (Kikuchi *et al.*, 2003), because they are a by-product of the sequence data from genes that are publicly-available. Additionally, the sequence data for EST-SSRs are often available in GenBank and so can be easily downloaded to screen them for the presence of microsatellite repeats.

EST-SSR markers have been developed for a number of plant species, including cultivated peanut (Liang *et al.*, 2009) grape (Cordeiro *et al.*, 2001), rice (Temnykh *et al.*, 2001), durum wheat (Eujayl *et al.*, 2002), rye (Hackauf and Wehling, 2002),

barley (Thiel *et al.*, 2003), barrel medic (Eujayl *et al.*, 2004), ryegrass (Faville *et al.*, 2004), wheat (Peng and Lapitan, 2005), and cotton (Han *et al.*, 2006).

EST databases can be processed rapidly to develop EST-SSR markers at relatively low cost (Liang *et al.*, 2009). As EST-SSR markers are derived from expressed genes, they are more conserved and have a higher level of transferability to related species than genomic DNA markers (Saha *et al.*, 2004). In addition these SSRs appear to be more closely related to important agronomical traits (Qi *et al.*, 2010). They are, therefore, useful as anchor markers for comparative mapping across species, comparative genomics, and evolutionary studies  (Cordeiro *et al.*, 2001; Eujayl *et al.*, 2004; Kantety *et al.*, 2002; Saha *et al.*, 2004; Scott *et al.*, 2000; Thiel *et al.*, 2003), however the conserved nature of EST-SSRs in transcribed regions may also limit their degree of polymorphism as lower levels of polymorphism are often reported compared with genomic SSRs in crop plants (Rungis *et al.*, 2004; Scott *et al.*, 2000). The transferability of EST-SSR loci across species within a genus has in several studies been above 50% (Dirlewanger *et al.*, 2002; Eujayl *et al.*, 2004; Gaitán-Solís *et al.*, 2002; Peakall *et al.*, 1998; Thiel *et al.*, 2003) whereas the transferability of genomic-derived SSR loci across genera was poor (Peakall *et al.*, 1998; Roa *et al.*, 2000; Thiel *et al.*, 2003; White and Powell, 1997). Recently, the application of next-generation sequencing (NGS) technology (Illumina and 454 sequencing) for genome sequencing leads to the discovery of a large number of genome-wide and genebased microsatellites in plant much more efficiently (Jun *et al.*, 2011; Wang *et al.*, 2012; Zalapa *et al.*, 2012).

**1.5.4.3 Application of microsatellites in different plant species**

Microsatellites as a DNA marker have been intensively used in a wide variety of organisms including humans (Beckmann and Soller, 1990), animals (Moore *et al.*, 1991), plants (Condit and Hubbell, 1991) and bacteria (*E. coli*; (Gur-Arie *et al.*, 2000).

However single nucleotide polymorphisms (SNPs) are replacing SSRs for some applications as more species are sequenced, their application in related minor species is limited due to the restriction of their transferability within species (Bakker *et al.*, 2006; Vezzulli *et al.*, 2008). In contrast, SSR markers will potentially continue to play an important role in different genetic studies of many minor plant and insect species in the future due to their efficiency, transferability to related species, multiple alleles and cost-effectiveness and ease of use for assessing small numbers of loci (Wang *et al.*, 2009).

Microsatellite markers can facilitate comparative mapping among different plant species to discover genome syntenies in order to develop DNA markers to specific chromosomal regions for marker-assisted selection. Anchoring the physical map of model plant species of *Arabidopsis* and soybean required SSRs in specific chromosome regions (Shoemaker *et al.*, 2008; Shultz *et al.*, 2007; Wang *et al.*, 1997) to assemble genome DNA sequences and joins together large pieces of overlapped DNA fragments.

SSR markers can be transferred to related species and to minor crops, which can facilitate the construction of genetic maps in minor species and enrichment of regions with low marker saturation. Placing barley SSR markers on rye and wheat genetic maps is an example of one such case (Varshney *et al.*, 2005b).

SSRs have become the marker of choice for evolutionary studies of natural plant populations as they highly polymorphic, genome-wide and multiallelic. Data collected from Chloroplast SSR molecular assays has been used to survey the cytoplasmic diversity among wild and cultivated barley populations. (Provan *et al.*, 1999).

There are many different applications of microsatellite in plant species. Over 80 genetic linkage maps have been constructed which include SSR markers with the first one instance being reported in tropical trees (Condit and Hubbell, 1991), followed by major and minor crops, fruits and vegetables (Akkaya *et al.*, 1992; Wu and Tanksley, 1993; Zhao and Kochert, 1993). In soybean, iron deficiency chlorosis associated with two common SSR markers was co-located in two separate populations (Wang *et al.*, 2008).

Germplasm classification for choice of parental selection has been calculated in different plant species through genetic distance and diversity assessment using the SSR marker system. Genetic diversity and phylogenetic relationships from germplasm collections for cultivated and wild peanut collection have been evaluated by SSR markers (Barkley *et al.*, 2007; Cuc *et al.*, 2008).

Microsatellite markers have also been used in bambara groundnut. Microsatellite primers developed from related legume species including cowpea, mungbean, common bean and soybean, were screened for identification of polymorphisms in the two parents 'DipC' and 'VSSP11' of a segregating $F_2$ mapping population in bambara groundnut. Out of 15 tested nine primers were amplified in both parents and only one (from soybean) was detected as a polymorphic marker and this was used with the polymorphic AFLP markers in constructing wide genetic map in this crop (Basu, 2005). In another investigation a set of 22 polymorphic SSR markers (from bambara

groundnut, azuki bean, cowpea and mungbean) were used to study genetic diversity in a collection of 240 accessions of bambara groundnut of Africa and Asia. A total of 166 alleles were detected, with a mean of 7.59 alleles per locus (Somta *et al.*, 2011b). Later a set of 75 microsatellites within species were characterized and used to investigate the genetic diversity of 24 bambara groundnut landraces of Africa (Molosiwa, 2012)

The level of inter- and intra-landrace polymorphism was assessed within a small collection of 10 Ghanaian landraces, using 10 reproducible SSR markers. A high level of polymorphism was detected among the landraces investigated (Siise and Massawe, 2012).

### 1.5.5 Single Nucleotide Polymorphism

SNPs are a marker system that can differentiate individuals based on variation detected at the level of a single nucleotide base (or more) in the genome. Such variation represents all sequence differences between individuals (Kumar *et al.*, 2009). Although SNPs can be used as a powerful and high throughput automated marker system in different applications of linkage disequilibrium and QTL analysis of plant species, they are only amenable in major crops which have already sequenced (Park *et al.*, 2009). Although SNP can now be developed in coding sequence through Next Generation Sequencing approaches at reasonable cost, they are more common in non-coding regions of the genome as coding sequences are often under selective constraints (Mondini *et al.*, 2009). On average, one SNP every 170 bp was identified comparing the sequences from two different rice cultivars, which makes this marker system an attractive tool in plant genomes in constructing linkage maps, QTL analysis and marker assistant selection (Gupta *et al.*, 2001; Rafalski, 2002). Generally, the

frequency of SNPs in plant species is estimated to range from 1 in 30 bp to 1 in 500 bp (Park *et al.*, 2009).

A set of 359 SNP markers derived from genic and intronic regions of the wheat genome were tested against 20 wheat cultivars. Reasonable levels of diversity were observed among wheat cultivars for SNPs located in two of the genomes; A and B. Markers in the D genome show approximately half the level of polymorphism of those for the A and B genomes. It is anticipated that SNP markers will play an increasingly important role in the genetics and breeding of wheat (Chao *et al.*, 2009).

### 1.5.6   Diversity array technology marker (DArT)

Diversity Arrays Technology (DArT) is a hybridization-based microarray platform, generic and cost effective as a genotyping technology and recognized as a suitable technique for genome-wide discovery and description of genetic variation. This technique was developed to overcome some of the limitations of other molecular marker technologies such as RFLP, AFLP and SSR (Akbari *et al.*, 2006).

This technology can be summarized as follows: complexity reduction using a chosen combination of restriction endonucleases, followed by adaptor ligation to amplify *Pst*I-*Pst*I fragments to obtain a genomic representation of the available genetic diversity, library creation so that each *E coli* colony contains one of the fragments, micro-arraying the library onto glass slides, treatment of individual genotypes with the same complexity reduction technique and hybridization of fluoro-labeled DNA onto the slides, scanning of the slides for hybridization signal and finally data extraction and analysis using DArT software (Mondini *et al.*, 2009).

DArT allows the simultaneous scoring of thousands of restriction site based polymorphisms between genotypes and does not require DNA sequence information or site-specific oligonucleotides (Alexander *et al.*, 2005). Screening a library of several thousand fragments from a sample of DNA fragments representing the whole genomic diversity in the species is key to the efficient discovery of new DArT markers, with all markers on a particular array being scored simultaneously (Jaccoud *et al.*, 2001). For these reasons, it is considered an important tool for scientists studying and managing genetic diversity with high throughput and low-cost data. (Semagn *et al.*, 2006b). DArT markers are applicable to model and non-model species/crops in addition to complex polyploid genomes, whereby the presence of individual fragment versus absence in genomic representative can be detected (Jaccoud *et al.*, 2001).

**The state of DArT development in plants**

DArT technology has been used successfully in genotyping of many species. Initially it was developed for rice, a diploid species with a small genome of 430 Mbp (Jaccoud *et al.*, 2001) and afterward it has been applied to many other species including *Arabidopsis* (Wittenberg *et al.*, 2005), cassava (Xia *et al.*, 2005), and for species with more complex genomes, e.g. barley (Wenzl *et al.*, 2006) and the allopolyploid wheat (*Triticum aestivum* L.) (Akbari *et al.*, 2006). This technology has been used recently to study the genetic diversity of bambara groundnut (Molosiwa, 2012; Olukolu *et al.*, 2012).

The first version of a high density functional DArT genotyping microarray has been reported in 64 different *Eucalyptus* species of flowering trees and shrubs, it has been identified that 56% of the total number of cloned fragments were found to be

polymorphic among 284 individuals. This array will provide a high-throughput platform for population genetics and phylogenetics in *Eucalyptus* where many of the hybrids used commercially are inter-specific (Sansaloni *et al.*, 2010).

Diversity Array Technology has been applied to oat (*Avena sativa* L.) allowing the construction of a high density genetic map, solving most of the problems of lower-throughput obtained using other technologies. Three discovery arrays were run and 2000 polymorphic markers were identified and 1010 of them were used to saturate and improve the 'Kanota' × 'Ogle' genetic map (Tinker *et al.*, 2009).

DArT markers have been integrated with other markers types in different crop species. A high-density genetic map for a doubled haploid population from a wide barley cross was constructed using 442 DArT markers along with 536 SSRs. High levels of polymorphism, co-linearity and telomeric coverage of the genome were obtained, compared to a previously constructed map (Hearnden *et al.*, 2007).

The DArT array has proved to be a useful marker system in *Sorghum bicolor*. Over 500 markers have been detected in this staple food and fodder crop, to be used in the analysis of genetic diversity and in construction of a medium-density genetic linkage map. This large number of DArT markers offers a more precise estimation of genetic relationships and provides valuable molecular breeding and genomic applications for this crop (Mace *et al.*, 2008).

### 1.5.7   Next generation sequencing

The original Sanger sequencing method developed in 1977 (Sanger *et al.*, 1977) and its modifications (Madabhushi, 1998; Smith *et al.*, 1986) after discovery of the polymerase chain reaction (PCR) in 1983 (Inoue and Orgel, 1983) have prevailed in

the DNA sequencing field for nearly three decades. Although in the past decade the length of Sanger sequence reads has increased dramatically from 450 bases to more than 1kb, the primary restrictions are cost per sequence and the need to prepare each sequencing template separately (Smailus *et al.*, 2005; Varshney *et al.*, 2009a). So to try to overcome these limitations alternative next generation sequencing technologies have been developed, which dramatically increase throughput (although are currently still limited in sequence read length) (Angeloni *et al.*, 2012).

Next generation sequencing can be defined as a high-throughput sequencing method that parallelizes the sequencing process, producing thousands or millions of sequences at once (Stapley *et al.*, 2010). Genome analysis has taken off with the advent of NGS systems with the production of vast quantities of sequence at relatively low cost and in a very short time compared to an automated Sanger sequencing (Schloss, 2008). However NGS technologies still produce shorter reads and have greater error rates than Sanger sequencing, sacrificing assembled sequence quality for speed and greatly reduced costs (Berkman *et al.*, 2012).

Next generation sequencing technologies, including Roche 454, Illumina GA, and ABI SOLiD, have been developed to generate more sequencing with lower cost than Sanger methods on ABI 3730xL platform (Bonetta, 2006; Schuster, 2008). The types and volumes of data produced through these technologies have increased greatly over the last few years, promising further accelerated progress (Pareek *et al.*, 2011). Sequencing more than 20 million base pairs in a 4 hours period and considered as the first available system for NGS was developed by 454 and commercialized by Roche as the GS20 (Margulies *et al.*, 2005). The greatest data volume of any NGS platform is generated with HiSeq2000 from Illumina with been up to 600 Gbp of usable data per run (http://www.illumina.com) (Barski *et al.*, 2007; Johnson *et al.*, 2007;

Varshney *et al.*, 2009a). The SOLiD System from Life Technologies based on sequential ligation with dye-labeled oligonucleotides, can generate 20 – 30 Gbp of data per day, with read lengths up to 75 bp for the latest 5500xl system (http://www.appliedbiosystems.com) (Berkman *et al.*, 2012).

The efficient use of two major NGS technologies of 454 pyrosequencing and Illumina has been demonstrated recently to generate large numbers of genome or transcriptome sequences. The sequence data generated from both techniques can be used efficiently to discover SSR loci and identifying the high quality loci for marker development, using the latest computational tools and high-throughput methods, with reduced cost and effort compared to the traditional Sanger approach (Zalapa *et al.*, 2012).

Next-generation transcriptome sequencing (either by *de novo* assembly of the transcriptome sequence data or by aligning reads to a genome sequence) can be applied normally to analyse gene expression and sequence variation of genomic loci (Morozova and Marra, 2008). With next generation sequencing, the whole genome of many related organisms can be achieved, allowing comparative and evolutionary studies in different organisms regardless of the state of previous genetic information (Metzker, 2010). With modern NGS technologies it is possible to identify and track genetic variation more efficiently and precisely, and hundreds or even thousands of candidate genes could be tracked within a large genebank (Kilian and Graner, 2012).

**Application of NGS in crop species**

Traditional approaches of genome sequencing are progressively being substituted by assemblies of NGS technologies and they are currently being used to explore *de novo* genome sequencing in several crop species such as wheat, pigeonpea and common

bean (Varshney *et al.*, 2009a). Genome sequencing and /or resequencing with the ongoing revolution of steady cost reduction and increased read length, could be extended beyond the major crops and model plant species to involve underutilized species and even the progeny lines of mapping populations. This will allow the breeder to follow the segregation of fragments easily from parents to their progeny. These fragments can be placed on the genetic and physical map with more certainty, which helps to confirm the introgression lines for the target traits and introduction of specific fragment from one species to another (Varshney *et al.*, 2009a). An overview of NGS applications relating to crop genetics and breeding is shown in Figure 1-5.
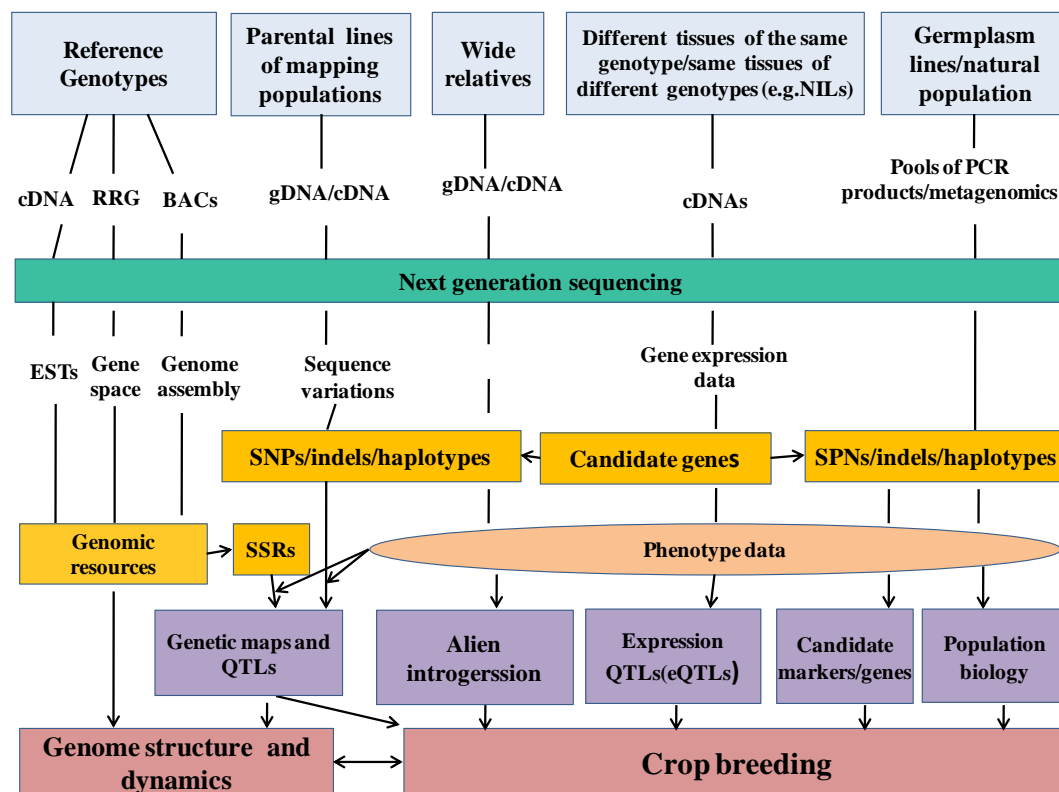


**Figure 1-5: An overview of potential NGS applications in crop genetics and breeding. The progress including the generation of genomic resources, marker development and QTL mapping, wide crosses and alien gene introgression, expression analysis, association genetic and population biology (Varshney *et al.*, 2009a).**

Using such innovate technologies as NGS will accelerate the possibility of importing genetic knowledge into underutilized species from major or model species and gives a new vision of working with the genomes of clusters of related species instead of single species. The significant sequence data generated through these techniques also allows DNA markers to be produced at reasonable cost in an unknown genetic background which can be exploited for quality control, genetic diversity analysis and practical breeding approaches (Mayes *et al.*, 2011).

RNA-seq is one of the most important applications of NGS technologies that is speedily replacing other methods of studying gene expression in non-model species without a sequenced genome (Morozova *et al.*, 2009). With this method all RNA transcripts can be sequenced. It is practical in underutilized crop species because a reference genome is not strictly required (Strickler *et al.*, 2012). Legumes *Pisum sativum* and *Cicer arietinum* were among the first transcriptome characterized species (Garg *et al.*, 2011; Iorizzo *et al.*, 2011).

## 1.6 Construction of genetic maps

The study of genetic linkage mapping dates back to the beginning of last century when scientists began to understand the recombinational nature and cellular behaviour of chromosomes (Wu *et al.*, 2008). The first genetic linkage map of chromosome *X* of *Drosophila melanogaster* was constructed by Sturtevant (1913). With the advent of different molecular markers derived from different techniques it is now possible to follow the segregation of valuable alleles in the populations and map them (Cholin, 2009).

By the end of last century the introduction of DNA based markers caused genetic maps to become much more densely populated, generally into the range of several hundred to more than a thousand markers per genome (Wu *et al.*, 2008).

Linkage maps indicate the position and relative genetic distances between markers on the chromosomes, allowing the identification of regions that contain genes and QTL related to interested traits (Paterson, 1996; Vinod, 2009). Genes which are tightly linked will be inherited together in the progeny much more frequently than those that are located further apart or even on different chromosomes (Semagn *et al.*, 2006a). The recombination frequency of genetic markers allows the genetic distance between the markers to be determined. There are three main steps to constructing a linkage map: production of a mapping population, identification of polymorphic markers for mapping and linkage analysis of the markers which includes statistical procedures and the use of appropriate software packages (Collard *et al.* 2005).

### 1.6.1   Mapping populations

A segregation population is required to construct the genetic linkage map. Differences should be present at least in one trait between the parents of the cross selected for the mapping population. If the marker-trait association is to be attempted the individual offspring of the population must be phenotypically evaluated (Collard *et al.*, 2005). Segregating populations such as $F_2$, $F_3$ or backcross populations are frequently used to construct a genetic linkage map, however the 'immortal' populations of recombinant inbred (RIL) and doubled haploid are preferable because they allow replicated and multi-location experiments (Collard and Mackill, 2008). It has been reported that the basic population size for genetic mapping ranges from 50-250 individuals; however

larger populations are often required for high resolution mapping (Ferreira *et al.*, 2006; Mohan *et al.*, 1997).

**Types of population used to investigate the genetic control of traits**

Different kinds of genetic population are often developed in autogamous species, such as: $F_2$, recombinant inbred lines (RILs), near isogenic lines (NILs), backcross populations and doubled haploid populations (DH) (Semagn *et al.*, 2010). These populations are developed by crossing two (often inbred) parents with clear contrasting differences in the phenotypic trait(s) of interest, to follow the segregation pattern of the desirable traits and subsequent genetic analysis (e.g. Quantitative Trait Locus analysis) or for direct breeding selection.

The steps to generate these populations involve selecting the parents, crossing them with each other through controlled transfer of pollen from one parent onto the receptive stigma of the other parent, then advancing the progeny in an appropriate manner to obtain a set of individual plants or lines with variation for the traits of interest. Besides differing for phenotypic traits, the parents of a mapping population must also be sufficiently different at the genotypic level for the chromosome recombination events to be tracked using molecular markers (Collard *et al.*, 2005; Young, 1994). Sometimes relatively little genetic polymorphism exists between the parents of crosses and this is often overcome through choosing genotypes which are highly genetically distinct, sometimes crossing to wild species to increase the level of polymorphism present (Kennard *et al.*, 1994; Muños *et al.*, 2011).

### $F_2$ and $F_3$ populations

These are quickly and easily developed by self-pollinating the $F_1$ hybrid. In inbreeding species, this is relatively simple, with a single controlled cross being the rate limiting step to develop the $F_1$ hybrid from inbred parental lines (Semagn *et al.*, 2006a). The primary advantage of both $F_2$ and $F_3$ populations is their ability to measure the effect of additive and dominant gene action at specific loci because they contain heterozygote individuals; however such populations are not fixable due to the heterozygous genetic constitution present in individual lines, which reduces with each generation of inbreeding, such that the proportion of heterozygous loci halves each generation, as the individual lines head towards being fully inbred. This lack of immortality of the lines restricts their use in analysis of QTL (Rakshit *et al.*, 2012). In an $F_2$ population a single plant represents each genotype, so replication over time or space cannot be carried out (Semagn *et al.*, 2010) and phenotypic evaluation on single plants are usually not considered reliable for quantitative traits like yield. This limitation of $F_2$ populations can be overcome by cloning $F_2$ plants in species that are easily cloned (e.g. by taking cuttings) to increase the replication of the traits. Another possibility is to produce $F_3$ plants from $F_2$ self-pollinated plants, each $F_3$ family consisting of multiple plants to be used for phenotypic evaluation. The assessment of single plants is replaced by an emphasis on the performance of $F_3$ families.

### Recombinant inbred lines (RILs) population

Recombinant inbred lines are homozygous (fixed) lines and can be developed from $F_2$ individuals using single seed descent or continuous inbreeding (Keurentjes *et al.*, 2011). One seed is taken from each $F_2$ individual and grown for a number of generations (until around $F_8$ to $F_{10}$) with single seed descent at each generation until

the lines are virtually homozygous at all loci. The result is a set of homogeneous, homozygous lines for almost all loci (Darvasi and Soller, 1995). These lines are true breeding and large amounts of seed can be produced for replicated trials, making such populations potentially immortal, for fully inbreeding species (Huang *et al.*, 2011). RILs (or doubled haploids, where available) are considered powerful to analyze quantitative traits effectively, with marker data only needing to be generated once (Collard *et al.*, 2005; Tanksley, 1993). Fully inbred lines can be readily circulated and new data could be continuously added to the previously constructed map. Genotype in this kind of population is represented by an inbred line, rather than by an individual, so genetic components of variance could be assessed more accurately when studying quantitative traits (Bailey, 1981). The main limitations in using this population type include the time and high cost for construction. Such populations will detect only additive effects without dominance effects. For species where the final product (such as the cultivar) is fully inbred, additive effects are likely to be sufficient. Where the final product is a hybrid between two inbreds, then dominance effects become important and these population cannot intrinsically reveal the effects of particular combinations of alleles within the hybrid (Haley and Anderson, 1997).

### *Backcross populations*

This type of population is easy to develop by crossing an $F_1$ plant to one of its parents (usually the 'better parent' for a trait). Similar to the $F_2$ population, each primary backcross genotype is represented by a single plant. This kind of population is less informative for linkage mapping compared to an $F_2$ population as recombination among markers occurs in only one set of gametes (either male or female), although interpretation is simpler (Lander *et al.*, 1987). BC1 populations can provide abundant

genetic information to analyze both additive and dominant effects. It can also be used to analyze QTL, however it is a temporary population and the population genetic structure will change with further self-pollination or backcrossing (Wang and Chee, 2010).

### *Doubled haploid populations*

Heterozygous $F_1$ individuals are used to produce doubled haploid lines (DH lines) using microspore culture or wide pollination techniques. They can also be produced by female gametophyte (usually unfertilized egg cells) culture (Devaux and Pickering, 2005; Zhang *et al.*, 2011). Production of haploids and chromosome doubling are two key steps in DH production. Haploid plantlets produced from embryos can undergo spontaneous doubling or this can often be induced through the use of Colchicine to produce doubled haploid lines (Forster *et al.*, 2007). DH lines have the advantage of producing homozygous or fixed lines that are no longer segregating in a single generation (Collard *et al.*, 2005; Semagn *et al.*, 2010). This kind of populations can be used efficiently in map construction and QTL analysis, particularly when the effect of the QTL is small (Devaux and Pickering, 2005). They have been used for QTL mapping in many species e.g. (Behn *et al.*, 2005; Semagn *et al.*, 2007; Semagn *et al.*, 2006c; Xu and Crouch, 2008).

### 1.6.2   Polymorphism identification

Identification of the polymorphic markers is the second step in genetic map construction. Enough polymorphic markers are required to be present between the parents of cross to construct the map (Young, 1994). The identified markers must then be screened against DNAs of the individuals of population and they have to reveal

differences between individuals to be called as polymorphic. Depends on their discrimination of homozygotes and heterozygotes they are classified in to dominant and codominant markers. Codominant markers can differentiate between homozygotes and heterozygotes, with the differences in the size 'alleles' apparent, while dominant markers are either present or absent within individuals (Collard *et al.*, 2005; Semagn *et al.*, 2006a).

### 1.6.3   Linkage analysis of the markers

Analysis of a linkage map composed of molecular markers makes it possible to detect the gene or QTLs controlling traits based on the genetic recombination during meiosis (Tanksley, 1993). Mapping programs are used to analyse the segregation patterns of DNA markers throughout coding the data for each marker in each individual of a population and conducting linkage analysis (Collard *et al.*, 2005). Logarithm of odds (LOD) value or LOD score is used to express the strength of linkage between markers (Risch, 1992). The Distance between these markers along a linkage map is measured in terms of the frequency of recombination between them, however  the genetic distance on a linkage map does not depend on the genome size of the plant species (the physical distance of DNA between genetic markers) but on the frequency of recombination between them (Paterson, 1996). Since recombination frequency is not linear, mapping functions are required to convert recombination fractions into additive centiMorgans (cM) distances. Two commonly used mapping functions are Kosambi and Haldane, assuming the presence of interference and assuming the absence of interference (i.e. no effect of crossing over in the neighbor regions, respectively (Hartl and Jones, 2001; Kearsey and Pooni, 1996)). Commonly used software programs and freely available online include Mapmaker/ EXP (Lander *et al.*,

1987; Lincoln *et al.*, 1993) and MapManager QTX (Manly *et al.*, 2001). The JoinMap software is another commonly-used program in constructing linkage maps, although it is not free (Stam, 1993; Van Ooijen, 2006).

Construction of a genetic linkage map is one of various applications of molecular markers in plant breeding. It allows detailed genetic analysis of qualitative and quantitative traits that enable the localization of genes or quantitative trait loci (QTL) (Doerge, 2002; Semagn *et al.*, 2006a; Yim *et al.*, 2002).

## 1.7 Quantitative Trait Loci

Genetic variation in nature often takes the form of a quantitative phenotypic range, with an approximately normal distribution, rather than of qualitative phenotypes that fall into discrete categories (Paran and Zamir, 2003). A QTL is a genome region that appears to contribute to a quantitative trait which can be localize by marker linkage map as first demonstrated by Paterson *et al.* (1988). It can be a single gene or it might be a cluster of linked genes that affect the trait. Mapping of QTL allows a statistical description of the effects of each genome region on quantitative traits (Vinod, 2009). QTL analysis studies the genetic variation in a controlled cross to locate the genes responsible and to explore their effects and interactions (Kearsey, 1998).

QTL mapping has facilitated investigations into the molecular basis of several traits in plant species such as *Arabidopsis* (El-Assal *et al.*, 2001), rice (Kojima *et al.*, 2002), maize (Thornsberry *et al.*, 2001), tomato (Frary *et al.*, 2000) and in soybean (Panthee *et al.*, 2007).

QTLs mapping analyses the association between the QTLs and genetic markers for linkage disequilibrium, requiring the information on both markers and phenotypic values of each individual in the population used (Huynh, 2009). When genes

controlling quantitative traits interact with each other or there is a genotype and environment interaction QTL mapping has to involve more sophisticated analysis (Mackay, 2001). With increasing complexity levels of QTLs, different techniques have been developed in plants to associate QTL with molecular markers using mapping population (Nagabhushana *et al.*, 2006). The commonly used techniques in QTL mapping are:

***Single Marker Analysis (SMA):*** Considered as the simplest method, does not require a complete linkage map. This analysis can be performed with basic statistical software programs to detect QTLs associated with single markers by testing the association between trait values and the genotype at each marker locus. Linear regression is the most commonly used statistical method for this analysis (Collard *et al.*, 2005). The programs used to perform this analysis include QGene (Nelson, 1997) and MapManager QTX (Manly *et al.*, 2001). The sample size is the first consideration as more robust associations can be obtained with larger samples of individuals and also provides greater sensitivity to detect QTL (Doerge, 2002). The main disadvantage of this method is that apparent QTL effects at a marker will be smaller than the actual QTL effect as a result of recombination between the marker and the QTL. Also when the markers are widely spaced (more than 15 cM) the QTL may be quite far from all markers causing the power of QTL detection to be decreased (Tanksley, 1993).

***Simple Interval Mapping (SIM or IM):*** Considered a more powerful method statistically compared to single marker analysis. It involves a test of association between trait values and positions within marker intervals along the linkage map to check for the presence of a QTL at many positions between two marker loci (Lander

and Botstein, 1989). The interval mapping (IM) approach uses information from informative markers simultaneously to estimate the QTL location and its effect (Knott *et al.*, 1996). The disadvantage of this sort of mapping is that the detection of QTLs could be misleading by giving positive or negative results due to possible interference of other markers (Rodriguez-Zas *et al.*, 2002). MapMaker/QTL (Lincoln *et al.*, 1993) and MapQTL (Van Ooijen, 2009) have been developed to carry out this analysis.

*Composite Interval Mapping (CIM):* This method has been developed to address problems with the previous two methods of QTL analysis. It combines interval mapping with linear regression and the possibility to allow additional genetic markers in the model in addition to an adjacent pair of linked markers for interval mapping (Jansen and Stam, 1994; Zeng, 1993; Zeng, 1994). In CIM the variance from other QTL is accounted for by including partial regression coefficients from markers in other regions of the genome (Basten *et al.*, 2001) and the effects of other QTL are not present as residual variance (Chakraborty and Zeng, 2011). Morever the bias that would normally be caused by QTL that are linked to the position being tested could be eliminated by performing this analysis (Nagabhushana *et al.*, 2006). To perform this analysis a number of programmes have been developed such as: Window QTL Cartographer (Basten *et al.*, 2001; Wang *et al.*, 2007), MapManager QTX (Manly *et al.*, 2001) and PLABQTL (Utz and Melchinger, 1996).

*Multiple Interval Mapping (MIM):* QTL mapping by MIM model was first proposed by Kao *et al.*, (1999). This model could fit multiple putative QTLs simultaneously to reveal their individual effects in addition to analyse epistatic QTL and heritability of quantitative traits. Thus, the components of genetic variance contributed by individual

QTL can be calculated and markers assigned to represent them (Chakraborty and Zeng, 2011). The software R/qtl can be used to estimate multiple QTL mapping for quantitative trait (Arends *et al.*, 2010).

However, none of these programs incorporates all the functions of QTL methodological analysis into one QTL mapping system, the QTLNetwork software package has been introduced to enable the detection of multiple QTLs which can simultaneously map quantitative trait loci (QTL) with individual effects, epistasis and QTL–environment interaction (Yang *et al.*, 2008).

## 1.8 Marker-Assisted Selection

In plant breeding, as quantitative traits are potentially controlled by several QTLs, and due to the cost of utilizing several QTLs only a few DNA markers that are tightly linked to agronomical important traits (no more than 3 QTLs) have be used as molecular tools for marker assistant selection in practice (Collard *et al.*, 2005; Ribaut and Betran, 1999). Up to 5 QTLs has been reported as being introgressed into tomato via MAS (Lecomte *et al.*, 2004). Markers detected to be associated with QTLs from preliminary mapping studies were not directly useful in MAS because of the possibility of inaccurate position of QTL due to sampling bias (Melchinger *et al.*, 1998). Confirmation and validation steps for these QTLs are preferable before using them without subsequent phenotyping in MAS (Langridge *et al.*, 2001), although there are some examples of highly accurate preliminary QTL mapping data determined by subsequent QTL mapping research (Price, 2006).

All QTLs selected for MAS should ideally be stable across environments and they should account for the majority of phenotypic variance for the trait (Hittalmani *et al.*,

2002; Ribaut and Betran, 1999; Tanksley, 1993). The pipeline of marker developments in QTL analysis and MAS can be described in Figure 1-6.
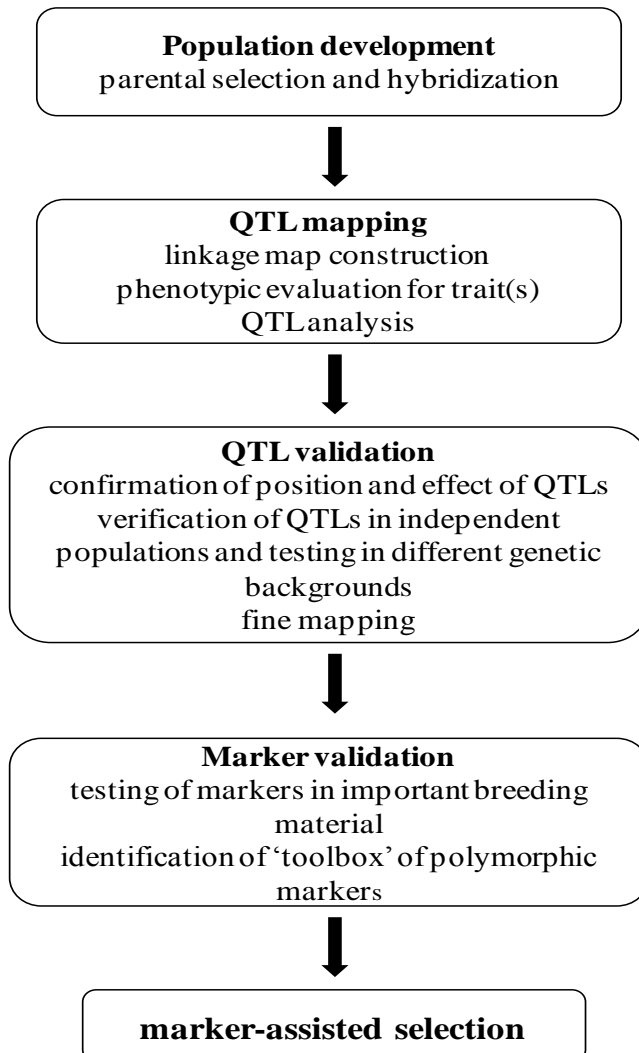


**Figure 1-6: Marker and QTL validation pipeline for MAS (Collard and Mackill, 2008).**

Other considerations for the markers used in marker assisted selection are they should be tightly linked to target loci at less than 5 cM genetic distance is preferable. Flanking markers or intragenic markers increase the reliability of the markers to predict phenotype. Good quantity and quality of DNA and simplicity of technical procedure for marker assay are also required. In addition they have to be highly

polymorphic in breeding material and be cost-effective (Mackill and Ni, 2000; Mohler and Singrun, 2004).

Marker-assisted selection (MAS) is an approach whereby a phenotype is selected based on the genotype of a marker. This approach has been developed to solve the problems of selection criteria in conventional plant breeding through shifting from phenotypic selection toward gene selection either directly or indirectly, which may greatly increase the efficiency and effectiveness in plant breeding compared to conventional breeding methods (Collard *et al.*, 2005). Once markers that are tightly linked to genes or QTLs of interest have been identified, prior to field evaluation of large numbers of plants, breeders may use specific DNA marker alleles as a diagnostic tool to identify plants carrying the genes or QTLs (Michelmore, 1995; Ribaut *et al.*, 1997).

Applying MAS in the early generations has a great advantage allowing the elimination of plants with undesirable gene combinations. Thus greater attention can be concentrated on fewer high-priority lines in subsequent generations. When the linkage between the marker and the selected QTL is not very tight, the greatest efficiency of MAS is in early generations due to the increasing probability of recombination between the marker and QTL in later generations (Collard and Mackill, 2008).

The trend of utilization of molecular markers arrays and genetic maps make MAS possible both for traits controlled by major genes and for quantitative trait loci (Francia *et al.*, 2005). Although the availability of public databases for markers and QTLs will lead to the greater adoption of MAS in major crops, the incorporation of MAS will still be very slow in plant breeding programmes for orphan crop species and in developing countries (Naylor *et al.*, 2004). Focusing on comparative genome

analysis for closely related crops especially between model crops and orphan species might be one of the solutions to facilitate the construction of high-density molecular linkage maps for gene localization and QTL mapping of important agronomic traits in these minor crops.

## 1.9 Aim and objectives of this study

The aim of this project is to construct and develop a genetic linkage map of an $F_3$ segregating population derived from the intra-subspecific cross (DipC x Tiga necaru). The cross has been developed from domesticated landraces which contrast for a number of agronomically important traits, including; extent of bunching, internode length, canopy spread, peduncle length, biomass and seed yield. A partial genetic map for an inter-subspecific cross between a domesticated landrace and a wild ancestor (Basu *et al.*, 2007c) will also be improved, traits analysed and the genetic maps compared. The mapping will develop, characterize and utilize microsatellites and Diversity Arrays Technology (DArT) markers to carry out a QTL analysis of economically important traits (Jing *et al.*, 2009; Semagn *et al.*, 2006b; Xue *et al.*, 2010; Zhang *et al.*, 2004). This should help to identify marker-trait linkages useful for the introgression of desirable genes or QTLs for marker-assisted selection in breeding programs of bambara groundnut. The analysis should also give a fundamental insight into the genetics underlying important agriculturally relevant traits (as well as the process of domestication, through a more detailed analysis of the 'wide' cross) which could help inform conventional breeding selection.

**The objectives of this study were:**

- Validation of the population under study using microsatellite markers to reveal the residual heterozygosity in the population.

- Development of polymorphic SSR and DArT markers to construct the genetic linkage map.

- Construction of a genetic map in the segregating population derived from an intra-subspecific narrow cross.

- Comparison of the 'narrow' intra-subspecific genetic map with the 'wide' inter-subspecific genetic map.

- Evaluation of important phenotypic and agronomic traits through selection in two different environments (Field and controlled environment glasshouses in Indonesia and the UK) for QTL trait analysis.

- QTL analysis of the important agronomic traits.


**Thesis outline**

Chapter1: Presents the introduction and the aims of the project. It also provides a review of the existing literature on underutilized crops, focusing on bambara groundnut. Different types of molecular marker and their application in different plant species, construction of genetic linkage maps, QTL analysis and MAS are also reviewed in this chapter.

Chapter 2: Explains the methodology of molecular marker analysis using SSR and Diversity Arrays Technology (DArT) markers, phenotypic evaluation in glasshouse and field experiments, data preparation, construction of the genetic maps and subsequent QTL analysis.

Chapter 3: Focuses on the results of screening the SSR primer pairs (derived from different microsatellite-enriched libraries) for polymorphism in the segregating populations, DipC X Tiga necaru and DipC x VSSP11. It also describes the development of DArT markers and the validation of the populations under study, through evaluating the presence of residual heterozygosity and evaluation of the populations for contaminants/out-crosses.

Chapter 4: Describes the distribution, segregation of agronomic traits and their associations for the $F_3$ population derived from the DipC and Tiga necaru landraces, in controlled environment glasshouses (Sutton bonington campus, UK) and in the Field (Bungah field, Gresik, Indonesia). Data collected previously from an $F_2$ population of the same cross in the Tropical Crops Research Unit was also involved alongside the $F_3$ data analysis.

Chapter 5: Presents the results of genetic linkage map construction for both narrow (DipC X Tiga necaru) and wide (DipC x VSSP11) crosses, with a shared maternal landrace (DipC) parent, using AFLP, SSR and DArT markers. This chapter also describes the initial attempts to combine the wide and narrow genetic linkage maps.

Chapter 6: Describes the QTL analysis of agronomically important traits and their inter-correlations.

Chapter 7: Overviewed the results and discussion of different chapters and also concluded some future works.

# Chapter 2.  MATERIALS AND METHODS

This chapter is divided into four sections. It starts with the molecular biology experiments in the laboratory (DNA and marker techniques). A segregating $F_3$ population derived from a narrow intra-subspecific cross and an $F_2$ segregating population from a wider inter-subspecific cross were used. The second section describes characterization of agronomic traits and their segregation patterns in a $F_3$ population derived from a controlled cross between the DipC and Tiga necaru landraces, assessed both in the field and the glasshouse. The third and fourth sections explain the methodology of constructing a genetic linkage map and QTL analysis, respectively, in both segregating populations from the narrow and wide crosses of bambara groundnut. Details of the materials and methods used for each aspect of the investigation have been described in their appropriate sections.

## 2.1 Development of polymorphic SSR and DArT markers and validation of the $F_3$ population

### 2.1.1 Plant materials

A segregating $F_3$ population derived from an intra-subspecific cross between the DipC (female) and Tiga necaru (male) landraces which were obtained originally from IITA was used as the plant material for molecular analysis using SSR and DArT markers. Although there was limited seed material available for the $F_3$ segregating population under study, it was possible to investigate a total of 73 lines in the glasshouse and in a field trial. Another segregating $F_2$ population, previously studied in the construction of an initial genetic linkage map for the cross of DipC x VSSP11, was also analysed.

## 2.1.2 DNA Extraction

For each line, approximately 5g of fresh leaf from the Glasshouse (25 leaflets in total) was picked from the bulk of plants within lines - from 4-10 plants per line were sampled, depending on the seed availability for that line - and flash frozen in liquid nitrogen. All 73 lines were transferred to the laboratory. The leaves were ground with liquid nitrogen using a pre-chilled mortar and pestle, after which the powder was transferred to a pre-chilled 50 ml falcon tube (Sarstedt) to begin DNA extraction. Genomic DNA was extracted from all the lines of the segregating population using the Dellaporta protocol (Dellaporta *et al.*, 1983) with some modifications. DNA extracted was re-suspended in 500ul of 1X TE buffer. Samples were kept in 1.5 ml Eppendorf tubes and stored at $-20^{o}C$. All work with DNA stocks was carried out on ice. For the individuals of the $F_2$ population GenElute Plant Genomic DNA kit (Sigma Aldrich) was used to extract DNA from silica dried leaves for all individuals and their parents.

## 2.1.3 DNA quantitation

Gel electrophoresis was used to quantify and test the quality of genomic DNA. A known volume of each sample was loaded onto a 1% Agarose Molecular Grade (Bioline) gel in 0.5 x TBE (Tris borate EDTA) buffer alongside a range of uncut lambda DNA standards containing 500 to 25 ng DNA. The gel contained 1µL of 10mg/mL ethidium bromide stock per 100ml of gel (Promega Corporation). This was added once the dissolved gel had cooled enough to be handled safely and the gel was swirled before pouring. The gel was allowed to fully set before submerging in the same buffer in a Biorad Mini Sub-Cell GT kit and the comb removed. In the first lane, 5µL of 2-log DNA ladder (New England Biolabs) was loaded, followed by DNA

samples of bambara groundnut premixed with 6x loading buffer (2µL DNA + 5µL 6x loading buffer) for each sample. Uncut lambda DNA (50ng/µL) was loaded in the following amounts; 10µL, 5µL and 2.5µL to act as a fluorescence comparison with the unknown samples, allowing an estimate of sample concentration to be made. The gel was run at 90 Volt for 2 hours. The concentration of each DNA sample was determined approximately by comparing the fluorescence of the unknown sample under UV light with lambda DNA bands for the same gel. The quality of DNA can also be determined by visualization, comparing the size of extracted fragments with the 2-log ladder and lambda samples. After quantitation DNA samples were diluted to 10ng/µL to be used in PCR reactions.

### 2.1.4 Molecular markers

Two molecular marker techniques, microsatellite (SSR) and Diversity Array Technology (DArT) were investigated in this study to analyze the differences at molecular level in the segregating $F_3$ population of the 'narrow' cross. Residual heterozygosity was also revealed with polymorphic SSR markers, allowing the validation of the identity and generation of the population under study. Microsatellite markers were also applied to the $F_2$ population derived from the 'wide' cross (DipC x VSSP11) and the population was also analysed by DArT, allowing additional markers to be added to the initial linkage map (primarily based on AFLP markers).

### 2.1.4.1 Microsatellite markers developed and used in this study

A total of 124 SSR markers were tested for segregation in the wide and narrow crosses of bambara groundnut. These markers were derived from a number of sources,

including a 454-sequenced microsatellite-enriched library and a bambara groundnut leaf transcriptome. One SSR marker from soybean was also used in this investigation.

Roche 454 pyrosequencing technology was used to facilitate microsatellite discovery and sequencing. The libraries were developed using the method of Edwards *et al.* (1996). The technique involved digestion of genomic DNA with a restriction enzyme. Digested fragments were ligated to an adapter, then they were amplified with PCR, before hybridization to immobilized synthetic SSR repeat motif oligonucleotides. After stringent washing, the repeat-enriched amplicons were recovered from the hybridization filters and amplified. The pool of PCR fragments enriched for simple sequence repeats was 454 Pyrosequenced (Roche) using a 1/16th plate run with Titanium reagents. This was expected to give an average read length 400 –450bp, as defined by the manufacturer.

Sequences of microsatellite repeat motifs were used to screen for repeats within individual sequence amplicons using the MISA.pl (Perl) script (pgrc.ipk-gatersleben.de/misa/). The M13 5'-Tag approach was used to allow generic labeling with a third (M13 sequence) primer which was directly labeled with one of the following dyes; D2-D4 (Black, Green, Blue; WellRED, Sigma Aldrich) (Schuelke, 2000).

The primers used in this investigation were classified to three primer sets according to their development sources:

The first set of 57 primers was previously developed. Primer pairs amplified PCR fragments of the expected size were screened for polymorphism in the segregating populations. These were developed using Sanger sequencing of individual clones. Labeling used a M13 5'-Tag approach. One microsatellite marker from the related

legume, soybean (*Glycine max* L.) (Peakall *et al.*, 1998) was also tested for polymorphism within the individuals of $F_3$ progeny (Appendix 1).

The second set of 37 SSR primer pairs were also designed from the bambara groundnut microsatellite-enrich library (Mayes, unpublished data), this time using a $1/16^{th}$ 454 Pyrosequencing run with Titanium reagents, allowing longer sequences (expected length 300-350bp) to be generated. A total of 37 primer pairs were designed by the current author with 32 fragment sequences being screened for polymorphism in the populations under study (Appendix 2). The first work on this library started with identifying the unique fragments containing microsatellite repeats from sequences contained SSRs. SSR parameters, including motif, size, repeat start point and end point were used to identify unique fragments from fasta dataset.

The third set of 29 primers was developed from bambara groundnut RNA deep sequencing transcriptome library. RNA-based sequence data was generated previously using a 1 plate 454 Pyrosequencing from the S19-3 genotype of bambara groundnut as part of an investigation into temperature effect on gene expression in bambara groundnut leaf by Sean Mayes' group. The sequences were screened in *silico* for microsatellite repeat motifs. Of 68 primer pairs synthesized 29 gave polymorphic SSR primers against a standard set of 24 genotypes representing Bambara groundnut germplasm. These were tested as the third set of microsatellites on both wide and narrow cross populations (Appendix 3).

**Detailed microsatellite methods**

*Primer Design*

Web-based programs Primer3 version 0.4.0 (http://frodo.wi.mit.edu/primer3/) by Rozen and Skaletsky (2000) was used to design the primers flanking the identified SSR. In the program the unique sequences were pasted into the input field. The target SSR was identified using a square bracket "[ ]". Triangular brackets "< >" were used to exclude parts of the sequence such as adapter or obvious repetitive or poor sequence regions. Some of the default parameters were adjusted manually in the software package to screen for the required primer pair; for instance the product size range was fixed to start from 100 bp instead of the default of 150bp, primer GC% minimum not to be less than 40 and an effort was made for Max 3' Self Complementarity not to be more than 3 (Figure 2-1).



**Figure 2-1: Default parameters condition to pick the primer using Primer3 (v. 0.4.)**

The result of designed primer by web-based program (Primer3 v.0.4.) was picked and

saved as a text file (Figure 2-2).

```
F4GBQGD03F5L0A_LEN=295_QL=36_QR=262, (GA)14, primer name BN 145


OLIGO              start  len       tm      gc%    any    3' seq
LEFT PRIMER         126    19    59.86   52.63   4.00   0.00 GGCACTGGTAGCAACGAAA
RIGHT PRIMER        248    22    61.33   50.00   4.00   0.00 CGTGGACGTAACAACACAACAC
SEQUENCE SIZE: 295
INCLUDED REGION SIZE: 295

PRODUCT SIZE: 123, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00
EXCLUDED REGIONS (start, len)*: 1,35

     1 tcagcgacgtgactCTCTTGCTTACGCGTGGACTACATACCCATTGTGCCAGAGGCATCG
       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

    61 CTTCCACCGTAGAAAGTAGCGTGGGCACCCTGCCATGCACCGCCGGTGTAGACACCGGGA


   121 ATTCTGGCACTGGTAGCAACGAAAAGGGTAACCAACGATGCAATGCAGAGCGAAACCGTT
          >>>>>>>>>>>>>>>>>>>

   181 CCAGCCATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTAGTTGTGTTGTGTTGTTA
                                                <<<<<<<<<<<<<<

   241 CGTCCACGACGTAAGACAAGAGctgagactgccaaggcacacgagggganaggnn
       <<<<<<<<
```

**Figure 2-2: Primer pairs designed with the Primer3 software. The SSR motif repeat has been highlighted, the forward and reverse primers were denoted by ">>>" and "<<<", respectively.**

The modified M13 (-21) tail sequence (CAC GAC GTT GTA AAA CGA C) was

added to the 5'-end of the forward primers (Schuelke, 2000), to give Forward-Tag

primers. Fluorescently labeled M13 universal primer, labeled with blue dye (D2-D4),

was ordered from Sigma Aldrich/Proligo at the 10 OD scale. Reverse and F-Tag

primers were all ordered from MWG Eurofins. The three primer labeling approach

(Schuelke, 2000) with modified M13 sequence was used to label the SSRs prior to

running the PCR products on a Beckmann CEQ$^{TM}$ 8000.

### *Preparation of main stocks of primer (200pmol/µL)*

To prepare a 200pmol/µL (1000x main stock) from the ordered primers, Sigma Molecular biology grade water was added into the lyophilized primer containing vials to give a final concentration of 200pmol/L, based on the synthesis report. The tubes were vortexed, spun down and placed on ice for 30 min. After that a 5µL volume was taken from the main stock of both Forward & Reverse primers, separately, and mixed with 495 µL SDW, to produce a 10x primer stock 10X ( 20 µM/µL) for each primer. These were kept in a freezer at -20°C.

### *Three primer design*

The system of three primer design was reported by Schuelke (2000). This approach allows a large number of potential microsatellites to be screened at reduced cost. Detecting the length of PCR product by capillary sequencer requires one of the primers to carry a fluorescent label. In this system, a specific sequence is added to the 5' end of the microsatellite-specific forward primer. This F-tag primer is used along with the previously used sequence–specific Reverse primer and 'universal' fluorescently–labeled M13 (-21) in the PCR reaction.

The forward primer used in the reaction was $1/10^{th}$ of the standard amount. The other $9/10^{th}$ of the forward primer are made up with the Universal fluorescent-labeled M13 primer. The reaction conditions were chosen such that during the early cycles, the forward primer with its M13 (-21) Tag (CAC GAC GTT GTA AAA CGA C) is incorporated into the accumulating PCR products. Thus, the universal M13 (-21) 'takes over' as a Forward primer and incorporates the fluorescent dye into the PCR product in later cycle.

## *Primer optimization*

All the primers were optimized to identify the optimal and working annealing temperature at the gradient range of 50-65$^{o}$C using a Thermo Hybaid Express PCR machine (Electron Corporation, Milford, MA, USA). Equal amounts of genomic DNA from six landraces at 10ng/µL concentration were mixed and used as a template in the primer optimization. The PCR reaction mixture is described as below:

2µL PCR Buffer (New England BioLabs; w/MgCl$_2$, final 1.5mM, 0.2µL Forward and 2µL Reverse primers (20 µM/µL), 1.8 µL of M13 Tag (20 µM/µL), 0.2 µL of dNTPs (0.25 mM each final concentration (Promega corporation), 0.1 µL of 5 unit Taq DNA polymerase (New England BioLabs), 2 µL of template DNA (10ng/µL) and 11.7 µL of sterilized distilled water. The master mix for each primer set was dispensed into 12 wells of 96-well plate (1-12) per primer pair. Eight primer pairs were screened per 96-well plate (Thermo Scientific). The plates were sealed with a Thermowell$^{®}$ Sealing mat (Fisher Scientific) and spun briefly in an Eppendorf refrigerated centrifuge (5180). Amplification was carried out across a gradient of 12 different temperatures with the following cycling regime: 94$^{o}$C for 3 minutes, 35 cycles of 94$^{o}$C for 1 minute, 12 temperatures from wells 1 to 12, ranging between 50-65$^{o}$C for 1 minute, 72 $^{o}$C for 2 minute; with a final extension at 72$^{o}$C for 10 minutes.

## *Gel electrophoresis of PCR products*

A 500ml agarose 2% gel was prepared in 0.5 X TBE.  2µL of ethidium bromide stock (10mg/ml) was added to the gel solution which was then poured into a taped Scot lab gel tray, with 26-well combs (Scotlab Maxi gel; model). 12 rows of combs (each row contained 26 wells) were fitted into the gel tray immediately after pouring. This allowed 3 x 96 well-plates to be analyzed at the same time. 5µL of 6x loading buffer

was added into each well of PCR products after amplification. The plate was spun briefly and 10μL from each well was loaded into the submerged gel, after removing the combs. Each primer pair reaction was loaded into one half of a row of the gel (each primer pair had 12 reactions across a 15$^o$C annealing gradient). 5μL of 2-log DNA ladder (New England Biolabs) was loaded in the first lane of each primer pair. The gel was run for 45 minutes at 120 Volt. It was visualized by illumination with UV light and images taken using a Biorad (Gel DOC 2000), then printed on a thermal printer (Mitsubishi P91). Based on the intensity and size of bands, the optimal annealing temperatures were determined for the primers.

### *PCR reactions for the individual lines of segregating populations*

The PCR reactions were carried out in a total volume of 20μL as above. The only difference was in the DNA template, with each well a single individual from the cross being analysed (plus parents and controls). The amplification was carried out in a ABI PCR 9700 Thermocycler machine (ABI, Carlsbad, CA, USA) with the following profile: Initial denaturation at 94$^o$C for 3 minutes was followed by 35 cycles of 94$^o$C for 1 minute, 50-65$^o$C for 1 minute (at the optimal annealing temperature derived from a gradient annealing temperature PCR reaction) and 72$^o$C for 2 minutes, followed by a final extension at 72$^o$C for 10 minutes.

### *Gel electrophoresis for PCR products of the individual lines*

Visualizing the PCR products of individual lines was conducted in the same way as described for primer optimization. Products were visualized to ensure that all samples had amplified before loading samples onto the Beckmann CEQ$^{TM}$ 8000.

### Capillary electrophoresis

The fluorescently labeled PCR products were loaded onto the Beckmann CEQ$^{TM}$ 8000 (Beckman Coulter Inc, Fullerton, USA). The Beckman Coulter CEQ$^{TM}$ 8000 Genetic Analyzer is a fully automated system for genetic analysis. It automatically fills the capillary array with a patented linear polyacrylamide (LPA) gel, denatures and loads the sample, applies the voltage program, and analyzes the data. All PCR products (SSR fragments) were sized on with a 400 bp size standard. The sample loading solution (SLS; Beckman coulter Inc, Fullerton, USA) was mixed with the size standard (SS; Beckman Coulter Inc, Fullerton, USA) in the ratio of 1:100 (v/v) and 25μL of the mixture was loaded into the individual wells of a new PCR plate. 2 - 4μl of the single or pooled PCR products (depending on the intensity of bands visualized on agarose gel and size range of SSRs under analysis) from each genotype was added into the mixture of SLS and SS solutions. Each well was overlaid immediately with a drop of mineral oil (Beckman Coulter, Inc Fullerton, USA).

### Capillary result analysis

The CEQ$^{TM}$ 8000 Fragments Analysis Software Version 8 (Beckman Coulter Inc., Fullerton, USA) was used to measure and analyze the fragment sizes of PCR products. The results were transferred to a Microsoft Office Word Document file. For each microsatellite marker the product sizes were scored manually in all individuals. The largest peak size (rounded to the nearest whole number) from the electrophoretograms was recorded as the allele size for each individual. Scoring was based on overall microsatellite allele patterns, rather than a specific peak, as the relative peak height can change between alleles and it is the overall pattern which is important to maintain a consistent scoring pattern.

**2.1.4.2 DArT marker**

Diversity Array Technology marker assays were developed and performed by DArT Pty. Ltd (Yarralumla, Australia; *www.diversityarrays.com* ) as previously described (Akbari *et al.*, 2006; Semagn *et al.*, 2006c; Stadler, 2009; Wenzl *et al.*, 2004). Genomic representations were prepared from 73 DNA samples from the individual $F_3$ plants of the cross (DipC x Tiga necaru), using both restriction enzymes representations (bgPabg and bgPt) as described in Stadler (2009). These representatives were hybridized to the existing 7700 clone array and scored by DArT Pty Ltd.

**2.1.5 Application of the microsatellite markers in the $F_2$ population**

A segregating $F_2$ population derived from the wide cross of DipC and VSSP11 were prepared from dried leaves that had been stored for 10 years. Genomic DNA was extracted from 98 individual lines of the $F_2$ population along with their parents and $F_1$ hybrid, using the GenElute Plant Genomic DNA kit (Sigma Aldrich). All microsatellite markers developed from different sources were tested for polymorphism in this population. The parents and $F_1$ hybrid were used first to detect potentially polymorphic markers and those showing different allele sizes on the Beckmann CEQ$^{TM}$ 8000 were screened against all individuals. Genomic DNA of individuals from this cross was amplified with a slightly increased cycle number (37-38) due to the relatively poor quality of the genomic DNA from this cross.

**2.1.6 Isolating DNA from Bambara groundnut seeds**

**2.1.6.1 DNA Extraction**

An experiment was carried out to extract the DNA directly from the endosperm of bambara groundnut seed, without affecting seed viability. The seeds were drilled on the opposite side to the embryo with a Dremel craft drill (Dremel 300 series, model 300 F013030046, mexico). The GenElute Plant Genomic DNA kit (Sigma Aldrich) was used (with some modifications) to extract the DNA from the endosperm powder. Three different incubation periods (10, 30 and 60 minutes at $65^0C$) were tested after adding both parts of lysis buffer A and B, to optimize the incubation period to obtain the highest yield of DNA from dried seed.

**2.1.6.2 Germination test**

The drilled seeds were tested for viability in a germination test, to check the influence of drilling the sample. Two drill diameters (3/64" and 1/16"), two different seed sizes (1 and 0.5 cm diameter) and untreated seeds, were used in a (complete randomized design) CRD experiment with three replicates of 5 seeds each. The seeds were treated with 10% Parazone bleach for 15 minutes, prior to the experiment, to surface sterilize them. Then they were rinsed three times with distilled water before use, to remove the effects of the Parazone. Germination testing was conducted in the germination room at $29^oC$ on Whatman filter paper in plastic petridishes, with 5 seed per dish. Water was applied once a day to ensure that the paper remained wet and the irrigation water was treated with File-X fungicidal (0.0625 ml/2.5L water) to protect the seeds from fungal disease.

## 2.2 Characterization of the segregating population

The segregating $F_3$ population lines derived from an intra-subspecific cross of DipC x Tiga necaru landraces were evaluated in a controlled environment and in a field experiment, to determine the segregation patterns and inheritance of morphological and agronomic traits identified in the two parents of the cross. $F_3$ seed derived from self-fertilization of 73 $F_2$ plants were utilized for both evaluations. Previous phenotypic data recorded for the second filial generation ($F_2$) of the same cross grown in the Tropical Crops Research Unit (TCRU) at Nottingham University were also used in this study for trait and QTL analysis.

Phenotypic evaluation was performed for the $F_3$ segregating lines and their parents in a fully controlled glasshouse (FutureCrop glasshouses) at the School of Biosciences, Sutton Bonington campus, University of Nottingham. The experiment was set in complete randomized design (CRD) with 4 replicates. The individual plants represent each replicate. The seeds were planted on 8th Aug 2011 after preparation of the soil bed and seed were planted at a 2-3 cm depth under the soil. The glasshouse experiment was run for 5 months and plants harvested on the 8th Jan 2012. They were grown in rows with a 25cm x 25cm spacing between individual plants. Photoperiod was set at 12 hours using an automatic blackout system with a $28^{o}C$ day time temperature and $23^{o}C$ during the night, set for a 12/12 split. Trickle tape irrigation was used to irrigate the plant for 20 minutes, twice per day. Irrigation was adjusted as needed to maintain well watered conditions (Figure 2-3). A field trial was also carried out for all $F_3$ progeny lines in Bungah field, Gresik, Indonesia on 30th May 2010. Seasonal constraints at that time of the year in Indonesia constrained the life cycle of this crop to around four months and it was harvested on 26th Sep 2010. The available seeds for each line (5-15 seeds) were planted directly into the soil with 40 cm between

rows and 40 cm between the plants within rows. The morphological and important

agronomic traits here were also characterized according to descriptors in the book

'Bambara groundnut (*Vigna subterranea* ) (IPGRI, 2000).



**Figure 2-3: Trickle tape irrigation for $F_3$ population of bambara groundnut grown in the FutureCrop glasshouses after 5 weeks from sowing.**

## 2.2.1 Characterization of the traits

Glasshouse and field traits were characterized according to descriptors in the book

'Bambara groundnut (*Vigna subterranea* ) (IPGRI, 2000) with some modifications.

Data was recorded for the individual plants at different growth stages and during

harvesting as follows:

### 2.2.1.1 Vegetative characters

*Days to emergence***:** Number of days from sowing to the appearance of first true leaf

on the soil surface.

*Flower no./plant*: Counted each 2-3 days from the first day of flowering for the

duration of study.

***Leaf no./plant***: Recorded at the 4th month after planting as total leaf no/plant in the field experiment. In the glasshouses it was recorded weekly from 5 weeks after emergence for a further three months.

***Days to flowering***: Recorded from seedling emergence to the appearance of the first flower(s).

***Plant height*** **(cm):** Recorded from 10 weeks after sowing and measured from the base of the plant at ground level to the highest point of the terminal leaflet.

***Petiole length*** **(cm):** Recorded as the average length of five leaves of each plant at 10th week after sowing.

***Terminal leaflet length and width*** **(cm)**: Maximum length and width of central leaflet was recorded for 5 leaves per plant at 10 weeks from sowing.

***Leaf area*** **(cm$^2$)**: Estimated based on the central leaflet length and width using the method of Cornelissen *et al.* (2002) in the following equation:

$$A_{plant} = 0.86 * \text{Leaf number} [0.91 * 3 (0.95 * \text{Length} * \text{Width} * \pi /4)]$$

Where leaf number = leaf number/plant; length and width being mean length and width of the terminal leaflet of five leaves/plant, and $\pi = 3.1416$.

***Plant spread*** **(cm):** Greatest distance between two opposite points at the base and top of plant, measured at harvest.

***Stem no./plant***: Recorded at harvest.

***Branch no./stem:*** Average number of branches on three stems/plant, counted at harvest.

***Node no./stem***: Average node number on three stems/plant, recorded at harvest.

***Internode length*** **(mm):** Average length of fourth internodes measured for five longest stems per plant at harvest.

***Growth habit*:** Recorded 10 weeks after sowing for all individual plants, based on the 4th petiole (P)/4th internode (I) length ratio (P/I) as measured in descriptors; Bunch type (P/I = >9), Semi-bunch type (P/I = 7 − 9) and Spreading type (open) (P/I = <7). In our investigation to follow the segregation pattern of growth habit in the segregating $F_3$ we used the following classification: Bunch types were recorded for the DipC parent class (P/I = >13.53), semi-bunch (intermediate) type (8.567< [P/I] <13.53) and spreading type for Tiga necaru parental class (P/I = <8.567). This modification of the standard classification was introduced to allow segregation in comparison to parental types to be scored.

***Peduncle length* (mm):** Average length of five peduncles per plant, measured at harvest.

***Days to podding onset*:** Number of days from seedling emergence to the discovery of the first pod(s) (at least 0.5 cm long).

## 2.2.1.2 Yield evaluation traits

***Pod no./plant*:** Counted at harvest. Number of pods with more than one seed was also determined.

***Pod weight* (g/plant):** Weight of dried pods (at 12% moisture content) was recorded after maintaining the harvest pods for three weeks at 37[o]C.

***Pod length* and width (mm):** Digital Vernier Caliper (model no. OD-15GP, serial no. 211810, Mitutoyo UK Ltd.) was used to measure the greatest length and width of five dried pods containing one seed and five pods containing more than one seed, when available.

***Seed length* and width (mm):** Digital Vernier Caliper was used to measure the greatest length and width of four dried seeds (at 12% moisture content).

***Seed no./plant***: counted after removing the shells of all pods.

***Seed weight*** **(g/plant):** Weight of dried seed (at 12% moisture content).

***Biomass dry weight*** **(g/plant):** The dry weight of all organic materials produced by the plant was measured.

***Shelling percentage (%)*:** measured as an average of all pods/plant, based on the weight of matured dried seeds compared to the weight of dried pods.

***100-seed weight*** **(g):** Recorded after harvest at 12% moisture content.

***Eye pattern around hilum*:** taken as a presence verses absence of *eye pattern around hilum* for the seeds of individual lines in the population under study.

## 2.2.2 Statistical analysis:

Anderson-Darling Normality tests (Stephens, 1974) were used to screen the distribution of the trait data. The inheritance and segregation of contrasting morphological traits were studied in the segregating population derived from the DipC and Tiga necaru landraces.

Chi-Square (Goodness-of-fit test) (Snedecor and Cochran, 1989) was used as an alternative to Anderson-Darling test for trait data that was non-normally distributed.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where:

$O_i$ = observed frequency.

$E_i$ = an expected (theoretical) frequency, based on the null hypothesis.

The Chi-Square value was calculated depending on the expected model for complete and incomplete dominance of the trait distribution which depend on filial generation.

For instance the expected ratio for the phenotype of Mendelian traits is 3:1 and 1:2:1 for complete and incomplete dominance in the second filial generation, respectively. The result was compared to the chi-squared distribution probability to determine the goodness-of-fit at k-1 degrees of freedom.

Box-Cox transformations were applied for the traits showing non-normal distribution and not following Mendelian patterns of segregation, before retesting to determine whether the transformed trait was normally distributed.

Statistical software GenStat 14th Edition and MINITAB (Release 16) were used to analyze the variance, construct residual plots and detect significant association between the traits (p< 0.05).

## 2.3 Construction of the genetic linkage maps

JoinMap4 software (Van Ooijen, 2006) was used to construct linkage maps for the two segregating populations of the narrow and wide crosses. All 73 individuals of segregating $F_3$ population derived from the cross of DipC x Tiga necaru were used as the first mapping population. The population was screened with SSR and DArT markers and polymorphic markers were scored for all individuals to be used in constructing the map. The segregating $F_2$ population of 98 individuals derived from the cross of DipC and VSSP11 was also screened with both marker types; SSR and DArT. Polymorphic markers for this population were used to add to the earlier version of the linkage map based largely on AFLP markers.

### 2.3.1 Preparation of the data files:

Plain text files 'Locus genotype file; .loc' were prepared for categorized segregating marker loci. These were prepared with the text editor program Wordpad and loaded

into JoinMap4 software to be analyzed. The locus genotype file (loc-file) contains the marker segregation patterns within the population. It has a sequential structure, containing four line header instructions. The data body contains the actual genotype information for all individuals grouped according to locus as defined in the manual. The four instructions defined the name of the population, the type of population, the number of loci, and the number of individuals. The syntax of the four instructions is:

name = NAME

popt = POPT

nloc = NLOC

nind = NIND


where NAME was the name given to the population, NLOC and NIND were the numbers of loci and individuals, respectively, and POPT was the code for the population type, which must be one of the codes given in 'Population type codes' in the manual for JoinMap 4.0.

In the segregating $F_3$ population derived from the 'narrow' cross (cultivated landraces) the parental data was not available by the time of running SSR and DArT markers.  Phase determination for DArT markers was carried out by analyzing the population initially as a Cross Pollinator (CP) to determine linkage phase for these markers, using the segregation type code of <hkxhk> (Figure 2-4).

Linkage analysis was carried out (as described below for the $F_3$ dataset) and all markers showing linkage were collected into a new linkage file, for conversion (using the determined phase data) into the new $F_3$ loc file.

```
DArTand SSR.txt - Notepad
File  Edit  Format  View  Help
name = DArT&SSR
popt = cp
nind = 73
nloc = 269
bgPabg-595038    <hkxhk> hh      hh      hh      hh      hh      hh      --      hh      hh      hh
hh      hh      hh      hh      hh      k-      k-      --      hh      hh      k-      hh      hh
hh      hh      hh      hh      k-      hh      hh      hh      hh      hh      hh      hh      hh
k-      k-      hh      k-      hh      hh      hh      hh      hh      --      hh      hh      hh
bgPabg-596655    <hkxhk> hh      --      hh      hh      hh      hh      --      --      hh      hh
hh      hh      hh      hh      hh      --      k-      --      hh      hh      --      hh      hh
hh      hh      hh      hh      k-      hh      --      hh      hh      hh      hh      hh      hh
--      k-      hh      k-      hh      hh      hh      --      hh      --      hh      hh      hh
bgPabg-601086    <hkxhk> hh      hh      --      --      hh      hh      --      hh      hh      --
hh      hh      hh      hh      hh      hh      k-      --      hh      hh      k-      hh      hh
hh      hh      hh      --      hh      hh      hh      --      hh      hh      hh      hh      hh
--      --      hh      k-      hh      hh      hh      hh      hh      --      hh      hh      hh
bgPabg-594508    <hkxhk> hh      hh      hh      hh      hh      hh      --      hh      hh      hh
hh      hh      hh      hh      hh      k-      k-      k-      hh      hh      k-      hh      hh
hh      hh      hh      hh      k-      hh      hh      hh      hh      hh      hh      hh      hh
k-      k-      hh      k-      hh      hh      hh      hh      hh      --      hh      hh      hh
bgPabg-423807    <hkxhk> hh      --      hh      hh      hh      hh      --      hh      hh      hh
hh      hh      hh      hh      hh      k-      k-      --      hh      hh      k-      hh      hh
hh      hh      hh      hh      hh      hh      --      hh      hh      hh      --      hh      hh
--      k-      --      k-      hh      hh      hh      hh      hh      --      hh      hh      --
bgPabg-594180    <hkxhk> --      hh      --      --      --      hh      --      k-      k-      k-
hh      k-      hh      k-      k-      hh      k-      hh      hh      hh      k-      k-      --
hh      k-      hh      --      k-      k-      k-      k-      k-      hh      hh      k-      k-
hh      k-      k-      hh      k-      hh      hh      k-      --      hh      k-      hh
bgPabg-596933    <hkxhk> hh      hh      hh      hh      k-      k-      k-      hh      --      hh
hh      hh      hh      hh      hh      k-      k-      k-      hh      hh      hh      k-      hh      hh
hh      hh      hh      hh      hh      k-      hh      hh      hh      hh      hh      --      hh      hh
k-      k-      hh      k-      hh      hh      hh      --      hh      --      hh      hh      hh
```

**Figure 2-4: The locus genotype file (.txt) used to detect the phases of DArT markers in segregating F$_3$ population derived from 'narrow' cross in bambara groundnut.**

Phase in the linkage groups was used to convert the DArT marker loci to different classifications in an 'RIL3' population type dataset ({0,0} giving a or c, while {1,1} gave b,d). The newly arranged marker locus loc file for the F$_3$ population was prepared again in plain text file to assign linkage groups.

In the wide cross population the F$_2$ population code was selected for use in mapping, as shown in Figure 2-5.

```
wide map.txt - Notepad
File  Edit  Format  View  Help
name=widecrossmap
popt=F2
nind=98
nloc=301
P1b67AAG+CTT1   (b,d)  b      b      d      d      b      d
b         d      d      b      d      b      d      b      d
d         b      b      b      b      d      d      d      b
b         b      b      d      d      b      b      d      b
b         b      b      b      b      d      d      b      d
d
P1b88AAG+CTT2   (b,d)  d      d      d      b      d      d
b         d      d      d      d      d      d      d      d
d         d      d      d      d      d      d      b      d
d         d      d      d      d      d      d      d      d
b         d      d      d      d      d      d      d      d
d
P1b161AAG+CTT3  (a,c)  a      a      a      a      a      c
c         a      a      a      a      c      a      c      c
c         a      a      a      c      c      c      c      a
c         a      c      a      a      c      c      c      c
c         a      c      c      c      a      c      c      a
c
P1b251AAG+CTT4  (a,c)  c      c      a      c      c      c
c         c      c      a      c      a      c      c      a
a         a      c      c      a      c      c      c      c
a         c      c      a      c      c      a      c      c
c         c      c      a      c      c      c      c      c
c
P2b83AAC+CTA1   (a,c)  a      c      c      c      c      c
c         c      c      c      c      a      c      c      c
c         c      a      c      c      c      c      a      a
a         c      a      a      c      c      a      c      c
a         c      c      c      c      c      c      a      c
```

**Figure 2-5: The locus genotype file (.loc) used the phases of DArT markers in segregating F$_2$ population derived from 'wide' cross in bambara groundnut.**

## 2.3.2 Genotyping codes for population types

It is very important that the order of the individuals is identical over all loci in the file.

The genotype code for population types RIL3 and F$_2$ were as follows:

| Code | Description |
|------|-------------|
| a | homozygote as the one parent |
| b | homozygote as the other parent |
| h | Heterozygote |
| c | not genotype a (dominant b-allele) |
| d | not genotype b (dominant a-allele) |
| - | genotype unknown |

Classification type codes at each locus were summarized for all individuals automatically by the software into different genotypic classes with the expected ratio adjusted according to the expected segregation ratios (Table 2-1).

**Table 2-1: Classification codes of F$_2$ and RIL3; the ratios of expected Mendelian segregation**

| Code | Ratio | | Classification into genotype classes |
|------|-------|---|--------------------------------------|
| | F$_2$ population type | RIL3 population type | |
| (a,c) | 1:3 | 3:5 | a and c; h and b will be included in class c |
| (b,d) | 1:3 | 3:5 | b and d; a and h will be included in class d |
| (a,h,b) | 1:2:1 | 3:2:3 | a, h, and b |

### 2.3.3 Linkage analysis of the markers

The linkage map was assembled for both populations, using the JoinMap4 software. The default calculation options were applied, started with calculating the LOD scores and pairwise recombination frequencies between markers. The loci for all individuals in both population types were grouped according to 'independence LOD'. They were manually selected using thresholds from LOD 3-5 and Regression mapping was carried out. The Haldane mapping function was used to convert recombination frequencies into map distances (centimorgan; cM). Segregation distortion was determined through a chi-square test for goodness of fit by comparing the observed ratios to those expected, at the threshold of $p = 0.05$. The third mapping round forces markers which have been grouped, but rejected in the first and second rounds of mapping into the genetic map. The markers those forced to be grouped after first round but showed negative map distances and large *jumps* (threshold value >5) in mean chi-square values were removed from the map.

### 2.4 QTL mapping

Phenotypic data for all agronomical traits for each individual across different environments were combined separately with genotypic data and the linkage map in

order to identify QTLs associated with these traits using MapQTL v.6 software, for each environment (Van Ooijen, 2009).
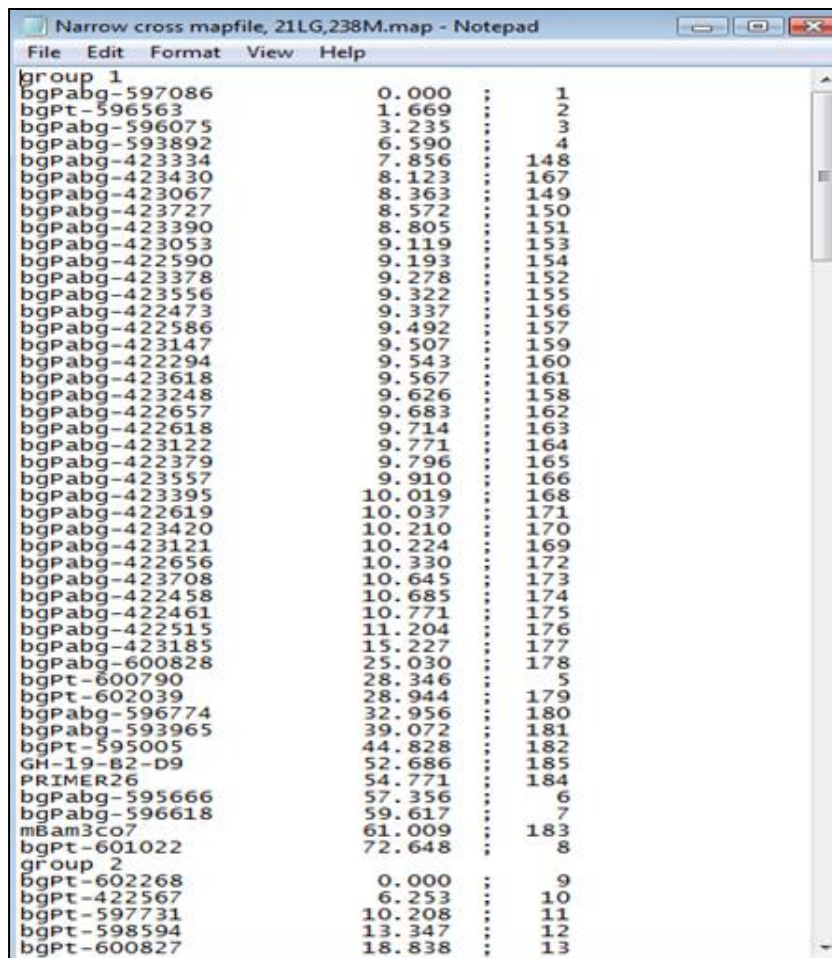
In the narrow cross, QTL analysis was conducted for the means of agronomic data for the segregating $F_3$ population in both glasshouse and field trials. Phenotypic data recorded previously for the segregating $F_2$ population of the same cross (TCRU, 2004) were also used for QTL analysis. In the wide cross QTL analysis was performed for the phenotypic data recorded for the $F_2$ population, analysed using the improved genetic map derived from the wide cross of DipC x VSSP11.

### 2.4.1 Data files for QTL analysis

Three types of data files were prepared for QTL analysis. They were prepared according to the instructions given in the software manual (Van Ooijen, 2009). All three data files were imported into the program and placed in the same directory to avoid confusion. These data files are:

1. *Locus genotype file*: It also called *loc-file*, being the same file as was used for JoinMap 4 software to construct the map. This file contains the genotype codes for the loci of segregating lines.

2. *Map file:* containing the mapped positions of all loci, resulting from grouping and mapping the markers to yield the constructed map. This file will be used by the software to dictate the positions of the markers relative to each other for applications such as Interval Mapping. In non parametric mapping the loci are analysed one by one and the positions of map are only used to sort loci, while in the interval mapping they are used to calculate recombination frequencies, necessary for likelihood calculation.

Map file has no header and starts with the instruction of group number. The loci with their map positions must be given on subsequent lines in ascending order (Figure 2-6).



**Figure 2-6: Part of a map file (.map) used for QTL mapping of narrow cross population.**

3. *Quantitative data file:* (also called *qua-file*) has the data for all quantitative traits for all individuals. The header of the file contains three instructions, followed by the names of the traits. The data body contains the trait information for all individuals. The three instructions define the numbers of traits (ntrt = NTRT), individuals (nind = NIND) and the text that indicated a missing value (miss = MISS), here denoted by '*' (Figure 2-7).

```
Fielddata.qua - Notepad
File  Edit  Format  View  Help
ntrt=29
nind=73
miss=*
Emergence
leafno
plantheig
plantspr
leafletlen
leafletwid
Leafarea
petiolelen
pedunclele
Podweight
Podno.
Podno.2seed
podwidth
podlength
flowering|
stemno
Branchno
Nodeno
seedsweight
seedno
seedlenght
seedwidth
shelling%
100-seedwei
biomass
internode
growthhabit
T-podlenght
T-Podno.
9      32.00    21.9    26.8    5.1     2.37    713.2150446    *        6
*      *        *       *       *       *       0.74     *      4.472135955
6      20.00    15.6    15.5    3.67    1.7     230.0895933    9.33     3
8      7.6      6.4     71.05263158     33.75   8.77     0.3   31.11111111
7      39.00    21.98   34.32   5.92    3.06    1302.745961    13.31    4.54
10.2   8.3      6.17    93.33333333     37.05882353      19.12  1.515714286
6      44.33    23.32   41.18   6.05    2.65    1310.640639    15.37    4.53
8.5    9.01     6.67    57.64499121     48.23529412      19.23  1.621809524
6      27.63    19.7    22.97   5.32    2.13    577.225512     10.46    5.13
5.75   7.94     6.23    45.71917808     38.69565217      14.50  0.652   16.0403
6      38.17    19.93   30.03   5.51    2.14    829.8538236    12.53    4.99
7      7.94     6.31    92.40506329     41.71428571      17.28  1.272   9.84888
6      39.92    18.69   30.45   5.35    2.06    811.3290066    10.61    3.62
8.33   7.11     5.67    86.4594498      32.26666667      14.66  1.425777778
6      28.27    17      20.77   4.32    2.04    459.4450886    9.05     3.09
7.8    8.63     6.54    67.81725888     42.82051282      10.29  0.546666667
```

**Figure 2-7: The trait file (.qua) of the field data used for QTL mapping in the narrow cross population.**

Transformation was used to try to normalize some of the traits which had not been distributed normally, before performing a QTL analysis. The Box-Cox transformation (Box and Cox, 1964) was performed with QI macros SPC Software for Excel.

## 2.4.2 The analysis of QTL

*Nonparametric mapping (Kruskal-Wallis analysis; KW)*

This analysis was done for all normal and non-normally distributed trait data. Interval mapping has an assumption of normally distributed residuals, but KW analysis does not rely on this assumption, so is suitable for both normally and non-normally distributed traits. The Kruskal-Wallis test is regarded as the nonparametric equivalent of the one-way analysis of variance. The test ranks all individuals according to the quantitative trait, while it classifies them according to their marker genotype. A segregating QTL (with big effect) linked closely to the tested marker will result in large differences in average rank of the marker genotype classes. A test statistic based on the ranks in the genotype classes is calculated. For individuals in ties, i.e. several individuals have equal values of the quantitative trait, the average rank (midrank) is used, while for the test the statistic adjusted for ties is used (indicated by K*) (Lehmann, 1975).

The Kruskal-Wallis statistic K* is a Chi-square like distribution, including the degrees of freedom (n-1 df for $F_2$ and $F_3$ populations) and the significance level in asterisks. [Significance level: *:0. 1, **: 0.05, ***: 0.01, ****: 0.005, *****: 0.001, ******: 0.0005, *******: 0.0001]

*Interval mapping*

In interval mapping a so-called QTL likelihood map is calculated. In each position in the genome (every centiMorgan) the likelihood for the presence of a segregating QTL is determined [the likelihood under the alternative hypothesis (*H1*)]. At the same time the genetic effects of the QTL and the residual variances are calculated. Coefficient determination is performed from the model of analysis to estimate the proportion of

phenotypic variance attributed to a particular QTL and the additive effects. This likelihood under *H1* is compared to the likelihood for the situation when a locus with zero genetic effects (no segregation) [the likelihood under the null-hypothesis (*H0*)]. This comparison is done with a likelihood ratio statistic called the LOD (or LOD score), which is the 10-base logarithm of the quotient of the two respective likelihoods. The maximal LOD position was determined as a most likely position of QTL and a one LOD decrease on either side of LOD peak was considered as a confidence interval for the effect. The appropriate threshold was estimated, based on the actual data under study, to detect the significance threshold for interval mapping. This can be determined by performing a Permutation Test using 10,000 reiterations to develop Genome Wide (GW) threshold equivalent to $p \leq 0.05$ (Churchill and Doerge, 1994).

# Chapter 3:   THE DEVELOPMENT OF POLYMORPHIC SSR AND DArT MARKERS FOR MAP CONSTRUCTION AND VALIDATION OF THE F$_3$ POPULATION

## 3.1 Introduction

Molecular markers act effectively as chromosomal landmarks which possess multiple alleles. These sequence variants are distributed in different individuals, so the pattern of inheritance of alleles from parental genotypes can be traced after the construction of a controlled cross between them.   This allows the testing of the patterns of inheritance of these loci and the surrounding chromosomal region that has been co-inherited. Generally, such markers are considered to be phenotypically neutral (Chauhan and Varma, 2009). They are passed on by the standard laws of inheritance through the generations, potentially allowing target allele association with trait inheritance (Semagn *et al.*, 2006b). They have several advantages over traditional phenotypic markers, particularly in terms of access to very large numbers of phenotypically neutral markers which can be used to objectively assess associations between marker and phenotypes. This creates opportunities to improve crop varieties in a shorter time-scale, through the association of marker alleles and trait genes (Farooq and Azam, 2002). There are several types of molecular markers which differ in principle, methodology, and application. Careful consideration is required to choose one or more classes of such markers (Semagn *et al.*, 2006b). In the current study detection and exploitation of DNA polymorphism used both microsatellite and DArT markers.

This chapter includes a discussion of the methodologies of SSR and DArT markers, their application, analysis and interpretation in the 'narrow' and 'wide' cross populations. As a starting point for mapping, validation of the $F_3$ population under study was carried out, both in terms of legitimacy and also to confirm the generation of inbreeding of the lines. As an early application relevant to breeding programmes, the isolation of DNA from dried bambara groundnut seed were also investigated with a view to rapid fingerprinting of putative hybrids in the dried seed state.

### 3.1.1 Microsatellite markers

A microsatellite contains many copies of a short tandem repeat sequence. These sites are usually in the size range of 16 to 60 bases in total, composed of individual repeat units of up to six base pair repeats in length. The sequence repeat unit and overall repeat distribution of different SSR motifs has been investigated in many organisms (Trivedi, 2004). In plants, the average distance between SSRs was found to be 6-7 kb which is similar to reports in mammals (Cardle *et al.*, 2000). Microsatellites are often very polymorphic, scattered throughout genomes, usually located in the same place on the genome across individuals within a species, but often with a difference in the number of repeat sequences (Kalia *et al.,* 2011; Trivedi, 2004).

### Amplification of SSR

Microsatellites can be identified by amplification with the polymerase chain reaction, using primers designed to unique sequences flanking the repeat region. As these markers are PCR-based, only small amounts of DNA are needed for amplification and polymorphism can be detected in a number of ways to reveal size differences between alleles (Hammock and Young, 2005).

**Limitations of SSR markers**

Microsatellites have proven to be adaptable molecular markers, particularly for population analysis, but have some limitations. Microsatellites developed for a particular species can often be applied to closely related species, but the percentage of amplified loci may decrease when the genetic distance between design and tested species increases (Jarne and Lagoda, 1996). PCR failure may result when a particular allele at a locus fails to amplify, whereas others alleles may amplify more efficiently and may appear homozygous on a gel assay, when they are actually heterozygous in the genome (Gholizadeh and Mianji, 2007). Such 'null' alleles are found in most taxa and complicate the interpretation of microsatellite allele frequencies. These are considered a technical limitation which might lead to high frequencies of false homozygotes and cause problems for diversity or parentage analysis (Dakin and Avise, 2004; Reece *et al.*, 2004). Large allele dropout and slip-strand mispairing for microsatellites during PCR are considered another problem that can cause reduced signal strength or increased stutter (Van Oosterhout *et al.*, 2004). Stutter bands are artifacts having less signal intensity and usually they are shorter than the actual SSRs. They are thought to arise due to the inherent instability of the repeat sequence in the PCR reaction, reflecting their instability *in vivo* which leads to their high mutation rates. These bands have to be taken into consideration during microsatellite scoring, as improper scoring of the bands will lead to incorrect results during data analysis (Wang *et al.*, 2009).

Alleles containing more repeat sequence bases are more likely to be mis-replicated as the long alleles have more chance of recombination instability during meiosis (Hancock and Simon, 2005).

**3.1.2 Diversity array technology markers (DArT)**

Due to a number of limitations for other markers types, such as; time, cost and resources required to discover a large number of markers for a species, the DArT system has been developed to overcome some of these limitations. DArTs have the advantage of being highly locus specific, due to their detection by hybridization (Jaccoud *et al.*, 2001) but unlike SSRs they are considered dominant, therefore unable to differentiate between homozygous and heterozygous loci. Although, SSR markers have an advantage over DArT markers as they are co-dominant, highly polymorphic and widely distributed in the genome (Yang *et al.*, 2006), the requirement of substantial sequence information to generate SSRs makes DArT more applicable, especially in underutilized crops where no sequence information exists, as is the case for bambara groundnut.

Diversity Arrays Technology offers a low-cost per marker high-throughput, robust system with minimal DNA requirements, capable of providing comprehensive genome coverage even in organisms with limited or no DNA sequence information. The "genotyping array" is hybridized with an individual DNA sample using slide microarray technology; the resulting hybridization signal is measured and converted into a genotype score for 'presence' or 'absence'. Such arrays have been developed by Dr. Kilian and his team for about seventy species, includes; sorghum, rice, barley, wheat, chickpea, pigeon pea and many others including animal species (www.diversityarrays.com; as at July 2012).

DArT uses an array of individual clones from a genomic representation prepared by amplification of fragments between 'rare' cutting restriction fragments (often *Pst*I) using adaptors to the cut sites. A second frequent cutting enzyme reduces the complexity of the pool and the number of amplifying *Pst*I-*Pst*I fragments. Labeled

genomic representations of the individuals to be genotyped prepared in exactly the same way are then hybridized onto individual arrays. The polymorphisms scored are the presence versus absence of hybridization to individual array elements. The platform allows high-throughput screening of tens of thousands of potential molecular markers in parallel, and it especially suited for the generation of genome-wide markers for genetic linkage mapping.

## 3.2 Results

A total of 37 primer pairs were designed in this study and used as additional microsatellite markers to those developed previously; here referred to as 'the second set'(as described in Chapter 2). The pool of PCR fragments enriched for simple sequence repeats was generated from a bambara groundnut microsatellite-enrich library using a 1/16th 454 Pyrosequencing run with Titanium reagents. A total of 5443 sequences were produced giving a total length of 1,697,965bp. As a result, 261 sequences were found to be unique and had enough sequence flanking the SSR repeat motif to design the primer pair. These unique sequences include the SSRs from 2-5 unit size, having allelic sizes from 10-60 bp for repeated unit and 19 of them were with compound format of SSR. The results of the microsatellite motifs search, using the MISA.pl script, and the distribution of repeated type for library fragments are shown in Table 3-1.

**Table 3-1: Result of microsatellite-enriched library and the distribution of repeated type.**

| Results of microsatellite search | | | | | | |
|---|---|---|---|---|---|---|
| Total number of sequences examined | | | | 5443 | | |
| Total size of examined sequences (bp) | | | | 1697965 | | |
| Total number of identified SSRs | | | | 1559 | | |
| Number of SSR containing sequences | | | | 1290 | | |
| Number of sequences containing more than 1 SSR | | | | 145 | | |
| Number of SSRs present in compound formation | | | | 227 | | |
| Distribution of different repeat type classes | | | | | | |
| Unit size | 1 | 2 | 3 | 4 | 5 | 6 |
| Number of SSRs | 68 | 740 | 669 | 58 | 21 | 3 |

Thirty seven primer pairs were designed in this study from a total of 32 fragments out of 261 unique sequences. These primer sets were used along with the other two sets of primers designed previously to give a total number of 124 primer pairs. These were test for segregation in the both 'wide' and 'narrow' crosses of bambara groundnut.

### 3.2.1 Microsatellite markers applied to the F₃ population of the 'narrow' cross

### 3.2.1.1 Primer optimization

The amplification was performed with a Thermo Hybaid Express PCR machine to optimize annealing gradient temperatures for all the primers. The best annealing temperature was identified for each primer based on the intensity of the amplified bands from 12 different temperatures along the gradient. To detect polymorphic alleles all individuals were amplified with the primers at the best annealing temperature using the M13-Tag Forward primer in a three primer reaction to allow fluorescent labelling.

In the first primer set all markers amplified as expected except Primer81, while bam2coL58 gave a weak band with fragment size larger than expected (more than 500bp). Generally, most of the primers amplified with a wide range of gradient temperature (Figure 3-1).

In the validation of the second primer set with bambara groundnut genomic DNA, the product sizes in general were less than 300 bp, however the fragments amplified with some of the primers were not in agreement with their predicted sizes. Based on original sequences the product sizes of primers BN11 and BN21 were expected to be 152 and 114 bp, while the product sizes for both showed fragments more than 400 and

500bp on agarose gel, respectively (Figure 3-2) and they eliminated from further test. In total the optimal annealing temperature was identified for 22 out of 37 primers as the rest were not amplified well or they had unexpected product sizes.
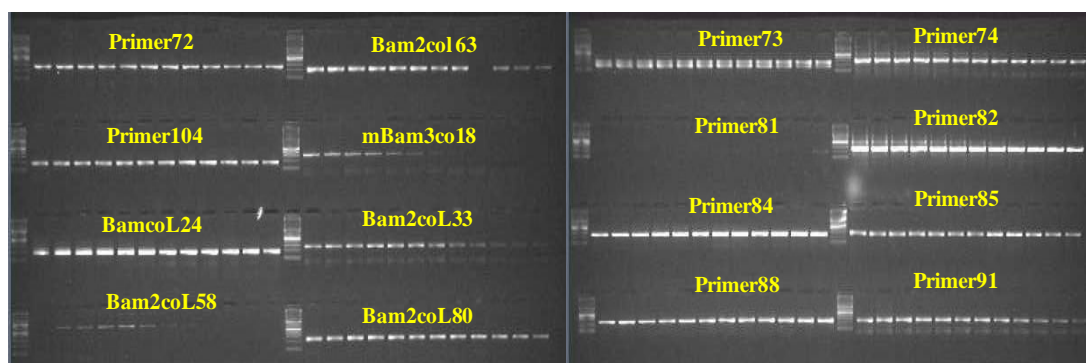


**Figure 3-1: Annealing gradient evaluation of a range of primers from the bambara groundnut microsatellite enriched-library, showing amplification across a wide range of temperatures (50-65°C) with different expected fragment sizes (168bp-294) and unexpected (incorrect) fragment size (>500bp) for primer bam2coL58.**
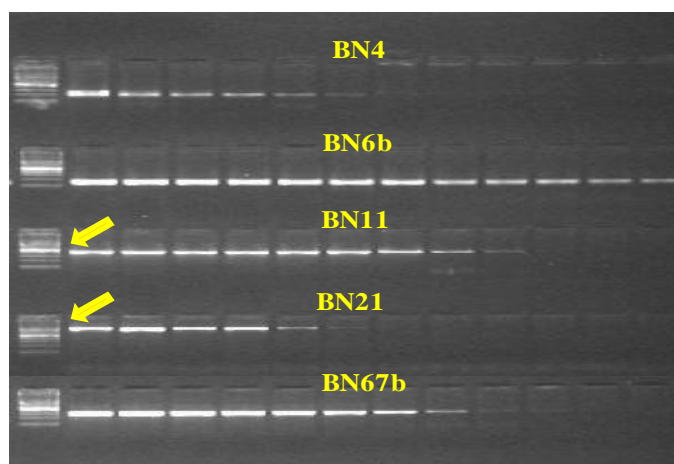


**Figure 3-2: Annealing gradient evaluation for a number of the second primer set of bambara groundnut genomic enrich-library designed using Primer3 from sequence generated with Titanium reagents. The temperature gradient ranges from 50°C to 65°C and uses mixed genomic DNAs representative of bambara groundnut germplasm. An expected fragment sizes of around 400bp and 500bp were amplified for both primers BN11 and BN21, respectively, which they were too big to be use in capillary analysis.**

All 29 primers developed from bambara groundnut transcriptome amplified products of the expected size in an annealing gradient test with bambara groundnut genomic DNA. They were then used to amplify individual lines at the best optimized temperature to screen for allelic size differences.

### 3.2.1.2 Detection of polymorphism

All primer sets with identified optimal annealing temperatures were used to screen for polymorphism within the F$_3$ segregating population. Each primer pair was run against 8 lines of the cross, initially. When polymorphism was detected for primers, the remaining lines were included to generate a full dataset for the mapping. The correct size of different alleles were determined with the CEQ$^{TM}$ 8000 Fragment Analysis Software Version 8 (Figure 3-3) and data scored manually for segregation.
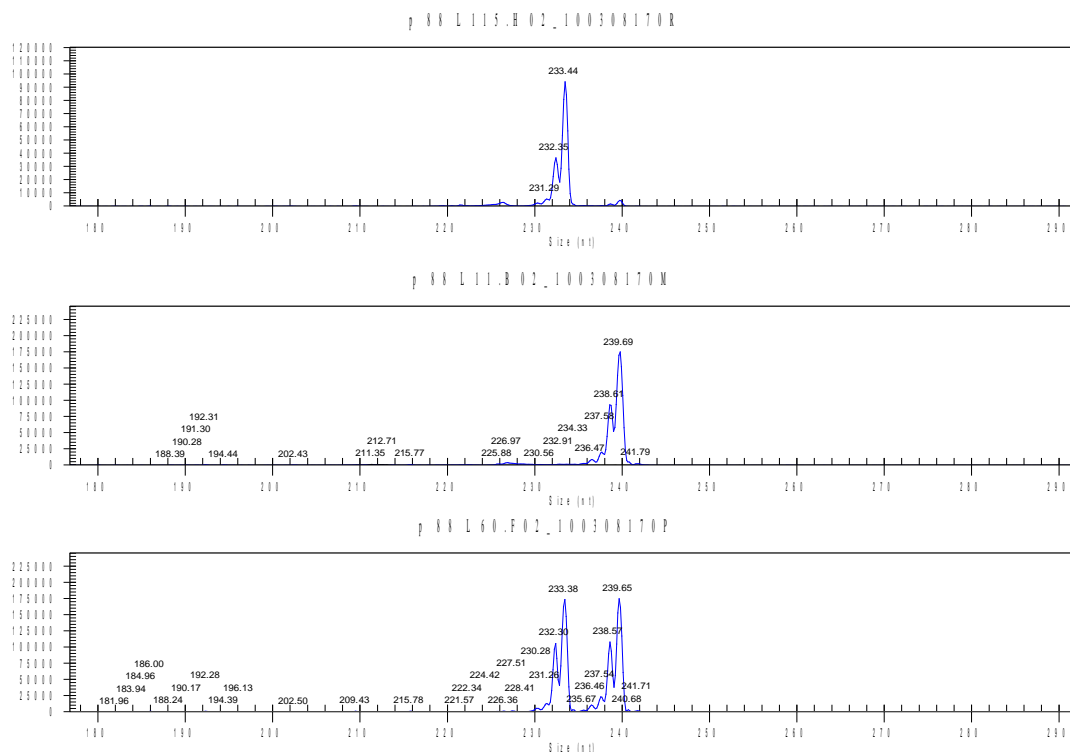


**Figure 3-3: Capillary result for the polymorphic Primer 88 showing different fragment lengths of parental alleles (DipC and Tiga necaru) within selected lines of cross. Line 11 and 115 contained one parental allele each, while line 60 has both parental alleles.**

The parents of $F_3$ population were confirmed by genotyping of available seed pools by the time of developing primers from the RNA deep sequencing leaf transcriptome for bambara groundnut. All 29 primers were run directly against the parents of cross (DipC and Tiga necaru). The PCR products were run on capillary electrophoresis and different fragment lengths for parental alleles were identified for 15 primer pairs (Table 3-2).

**Table 3-2: Polymorphic microsatellite primers derived from the RNA sequence data in the $F_3$ population of 'narrow' cross**

| No. | Primer name | Optimal annealing temperature $^oC$ | 'narrow' cross | |
|---|---|---|---|---|
| | | | DipC | Tiga necaru |
| 1 | D.5953 | 50 | 315 | 297 |
| 2 | D.25551 | 60 | 200 | 198 |
| 3 | D.42026 | 60 | 238 | 229 |
| 4 | D.8148 | 65 | 244 | 229 |
| 5 | D.8999 | 55 | 203 | 205 |
| 6 | D.37053 | 55 | 181 | 178 |
| 7 | D.12522 | 60 | 328 | 324 |
| 8 | D.14265 | 55 | 182 | 176 |
| 9 | D.16501 | 50 | 255 | 267 |
| 10 | D.24269 | 60 | 246 | 238 |
| 11 | D.35497 | 55 | 168 | 202 |
| 12 | D.51646 | 60 | 185 | 189 |
| 13 | D.7215 | 62 | 208 | 202 |
| 14 | D.125 | 60 | 277 | 273 |
| 15 | D.2094 | 60 | 224 | 227 |

Out of these 15 polymorphic primers, one of them (D.35497) was screened by agarose gel electrophoresis. A 2.5% agarose gel in 0.5 X TBE (500ml tray size) was used (adding 3µL of ethidium bromide stock; 10mg/mL) and the gel was run for 3 hours on 90Volts (Figure 3-4).
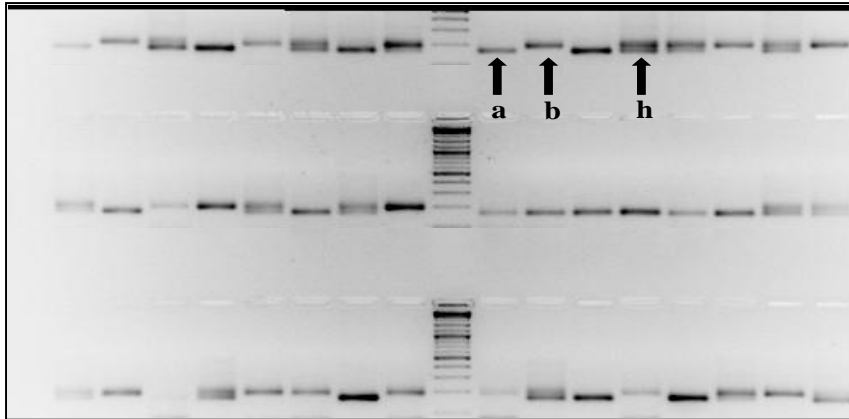
**Figure 3-4: Reverse gel plate image of PCR products amplified by primer D.35497 (derived from the leaf transcriptome library) with a selection of the individual F$_3$ lines of the 'narrow' cross. (a) allele from DipC parent, (b) allele from Tiga necaru and (h) for a heterozygote line within the cross. Fragments are visualised on a 2.5% agarose gel run on 90 Volt for 3 hours.**

### 3.2.1.3 Validation of the F$_3$ population

All 73 individual lines of the F$_3$ population derived from the cross of DipC x Tiga necaru were screened with the 33 polymorphic SSR markers derived from these three sources of SSRs. Heterozygote alleles were scored across all 73 individual lines by each marker and the average level of residual heterozygosity in this population was identified to be 24.86% (Table 3-3). This result shows that this population is effectively an F$_3$ population, as expected from planting history. In addition, no lines which were not consistent with the expected origin were identified.

### 3.2.2 DArT markers applied to the F$_3$ population 'narrow' cross

The slide microarray was developed as previously described (Mayes *et al.*, 2009; Olukolu *et al.*, 2012; Stadler, 2009) . Of the 7680 fragments detected in DArT array, 236 (3.1%) were assigned with at least one different score within the population and so identified as putative polymorphic markers in the 'narrow' cross population derived from the cross of DipC x Tiga necaru landraces. Later the DArT marker data

**Table 3-3: Primers used in the initial screening of the segregating F₃ population, showing optimal annealing temperature, heterozygosity and parental alleles.**

| Name | Forward sequence (5′-3′) | Length | Reverse sequence (3′-5′) | Length | Optimal annealing temperature (°C) | Hyterozygosity % | Fragment length (bp) Parent 1 | Parent 2 |
|---|---|---|---|---|---|---|---|---|
| PRIMER2 | CGTGGATACCCATACCGTCT | 20 | TAAGTCCATTTTGTCCGATTGA | 22 | 51 | 34.25 | 171 | 173 |
| PRIMER7 | GTAGGCCCAACACCACAGTT | 20 | GGAGGTTGATCGATGGAAAA | 20 | 54 | 32.88 | 210 | 212 |
| PRIMER10 | TCAGTGCTTCAACCATCAGC | 20 | GACCAAACCATTGCCAAACT | 20 | 54 | 27.40 | 260 | 234 |
| PRIMER15 | AGGAGCAGAAGCTGAAGCAG | 20 | CCAATGCTTTTGAACCAACA | 20 | 58 | 30.14 | 238 | 212 |
| PRIMER16 | CCGGAACAGAAAACAACAAC | 20 | CGTCGATGACAAAGAGCTTG | 20 | 55 | 16.9 | 189 | 187 |
| PRIMER19 | AGGCAAAAACGTTTCAGTTC | 20 | TTCATGAAGGTTGAGTTTGTCA | 22 | 57 | 28.77 | 273 | 235 |
| PRIMER26 | CGCTCATTTTAACCAGACCTC | 21 | CAAACAAACCAACGGAATGA | 20 | 55 | 20.83 | 183 | 185 |
| PRIMER32 | TTCACCTGAACCCCTTAACC | 20 | AGGCTTCACTCACGGGTATG | 20 | 55 | 19.18 | 247 | 251 |
| PRIMER37 | CCGATGGACGGGTAGATATG | 20 | GCAACCCTCTTTTTCTGCAC | 20 | 60 | 35.62 | 258 | 260 |
| PRIMER38 | TCACACTTGCAATGGTGCTT | 20 | TCGTTGTTTCTCTTTTCATTGC | 22 | 57 | 31.51 | 194 | 191 |
| PRIMER43 | CTTGATGCTACCGAGAGAGAG | 21 | AGGCTCCAACAATGCGATAG | 20 | 55 | 20.55 | 199 | 205 |
| PRIMER45 | CGTGGATACCCATACCGTCT | 20 | AAGTCCATTTTGTCCGATTGA | 21 | 52 | 36.99 | 171 | 173 |
| PRIMER48 | TACCTGCATTCGGGACAGTT | 20 | TTCACTCTTTCTTGATCACATGC | 23 | 60 | 20.55 | 238 | 230 |
| PRIMER65 | GGACGTGAATCGATGGAGAT | 20 | TCCTTCCCCCTTCTCTGATT | 20 | 55 | 35.62 | 172 | 176 |
| PRIMER66 | CGTTAGATCTGAGACGCCATT | 21 | CATCCATCACCTGTCACCAG | 20 | 60 | 27.40 | 225 | 213 |
| PRIMER85 | TTTCCAGATTGGATCGTTGA | 20 | TGTCTTCACACCGGAATTTG | 20 | 58 | 22.54 | 248 | 252 |
| PRIMER88 | TGTGGTTGTGCTCCTTCTCA | 20 | GGGAAGAAGAGTGAAGTTGGAA | 22 | 62 | 28.77 | 233 | 239 |
| PRIMER95 | AAGTCCATTTTGTCCGATTGA | 21 | CGTGGATACCCATACCGTCT | 20 | 58 | 31.51 | 168 | 170 |
| PRIMER98 | TTTTGTCACTGTTTGCCACAA | 21 | AGATTTATATCTGGATGAGAGAGAGAG | 27 | 57 | 15.07 | 264 | 294 |
| PRIMER103 | AAATTCAAAGGCCTGGAAAAA | 21 | TTTTTGAGTTCTGCGAGCAA | 20 | 57 | 26.76 | 210 | 220 |
| GH-19-B2-D9 | ATCAAAATCAAGCAAATGAGA | 21 | ACCTTTTACGCTCATTTTAACCAG | 24 | 50 | 26.03 | 236 | 238 |
| BamcoL17 | AACCTGAGAGAAGCGCGTAGAGAA | 24 | GGCTCCCTTCTAAGCAGCAGAACT | 24 | 58 | 28.77 | 162 | 166 |
| Bam2coL33 | ATGTTCCTTCGTCCTTTTCTCAGC | 24 | AAAACAATCTCTGCCCCAAAAAGA | 24 | 54 | 20.55 | 253 | 255 |
| Bam2coL63 | AAAATCTCACTCGGATGGCATGTG | 24 | TGGAATCACCTGATAGTAGTGTATTGG | 27 | 55 | 23.29 | 293 | 295 |
| Bam2coL80 | GAGTCCAATAACTGCTCCCGTTTG | 24 | ACGGCAAGCCCTAACTCTTCATTT | 24 | 58 | 13.70 | 220 | 224 |
| mBam3co7 | GGGTTAGTGATAATAAATGGGTGTG | 25 | GTCATAGGAAAGGACCAGTTTCTC | 24 | 59 | 28.77 | 267 | 275 |
| mBam3co33 | TGTGTCTGTTTGTGGGGATATGTA | 24 | TTATCCCGGTCCTAATTCATCTTA | 24 | 58 | 21.92 | 295 | 319 |
| AG81 | ATTTTCCAACTCGAATTGACC | 21 | TCATCAATCTCGACAAAGAATG | 22 | 52 | 10.96 | 202 | 190 |
| BN 6b | CACTACCCTGTTCTTCATCCGT | 22 | CATTGCACGTCATAGAATTTGG | 22 | 53 | 26.03 | 146 | 150 |
| BN 145 | GGCACTGGTAGCAACGAAA | 19 | CGTGGACGTAACAACACAACAC | 22 | 50 | 13.70 | 150 | 154 |
| BN 259 | CGATTGCACGTCATAGAATTTG | 22 | GTTCCAGACACTACCCTCGTTC | 22 | 50 | 21.92 | 159 | 163 |
| D.24269 | AGGTTCATGATCGTAGATGTGGAT | 24 | ACGATaTCATACTGACaTgtTTCATAC | 27 | 60 | 17.81 | 246 | 238 |
| D.35497 | ACTTTTAGCTCTTGTCAGGAAACG | 24 | TCTTTCTACTTTTCTCTGGCTGGT | 24 | 55 | 23.61 | 168 | 202 |

was combined with the SSR marker data to construct the genetic linkage map. As the parental seed sources were not confirmed by the time of the preparation of DNA representations from the individual lines, the data for the DArT markers had only two genotypic classes, present and absent, without knowing which parent was transmitting the absence of the band (the fully informative state). For this reason, the coding for unknown phase, two genotypic classes, was used (k-, hh) (Figure 3-5).

| DArT marker | Individual lines | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| bgPabg-593965 | hh | k- | k- | k- | hh | hh | hh | hh | hh | k- | k- | k- | hh | hh | k- |
| bgPabg-596335 | k- | hh | k- | k- | k- | hh | k- | k- | k- | k- | hh | hh | hh | hh | k- |
| bgPabg-596774 | hh | k- | hh | k- | hh | hh | hh | hh | -- | k- | k- | k- | hh | k- | k- |
| bgPabg-597624 | hh | hh | k- | k- | k- | hh | k- | k- | k- | k- | hh | k- | k- | k- | hh |
| bgPabg-597858 | hh | hh | k- | k- | k- | k- | k- | hh | hh | hh | k- | hh | k- | k- | k- |
| bgPabg-594335 | k- | k- | hh | hh | k- | hh | k- | k- | k- | k- | hh | k- | hh | k- | k- |
| bgPabg-594261 | k- | k- | -- | k- | k- | hh | k- | k- | hh | k- | k- | hh | k- | k- | -- |
| bgPabg-595641 | hh | hh | k- | hh | hh | -- | hh | hh | hh | hh | k- | k- | hh | k- | k- |
| bgPabg-595315 | k- | k- | hh | hh | hh | k- | k- | hh | k- | hh | hh | k- | k- | hh | k- |
| bgPabg-595273 | k- | k- | hh | hh | hh | k- | k- | hh | k- | hh | hh | k- | k- | hh | k- |
| bgPabg-594877 | k- | k- | k- | k- | k- | k- | k- | k- | hh | hh | k- | hh | k- | k- | k- |
| bgPabg-597436 | k- | hh | k- | hh | k- | hh | k- | k- | k- | k- | hh | -- | hh | hh | k- |
| bgPabg-593983 | hh | k- | hh | hh | hh | hh | hh | k- | k- | k- | hh | k- | hh | hh | hh |
| bgPabg-594142 | k- | -- | k- | hh | k- | hh | k- | k- | k- | k- | hh | k- | hh | hh | k- |

**Figure 3-5: Allelic differences at loci for different DArT markers among in a selection of the individual F₃ plants derived from the narrow cross of DipC x Tiga necaru. Alleles were scored depending on the presence and the absence of the allele, with the absence being fully informative and presence unable to distinguish between a homozygous or heterozygous band, due to the dominant nature of the hybridization-based DArT array marker system.**

### 3.2.3 Microsatellite markers scored in the F₂ population of the 'wide' cross

The population lines and their parents were extracted from dried leaves. The gel picture of extracted DNA showed a smear, suggesting relatively low DNA quality for

the parents and the individual $F_2$ plants (Figure 3-6), as might be expected from old silica gel dried leaf material. These DNA templates could be amplified with different primers in PCR using greater numbers of cycles (37-38 cycles). Different strengths of amplification were produced depending on the primer pair, but they all had enough PCR products to allow visualization on agarose gel and size calling on the Beckman CEQ$^{TM}$ 8000 Fragments Analyzer (Figure 3-7)
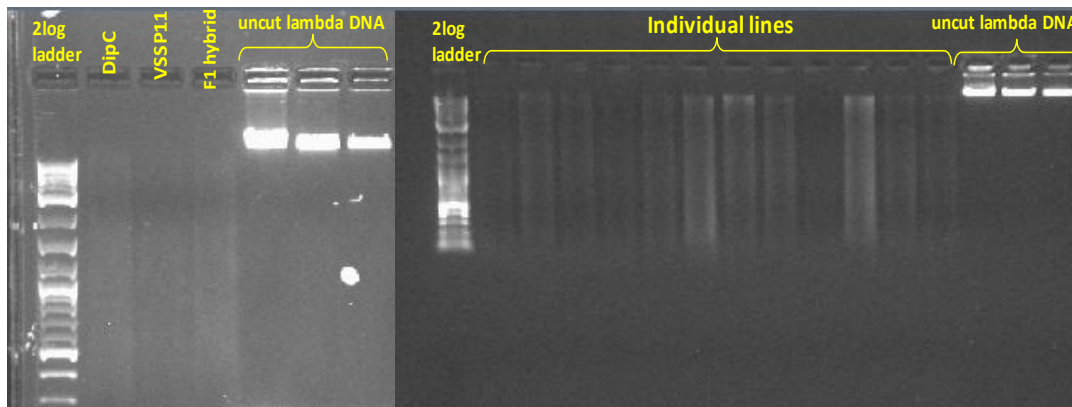


**Figure 3-6: Gel picture of extracted DNA from silica dried leaf tissue of parents (DipC and VSSP11), their $F_1$ hybrid and selected individuals of the $F_2$ population for the 'wide' cross, showing poor DNA quality and quantity.**
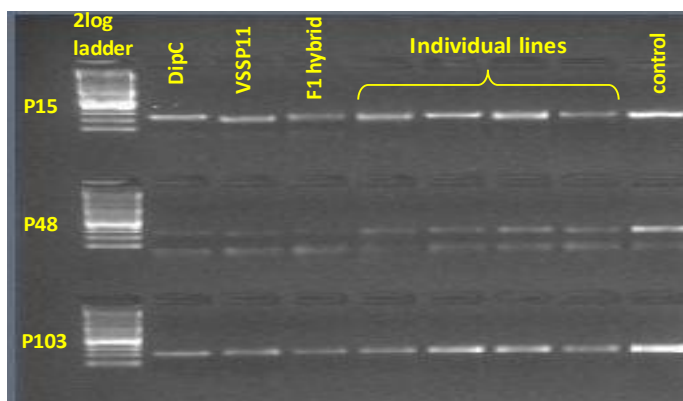


**Figure 3-7: Amplification of parents and selected individuals of the wide cross population for a number of different primer pairs (Primer15, primer48 and primer103) compared with amplified DNAs extracted from DNA derived from a single fresh leaf sample (control; narrow cross individual).**

This population was also screened with the same primer sets used for the 'narrow' cross population to identify polymorphic markers. The aim of this application was to further develop the initial map of the 'wide' cross population with more markers (in addition to the current AFLP markers) and then compare it with the 'narrow' cross population through a number of common markers. Out of 95 SSR markers established from the first and second sets genomic library of bambara groundnut and soybean , 22 markers were polymorphic when amplified in the parents of the cross (DipC and VSSP11) (Table 3-4). Among these polymorphic primers four of them did not show polymorphism when applied to 'narrow' cross population. All of these polymorphic primers except Primer81 and Primer96 were screened with the individual lines of $F_2$ population.

**Table 3-4: Polymorphic primers for the wide cross population recording the different allelic size for the parents of the cross.**

| No. | Primer name | Optimal annealing temperature $^o$C | 'wide' cross | |
|---|---|---|---|---|
| | | | DipC | VSSP11 |
| 1 | PRIMER10 | 53 | 268 | 237 |
| 2 | PRIMER15 | 59 | 240 | 212 |
| 3 | PRIMER16 | 55 | 189 | 185 |
| 4 | PRIMER19 | 59 | 270 | 244 |
| 5 | PRIMER26 | 55 | 183 | 189 |
| 6 | PRIMER32 | 50 | 220 | 218 |
| 7 | PRIMER38 | 56 | 194 | 191 |
| 8 | PRIMER48 | 55 | 244 | 228 |
| 9 | PRIMER65 | 55 | 172 | 176 |
| 10 | PRIMER66 | 55 | 219 | 225 |
| 11 | PRIMER73[a] | 56 | 236 | 234 |
| 12 | **PRIMER81**[a] | 53 | 189 | 185 |
| 13 | PRIMER82[a] | 53 | 270 | 242 |
| 14 | PRIMER85 | 55 | 248 | 240 |
| 15 | **PRIMER96**[a] | 53 | 182 | 184 |
| 16 | PRIMER98 | 59 | 274 | 276 |
| 17 | GH-19-B2-D9 | 50 | 236 | 242 |
| 18 | Bam2coL80 | 55 | 220 | 218 |
| 19 | Bm2coL33 | 50 | 239 | 249 |
| 20 | mBam3co7 | 55 | 267 | 271 |
| 21 | AG81[b] | 52 | 202 | 190 |
| 22 | BN145 | 50 | 143 | 147 |

a: polymorphic only in the wide cross population

b: primer from soybean

Amplified fragments of the individual $F_2$ plants by primers Primer10 and Primer82 could be scored for allelic sized differences using agarose gel electrophoresis. A 2.5% agarose gel in 0.5 X TBE (500ml tray size) was made adding 3µL of ethidium bromide stock (10mg/mL) and the gel was run for 3 hours on 90 Volts (Figure 3-8).
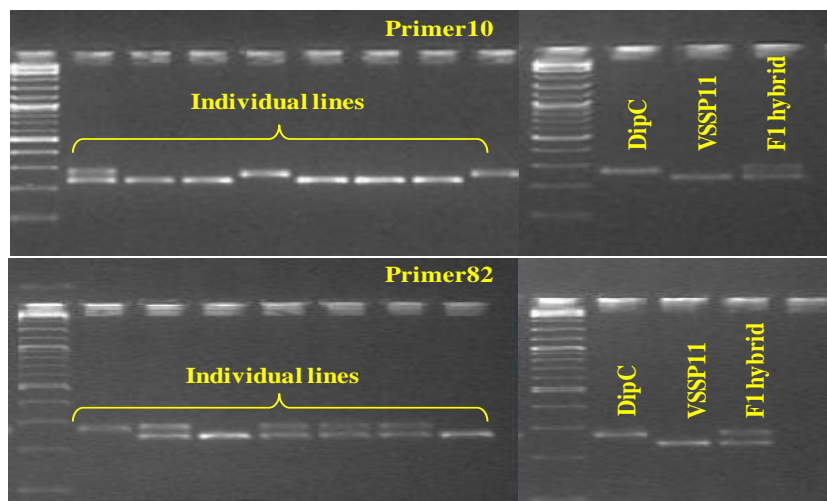


**Figure 3-8: Scoring the SSR markers in lines of wide cross population ($F_2$) for parental alleles using agarose gel 2.5% for primer10 and primer82. Two different fragment sizes for both parents in the $F_1$ hybrid and a selection of lines from the "wide" cross population were distinguishable on this system.**

Scoring the $F_2$ population with the other identified polymorphic primers was conducted through running the PCR products on the Beckman CEQ™ 8000 capillary electrophoresis machine. The electrophoretograms of the results were analyzed with the CEQ system software and the product sizes called manually from the traces (Figure 3-9).
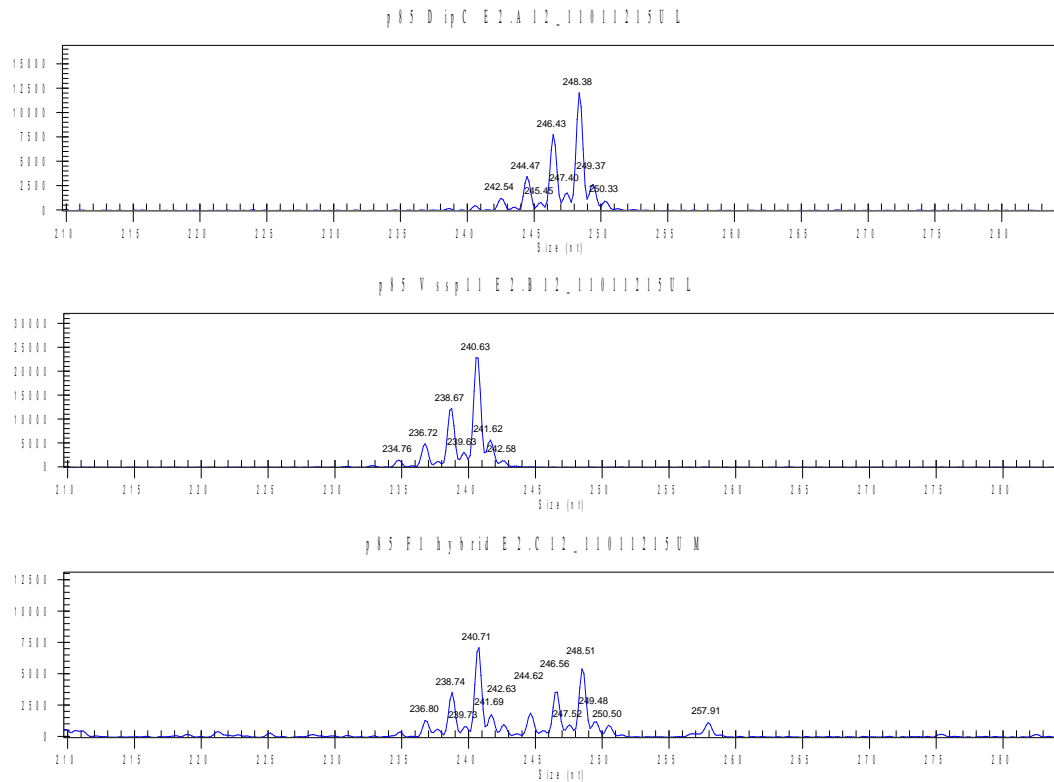
**Figure 3-9: Electropherogram for polymorphic Primer85 using the Beckman CEQ$^{TM}$ 8000 capillary electrophoresis machine. Different alleles can be seen at 248bp for DipC, at 240bp for VSSP11 and the presence of both parental alleles in the F$_1$.**

Primers developed from the bambara groundnut transcriptome were run directly against the parents (DipC and VSSP11) of the F$_2$ segregation population. Out of 29 primers 19 of them showed polymorphism when amplified from the parents and F$_1$ hybrid. Only 3 primers (D.32936, D.48339 and D.51646) were used to screen the whole population of individual lines of 'wide' cross population (Table 3-5) due to cost constraints in this part of the work. The population was screened with the second primer (D.48339) using agarose gel electrophoresis. Additional primers could be added when funding allows to increase the density of SSR markers in the wide cross.

**Table 3-5: Polymorphic microsatellite primers derived from the RNA sequence data for the parents of 'wide' cross.**

| No. | Primer name | Optimal annealing temperature °C | 'wide' cross | |
|---|---|---|---|---|
| | | | DipC | VSSP11 |
| 1 | D.15508 | 62 | 307 | 311 |
| 2 | D.36186 | 62 | 172 | 163 |
| 3 | D.42026 | 60 | 238 | 229 |
| 4 | D.8148 | 65 | 244 | 229 |
| 5 | D.8999 | 55 | 203 | 195 |
| 6 | D.37053 | 55 | 181 | 190 |
| 7 | D.12522 | 60 | 328 | 326 |
| 8 | D.15619 | 55 | 221 | 217 |
| 9 | D. 21310 | 50 | 178 | 186 |
| 10 | D.24269 | 60 | 246 | 250 |
| 11 | **D.32937*** | 60 | 291 | 305 |
| 12 | D.35497 | 55 | 168 | 166 |
| 13 | **D.48339*** | 60 | 197 | 161 |
| 14 | **D.51646*** | 60 | 185 | 195 |
| 15 | D.1006 | 60 | 161 | 158 |
| 16 | D.7215 | 62 | 208 | 202 |
| 17 | D.8387 | 55 | 187 | 190 |
| 18 | D.11860 | 55 | 271 | 283 |
| 19 | D.2094 | 60 | 224 | 227 |

* : primer used to screen the whole population individuals

### 3.2.4 Genotyping errors in SSR analysis

During the course of the present study a number of issues were encountered with the results of Beckman Coulter CEQ$^{TM}$ 8000. These problems were:

- Stutter bands: One of the potential genotyping errors in analyzing the electropherogram data of microsatellites is the production of stutter bands. In primer48, heterozygosity was difficult to score in line 106 for the $F_3$ derived from the cross DipC x Tiga necaru (Figure 3-10), but it became clear when the peaks of both alleles were compared. The results of parental alleles enabled a clear scoring of line 106 as heterozygous.
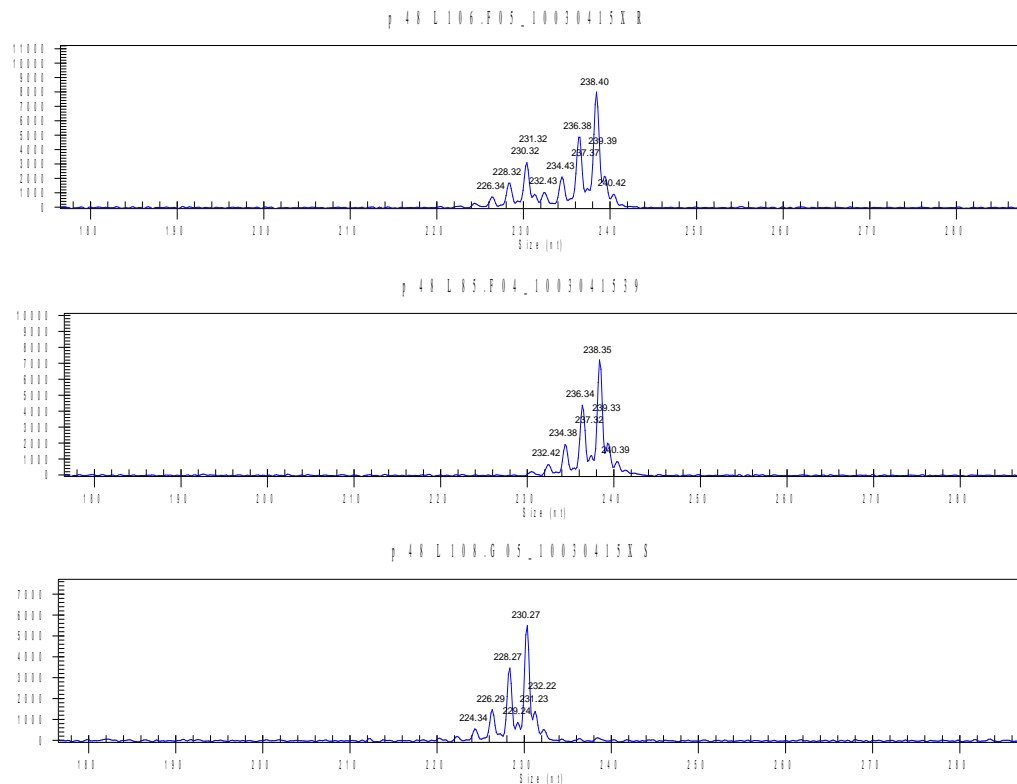
**Figure 3-10: Capillary electrophoresis showing a potential scoring error to detect the both parental alleles in the heterozygote sample of Primer48 L 106.F05 due to the effects of stutter bands and overlap. While both alleles showed normal band separately in different homozygote lines of L85 and L108.**

- Allele size calling: Allele sizes should be a whole number of bases, but as the mass of bases is not identical, there is some degree of genuine mass difference between sequences which may be the same length, but which have different base compositions. In addition, the capillary electrophoresis also can experience technical issues relating to operating temperatures between runs, quality of reagents and age of gel matrix which can cause minor shifts between runs or even within runs, if the calibration ladder is poor. These non-integer values can lead to mistyping errors due to rounding off of alleles during binning. As there were only 2 different sizes of parental alleles within the individuals for a particular SSR product, sample peak sizes was recorded

to the nearest whole number and it is not believed that this will be a major source of mistyping, for this application.

- Noisy peak shape: some samples had a high background noise which made scoring more difficult, especially when allele size differences between parents are small. Sample (a) of Figure 3-11 is line 56 in the 'narrow' cross population amplified using Primer 37 and shows an irregular peak shape, making it difficult to score, as there is only a 2bp difference between the parental alleles. A new PCR product gave a clear peak in sample (b), suggesting that there is a technical basis in PCR for this effect.
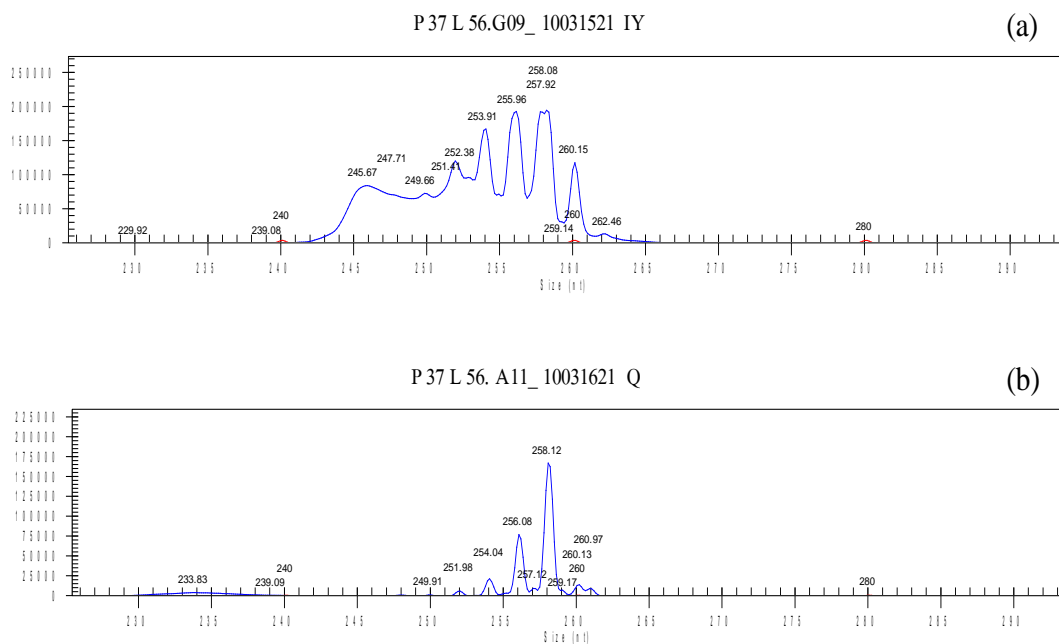


**Figure 3-11: Capillary electrophoresis showing a noisy background that made the scoring of allele size more difficult for Primer37 in the 'narrow' cross population. Sample (a) shows irregular peak shape, while in sample (b) the reamplified PCR product has a regular peak shape.**

**3.2.5 DNA Isolation from endosperm**

**3.2.5.1 DNA Extraction and quantitation**

Endosperm powder was made from bambara groundnut seeds with a Dremel craft drill. Genomic DNA was extracted using the GenElute Plant Genomic DNA kit (Sigma Aldrich). Isolated DNA with three incubation periods was quantified using gel electrophoresis (1% agarose). The result of different incubation periods revealed that the 30 minutes incubation had the best result for extraction of genomic DNA from dried endosperm, as shown on the agarose gel (Figure 3-12).
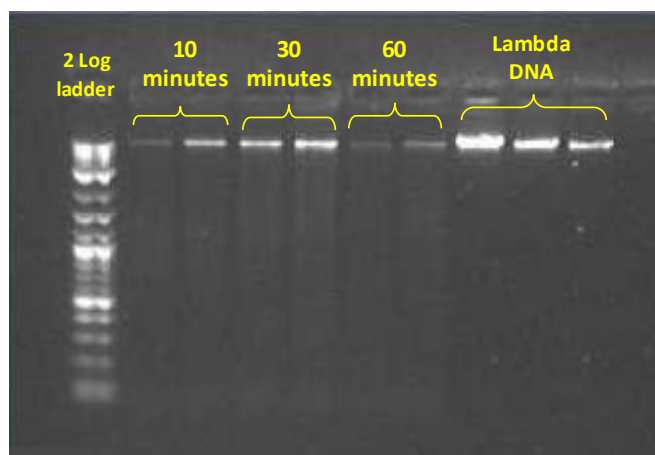


**Figure 3-12: Quantified DNA extracted from the endosperm of bambara groundnut seed with three incubation periods for extraction (10, 30 and 60 minutes), using the GenElute Plant Genomic DNA Extraction kit. Thirty minutes incubation produced the best results for quantity of DNA extraction.**

**3.2.5.2 Germination tests for treated seeds**

Analysis of variance was conducted using the Minitab software (release 16). Three levels of drilling treatment (not drilled and two drill diameters (3/64" and 1/16")) were applied against two seed sizes of bamabra groundnut, big and small seed sizes of 1cm and 0.5 cm in diameter. According to the results, the germination percentage was 48.9% and an analysis of variance shows that there was no significant difference for germination between the levels of drilling treatment [$F_{(2, 12)} = 2.8$, $p < .01$], seed size [$F_{(1, 12)} = 0.4$, $p < .01$] and also for their interaction [$F_{(2, 12)} = 0.4$, $p < .01$] (Table 3-6).

**Table 3-6: Analysis of variance for germination test of drilled seeds.**

| Source of Variation | df | SS | MS | F value | probability |
|---|---|---|---|---|---|
| Drilling | 2 | 1244 | 622 | 2.8 | 0.1 |
| Seed size | 1 | 89 | 89 | 0.4 | 0.539 |
| Drilling*seed size | 2 | 178 | 89 | 0.4 | 0.679 |
| Error | 12 | 2667 | 222 | | |
| Total | 17 | 4178 | | | |

## 3.3 Discussion

### 3.3.1 Microsatellite markers

Microsatellites essentially consist of short tandem repeated units and can be found in both coding and non-coding regions in all prokaryotic and eukaryotic genomes (Kalia *et al.*, 2011; Trivedi, 2004). It is considered to be a marker of choice to study molecular genetics in plant species (Somta *et al.*, 2011b). Microsatellite markers used in this study were developed from a genomic-enriched repeat library sequenced with non-titanium and titanium reagents (Roche 454 Pyrosequencing) and from a bambara groundnut leaf transcriptomic library (Roche 454 Pyrosequencing). Only one primer derived from another legume genome (soybean) was also mapped in the current study.

The RNA sequence data was generated as a part of an experiment to look at the effect of temperature on gene expression in leaf. Primer pairs were designed for microsatellite repeat motifs after their screening *in silico*. The development of a comprehensive set of genic-SSR markers has been reported in pigeon pea using deep transcriptome sequencing. They are considered as an important genetic resource in their application for understanding the genetic relationship among different accessions and regional origins in this crop (Dutta *et al.*, 2011).

The soybean microsatellite marker (AG81) was able to identify polymorphism in the parents of both populations. High sequence similarity at the amino acid level was identified between the AG81 sequence and a seryltRNA synthetase gene (*Arabidopsis thaliana*, Z70313), suggesting that the conservation of this locus could be due to functional constraints on its evolution (Peakall *et al.*, 1998).

However the cross-genus amplification rate of SSRs among *Medicago* and other forage legume (32%) was low in comparison to some other non-legumes (Dirlewanger *et al.*, 2002), the transferability of SSR markers in genetic studies of legume crops potentially opens a new approach to utilize them for the applications of comparative mapping, functional analysis and for assessing genetic diversity (Chandra, 2011).

The transferability of SSR markers across legume families has been investigated. SSRs from existing databases of different legumes (including *Medicago*, soybean, cowpea, and peanut) were used to screening 24 different legume accessions. About one-third of DNA markers from *Medicago* and soybean were reproducible and amplification was obtained in peanut, clover, cowpea, guar, and other legumes, suggesting that transferring SSRs across species and/or across genera could be an efficient approach to develop DNA markers, especially for minor crops (Wang *et al.*, 2004a).

Although a few SSR markers (10 markers) have been reported for bambara groundnut (Basu *et al.*, 2007a), markers from counterpart species such as azuki bean (Wang *et al.*, 2004b), cowpea (Xu *et al.*, 2010), and mungbean (Somta *et al.*, 2008) show a high rate of amplification when applied in bambara groundnut and their utilization was considered useful for genomics study in this crop (Somta *et al.*, 2011a).

DNA extracted from the individuals of the $F_3$ 'narrow' cross and the $F_2$ population of the wide cross using both the protocols of Dellaporta and the GenElute Plant Genomic DNA kit (Sigma Aldrich), respectively, were used to generate DNA stocks. DNA stocks of the 'wide' cross ($F_2$ population) were used directly in the PCR reaction without dilution because of the poor quality of the DNA. However, with additional cycles of amplification (+2/3 cycles) strong amplification was generally achieved,

making it unlikely that lack of template could be leading to issues with null alleles (allelic dropout and false alleles) caused by quality alone (Pompanon *et al.*, 2005)

The correct size of microsatellite amplification is considered to be critical for the confirmation of identity and correct locus amplification of SSR markers. In PCR amplification for a limited number of SSR (three) the product sizes of target sequences was greater than expected (bam2coL58, BN11 and BN21). Insertion/deletion of additional DNA between the designed primers could lead to differences in size for particular genotypes, compared to the design sequences. (Hackauf and Wehling, 2002). Incorrect priming or the abolition/mutation of the original priming site might also lead to an additional product being amplified and with transcriptome derived SSRs, the presence of an intervening intron when amplifying from genomic sequences is always a possibility (although primer pairs were tested against genomic DNA during the characterization phase). Genomic SSR fragments of unexpected size are usually monomorphic (Roder *et al.*, 1998) and these primers have been eliminated from further analysis and scoring to avoid genotyping errors. Furthermore running such fragments with capillary electrophoresis was not possible as all PCR products were sized with a 400 bp size standard and the software cannot provide a size estimate beyond this range (as the calibration curve would not extend this far).

### *Detection of SSR alleles*

Although agarose gel electrophoresis has relatively low resolution (Wang *et al.*, 2009), it was possible to score allelic size differences for some of the markers. Parental alleles were scored within the individuals of $F_3$ population derived from the 'narrow' cross on agarose gel using primer D.35497, as well as in the $F_2$ population of

'wide' cross for Primer10, Primer82 and D.48339. Parental allele differences within the individuals using these primers were in the order of 28-36bp, allowing their detection easily with 2.5% agarose gels and 1-3 hour run times on 80-90 Volt. The agarose gel system using a 3% concentration gel for 3.5 hours has been applied by other researchers to screen for polymorphism of SSR markers (Beyene *et al.*, 2005). The CEQ$^{TM}$ 8000 (Genetic Analysis System, Beckman Coulter, USA) was used to estimate different allele sizes for other primers in both 'narrow' and 'wide' cross populations. Although the CEQ$^{TM}$ 8000 contains an automated allele binning wizard to determine the sizes of alleles of markers, product sizes were scored manually from the electrophoretograms. Visual inspection is recommended to score sized fragments with the CEQ system to avoid mis-reporting of automated sizing caused by scoring stutters as a real peak in automated scoring. Combining automated allele size calling and human inspection of electropherogram data has been suggested to provide efficient and accurate detection of novel alleles (DeWoody *et al.*, 2006), but only peak size estimation can be relied upon. The automated calling system is unable to cope with changes in relative peak heights and it is essentially that the entire microsatellite profile is taken into consideration. Changes in this more realistically reflect genuinely different alleles. The heterozygosity of the individuals may have to be confirmed by direct comparison of their peaks to the parental allele traces.

All markers were screened individually before multiplexing, using the blue M13 WellRed Dye to provide information on peak patterns and size ranges of the expected alleles. This helped to avoid unidentified large-allele dropout in multiplexing, when one allele of a heterozygote randomly fails to amplify and leads to mistyping the heterozygote loci as homozygotes (Johnson and Haydon, 2007). Large allele dropout has been reported a common error in microsatellite scoring (Bonin *et al.*, 2004). The

amount of PCR products added to a post-PCR multiplex depended on the strength of the amplification products before pooling, as assessed on agarose gel. All primers were fluorescently labeled with the D4 (blue) dye to avoid the problem of false peaks due to colour bleed through and the PCR products were multiplexed for 2-3 primers which showed a wide allelic size difference between microsatellites in their pre-screening. The problem of false signal was reported in multiplexing different PCR products with different labeled dyes of D3 green and D4 blue (Molosiwa, 2012) where product allele sizes were similar.

### SSR polymorphism

Polymorphism of the microsatellite markers was estimated to be around 36.3% in the segregation population derived from the cross of DipC x Tiga necaru, while it was about 33.1% in the 'wide' cross population from the cross of DipC and VSSP11. A similar percentage of polymorphism (40%) was identified for bambara groundnut SSR markers in the eight tested Bambara groundnut single genotype accessions from all major growing regions of the world, while a lower level of polymorphism (17.4%) was identified for these accession using SSR primers from the relatives of bambara groundnut, azuki bean, cowpea and mungbean (Somta *et al.*, 2011b). A high polymorphism level for SSR markers (70.3%) was identified between the two chickpea cultivars (BG 256 and WR 315). These were used as the parents of mapping populations in linkage and quantitative trait loci (OTL) analysis to determine the position of a *Fusarium* wilt resistance gene (Qadir *et al.*, 2007). A higher polymorphism rate (81.3%) was presented in an investigation of pigeon pea. Sixteen newly developed SSR markers were screened on 40 genotypes representing different *Cajanus* species including eight wild pigeon pea genotypes (Saxena *et al.*, 2010).

The higher polymorphism observed in the 'narrow' cross compared to the 'wide' cross population could be due to the fact that parents of the 'narrow' cross came from different origins and domestication backgrounds. The results of a genetic similarity analysis of 87 bambara groundnut genotypes based on the similarity matrix of 296 unique polymorphic DArT markers revealed considerable structural differences between DipC and Tiga necaru populations (Stadler, 2009), while the wild accession of VSSP11, used as a male parent in constructing the 'wide' cross population, came from a similar background (despite being a wild ancestor/undomesticated individual) as DipC . This would argue that the 'wide' cross segregates mainly for domestication characters, but otherwise is genetically relatively uniform, while the 'narrow' cross does not segregate for domestication loci, but is genetically quite diverse.

Thirty SSR markers developed from the genomic microsatellite-enriched library of bambara groundnut and a single SSR from soybean were polymorphic in the 'narrow' cross population, while 22 markers were polymorphic in the 'wide' cross. While 15 markers derived from the transcriptome data were polymorphic in the narrow cross and 19 were polymorphic in the 'wide' cross, giving overall figures of 45 SSRs polymorphic in the narrow cross and a similar number, 41 polymorphic in the wide cross, which suggests levels of overall polymorphism in these crosses are similar, despite the wide cross being from the domesticated subspecies to the ancestor subspecies, spontenea

The DipC landrace was used as the female parent in both 'wide' and 'narrow' cross populations when crossed with landraces of VSSP11 (*V. subterranea* spontenea) and Tiga necaru (*V. subterranea* subterranea), respectively. The legitimacy of the DipC landrace as the source of the maternal parent was confirmed through production of the same allelic size in both populations with all the polymorphic transcriptome primers

and around half of the genomic microsatellite-enriched library primers applied. The fact that the maternal alleles are not identical in both crosses is likely to be a consequence of different individuals of the DipC landrace being used for the maternal parent (Massawe *et al.*, 2005).

### 3.4.2 Validation of the segregating $F_3$ population

The $F_3$ population derived from the cross between DipC and Tiga necaru landraces was checked for the residual levels of heterozygosity before using it to construct a linkage map and to carry out the subsequent QTL analysis. Segregating patterns of parental loci were followed and screened for all individuals using 33 polymorphic SSR markers. Residual heterozygosity was confirmed with these markers to match with the predicted Mendelian ratios expected from the population history.

### 3.4.3 DArT Assay

DArT is a genetic technology which has the ability to type thousands of loci in parallel, with reduced costs per data point compared to other genetic marker systems (Kilian *et al.*, 2005). Array development and genotyping was carried out in collaboration with Diversity Arrays Technology Pty Ltd (Yarralumla, Australia). In this study DArT arrays were developed and a total of 3.1% polymorphism per array element was obtained. Approximately 2% polymorphism of DArT array generated previously for the inter-subspecific cross of bambara groundnut was also used in this study. DArT as a high-throughput approach and has proven to be most cost- and time-efficient in studying genetic diversity in bambara groundnut and it is considered as a useful approach for other under-utilized crops (Mayes *et al.*, 2009; Olukolu *et al.*, 2012; Stadler, 2009). A total of 236 polymorphic markers were detected for the

117

individuals of the $F_3$ population derived from the cross of DipC x Tiga necaru. Due to the high locus specificity of DArT markers, these can be easily placed into genetic linkage maps with reasonable confidence that the same array element will usually detect the equivalent chromosomal location in all individuals (Wenzl *et al.*, 2006). Thus these developed markers along with SSR markers will be used for constructing a genetic linkage map of bambara groundnut in the segregating $F_3$ population. This was also used to improve the existing initial map from the cross of DipC and wild accession (VSSP11).

### 3.4.4 DNA isolation from the endosperm

Many attempts were made to develop a successful hybridization technique for bambara groundnut (Basu, 2005; Massawe *et al.*, 2004; Suwanprasert *et al.*, 2006). Hybrid seed production in bambara groundnut is by hand emasculation and pollination, the number of successful $F_1$ hybrids developed by artificial cross pollination can be reduced by self-pollination and it is not possible to distinguish genuine $F_1$ hybrids from self-pollination in the $F_1$ seed, as the testa is genetically maternal. It is only with the development of $F_2$ seed that non-maternal characters could be expressed and only with $F_3$ seed that segregation for these can be observed (Basu *et al.*, 2007c) although if there are substantial differences in morphology between the parents, these factors may be informative in the $F_1$ plant. PCR-based molecular marker techniques have been used to rapidly verify $F_1$ hybrids in plant crops (Sundaram *et al.*, 2008). The standard method has been by growing all the presumed hybrids and extracting genomic DNA from all of them to check their status as $F_1$ hybrids using molecular markers. In order to accelerate this process we have investigated the possibility of extracting DNA directly from the seed endosperm

before planting, in a way that does not reduce their viability. Genomic DNA was extracted from the endosperm of the seeds after optimizing the period of incubation in the extraction process. The DNA extracted with this process could be amplified in the polymerase chain reaction using molecular marker to identify genuine $F_1$ hybrids. The seeds tested showed no effect on their germination frequency compared to controls, suggesting that there was no reduction in viability (although viability was comparatively low in this experiment and the experiment needs to be confirmed with a large-scale analysis on recently harvested seed). This approach seems feasible to allow a rapid determination of hybrid status before planting of $F_1$ seed under optimal conditions to ensure good seed production for $F_2$ populations for research and breeding. Alternatively, confirmed $F_1$ seed could be sent for germination in countries with in-field programmes of bambara groundnut research.

# Chapter 4: SEGREGATING POPULATIONS AND TRAIT INHERITANCE

## 4.1 Introduction

The genetic improvement of crops is one of the most promising approaches for addressing future resource constraints (particularly land pressure and environmental concerns) and to enhance food production to meet the pressure of expected population growth in the near future.

Genetic improvement of crop species and genetic resource conservation and evaluation using both conventional and biotechnological approaches is an on-going activity, focused on improving tolerance to biotic and abiotic stresses while maintaining high levels of productivity, profitability and quality. Estimating the heritability of morphological and agronomical traits is an important part of effective selection for breeding traits (Firouzian, 2003). The development of molecular markers to such traits can assist the integration of desirable alleles into the crop genotype to accelerate progress in agriculture and the delivery of new cultivars and technologies.

### 4.1.1 Inheritance of agronomic traits in crop plants

Traits inheritance falls into two main classes; continuous variation and discrete classes. In crop plants most traits of agronomic important are quantitative and are likely to be controlled by many genes, often with an approximately normal distribution, in contrast to qualitative phenotypes which usually form discrete classes. The multiple genes often involved in many quantitative traits makes their control

genetically more complex as each gene may have a minor phenotypic effect and can be very sensitive to the environment (Choudhary *et al.*, 2008). The combination of many genes and their environmental interaction makes the study of quantitative traits more complicated (Chenu *et al.*, 2011; Ikram and Chardon, 2010; Koornneef *et al.*, 2004; Paran and Zamir, 2003).

Breeders often have limited information on the gene-to-phenotype architecture for the traits under study (Cooper *et al.*, 2005). Predictions of gene properties in populations of genotypes can be obtained through understanding the genetic basis of variation for simple and quantitative traits.

Although, very few morphological and developmental studies have been carried out in bambara groundnut, studies conducted on genome conservation and information gained from other legume crops could shed light on the inheritance patterns of important traits that are also present in bambara groundnut.

## 4.1.2 Inheritance patterns of phenotypic traits in selected legume crops

### 4.1.2.1 Qualitative trait samples

Thirty three domestication-related traits in azuki bean (*Vigna angularis*) were studied by Isemura *et al.* (2007). The traits were evaluated for the segregating populations ($BC_1$ and $F_2$) which were derived from the cross between the wild species *Vigna nepalensis* (female) and the cultivated azuki bean (male and recurrent parent for backcross). Three traits of *epicotyl colour*, *seed coat colour* and *black mottling of seed coat* were identified as qualitative traits. The segregation ratios for the three qualitative traits matched the ratios expected for control by a single gene in both

populations, except that *seed coat colour* in the $F_2$ population which showed distorted segregation significant at the 5% level, perhaps suggesting that the single gene model did not adequately explain the genetic control of this trait.

The expression of morphological traits was studied by Damayanti *et al.* (2010) in hybrids of the cross of domesticated var. *macrosperma* x wild Australian of tropical tuberous legume (*Vigna vexillata* (L.) A. Rich). Several traits relating to domestication were evaluated, including; *broad leaflet size, ovate leaflet shape, non-pigmented stems, non-dehiscent pods, uniform seed testa colour* and *green seed testa.* The standard chi-square goodness of fit test suggested that the segregation ratios for these traits except *seed testa colour* were consistent with control by single dominant genes, with the domesticated version acting recessively compared to the wild type traits. *Seed testa colour* was consistent with being controlled by two genes in dominance-epistatic relationship (Damayanti *et al.*, 2010).

In common vetch (*Vicia sativa* L.) a full diallel cross was conducted by Chowdhury *et al.* (2004) using five parents of different *cotyledon* and *testa colour* . The segregation patterns of *cotyledon colour* was studied in $F_2$, $BC_1$ and $BC_2$ populations and found to be governed by a single gene of two alleles with incomplete dominance. *Testa colour* was also found to be a qualitative trait and to be regulated by a single gene with complete dominance. Similar results of monogenic inheritance for seed coat color were reported by Milczak (1971) based on the analysis of $F_1$ population data.

Hasan *et al.* (2006) analysed a half diallel cross in French bean (*Phaseolus vulgaris* L.) for yield related traits. The result indicated that recessive alleles govern *days to 50% flowering* while dominant alleles are more important in controlling *days to maturity*.

In a diallel experiment using dry bean (*Phaseolus vulgaris* L.) the importance of both additive and dominance genetic components was emphasized in the control of different yield related traits, including *days to first flowering*, *days to podding* and *days to maturity*, however the high value of heritability in the narrow sense (explaining more than 50% of the trait variation) for *days to maturity* indicating the largely additive control of this trait and possibility of selection for this trait to improve this crop (Islam and Newaz, 2000).

### 4.1.2.2 Quantitative trait examples

In a genetic study of agronomic traits of 82 genotypes of mungbean (*Vigna radiata* (L.) Wilczek.), major additive gene effects were reported for *100-seed weight*, *seed yield/plant*, and *plant height* (Rohman *et al.*, 2003). Another investigation by Khattak *et al.* (2002) reported the importance of additive gene action in specifying *seed yield* and its components in mungbean. Similarly, inheritance analysis of agronomic traits of seven long bean varieties (*Vigna sesquipedalis*) found higher additive genetic variation compared to dominance variation in the control of *seed weight* and *pod length*. However, dominance was more important for *pod no./plant* and *seed no./pod* (Mak and Yap, 1980).

Phenotypic parameters were estimated in cowpea by Lopes *et al.* (2003) to understand the genetic control of seed size. The $F_1$, $F_2$, $BC_1$ and $BC_2$ populations from the cross TVx5058-09C X Manteiguinha along with their parents were analysed. The additive-dominance model fitted the data for *100-seed weight* and the additive effects were the more important. Five genes were predicted to control the expression of *100-seed weight* and the authors suggested that selection could be made in the early segregating generations for seed size in cowpea.

An inheritance study in French bean (*Phaseolus vulgaris* L.) by Hasan *et al.* (2006) estimated narrow sense heritability to be greater than 50% for *plant height* and *100-seed weight*, indicating that the majority of the phenotypic variation was due to additive gene action. It was concluded by the authors that selection would be effective for the improvement of these characters in French bean. *Seed yield/plant* was identified as a trait largely controlled by non-additive effects.

In dry bean (*Phaseolus vulgaris* L.) a diallel experiment was carried out using an $F_2$ population and evaluated in two different environments by Islam and Newaz (2000). They emphasized the importance of both additive and dominance genetic components in controlling the different yield related traits of *plant height*, *pods/plant*, *pod length*, *seeds/pod*, *seeds/plant* and *20-seed weight*. Heritability, in the narrow sense, showed high values (above 50%) for *pod length* in both environments, and *20-seed weight* in one environment, indicating that the major portion of the phenotypic variation of these two traits was additive in nature and selection of these traits could be an effective means to improve yield in dry bean.

Roy *et al.* (2006) studied ten traits in bush bean (*Phaseolus vulgaris* L.) to identify suitable traits for yield improvement. High heritability estimates in the broad sense were observed for *pods/plant* (90.76), *100-seed weight* (91.13), *seed yield/plant* (90.54) and *protein percentage* (99.78) in the genotypes, which indicated the presence of additive gene effects in addition to dominance effects. The authors stated that this crop could be improved through selection of these traits. An inheritance study on *days to flower* and *seed size* in pigeon pea (*Cajanus cajan* (L.) Millsp.) was conducted by Gupta *et al.* (1981). The authors used $F_1$, $F_2$ and backcrosses populations of a cross between two parents with large-seed (ICP-8504) and small seed (Prabhat). Additive gene effects were found to be most important in the expression of both *earliness* and

*seed size*. The result showed that selection could be used to improve the *seed size* in early maturing cultivars of pigeon pea.

In an inheritance study of seed yield and related traits in 15 lentil genotypes (*Lens culinaris* Medik) significant genetic variation was observed by Rasheed *et al.* (2008) for all traits studied. High heritability values and expected percent of genetic advance (respectively) were observed for *seed yield* (98.30%, 128.20%), *harvest index* (97.10%, 79.40%), *biological yield* (94.30%, 56.10%) and *100-seed weight* (88.30%, 50.80%) indicating the role of additive genes in controlling these traits. It was concluded that the traits of *100-seed weight, harvest index* and *biological yield* could be exploited for the improvement of seed yield in this crop.

The studies made in legume crops suggest the differences in number of genes controlling different traits. The effect of single or a few genes has been identified for some of the phenology traits such as days to emergence, days to flowering and maturity, seed and coat colour, while the other traits (yield related) were found to be under multiple gene effect. Identify the genetic control of these traits will delight the process to study their inheritance with the suitable breeding program whether to be through selection of hybridization. The inheritance studies of these legume related crops could be exploited for the improvement of these traits in bambara groundnut.

## 4.1.3 Trait correlation studies in bambara groundnut

A number of studies have been made in bambara groundnut on cultivars development and the relationship between yield and its components. However, it is important to note that bambara groundnut generally exists as landraces. As a strongly inbreeding species, the landraces are likely to be mixtures of inbred lines. How genetically diverse landraces are will depend upon the cultivation history of the landrace.

Ofori (1996) estimated the effects of yield components on the *seed yield* in bambara groundnut germplasm from Ghana. Both *pod no./plant* and *100-seed weight* were identified to have a positive effects on *seed yield*, as might be expected. For the bunch morphology type, *leaf no./plant* was more important in determining *seed yield*, while *leaf size* was more important in the spreading morphology types. Association was detected between the traits of *seed size* and *leaf size*. Thus selection of genotypes with bigger leaf size could contribute to improve the productivity of bambara groundnut.

Another correlation coefficient analysis for pod yield components of bambara groundnut was made by Makanda *et al.* (2009). Twenty bambara groundnut landraces were evaluated at four planting dates in Zimbabwe outside the traditional growing season of bambara groundnut. A significant correlation was indicated for most of the traits studied in different planting dates.

Another experiment was conducted under screen house conditions at the Sokoine University of Agriculture, Tanzania. *Leaflet length* was found to be the only vegetative variable which had a significant and positive correlation with *seed yield*. This positive relationship suggests plant architecture for bambara groundnut that favors longer leaves, results in more pods and higher *seed yield*. *Plant height* and *leaflet length* were both positively correlated with number of pods produced. The authors also reported that *pod no./plant* was highly correlated with *seed yield*. Whereas *days to flowering* was negatively correlated with *seed yield* and *petiole length* was negatively correlated with *100-seed weight*; this was largely due to its negative  effect on *seed size* (Misangu *et al.*, 2007).

Chijioke *et al.* (2010) performed a comparative study in bambara groundnut to evaluate the contribution of agronomical traits to yield in this crop. They evaluated 30 local bambara groundnut cultivars over two cropping seasons in Nsukka, Nigeria. The

correlation coefficients suggested that seed weight was significantly associated with *leaves per plant*, *plant height*, *flowers per plant* and *pods per plant*, considering these traits as major determinant of yield in bambara groundnut.

Inheritance patterns in bambara groundnut were investigated for a number of traits using a segregating $F_2$ population and 14 small $F_3$ families derived from the inter-subspecific cross of a domesticated DipC and wild VSSP11 accession. The additive gene effects and several genes were suggested to control the traits of *leaf area*, specific *leaf area*, *carbon isotope discrimination* (CID) and *100-seed weight*. In contrast, the variation for traits such as *internode length*, *stems/plant*, *days to emergence* and *seed eye pattern around the hilum* was likely to be under largely monogenic control (Basu *et al.*, 2007c).

The aim of the present study was to identify the inheritance patterns of domestication and agronomically important traits in an $F_3$ progeny segregating from the cross of a DipC (female) landrace with a Tiga necaru (male) landrace grown under both glasshouse and field conditions. The population itself was used to construct a genetic linkage map using DArT and SSR molecular markers. The combination of this linkage map with the trait data from current and previous studies was used to conduct a QTL analysis in this cross. An analysis of the traits and their inheritance patterns and distribution is presented here, before subsequent chapter go on to describe the mapping and QTL analyses.

**4.2 Results**

The results presented in this chapter represent the analysis of the distribution and associations of morphological traits for the segregating $F_2$ and $F_3$ populations derived from the intra-(sub-specific) cross between the DipC and Tiga necaru landraces.

The $F_2$ population data of the "narrow" cross was collected previously in the Tropical Crops Research Unit from the individual plants for each line. Data measured in the FutureCrop glasshouse were recorded for four replicates in a CRD design. A single plant represented each replicate and the QTL analysis was based on the means of replicates however the variance between the lines for all 4 replications is also presented for the glasshouse data. Data recorded for the Indonesian field trial of the same cross ($F_3$) came from the mean of (5-15) individual plants per line (depending on the availability of seed).

**4.2.1 Data distribution of morphological traits for the $F_3$ population**

Traits measured are (in concordance with the IPGRI descriptors for Bambara groundnut), describing below a number of them evaluated in both FutureCrop glasshouse and the field.

***Days to emergence*:** Data recorded for the number of days from sowing to attaining the first true leaf for the segregation $F_3$ population lines of FutureCrop glasshouse showed non-normal distribution (Table 4-1). As the P-value (P = 0.000) for the Anderson-Darling normality test was less than an alpha of 0.01, significant departure from normality is indicated. The parental data in the glasshouse revealed that the Tiga necaru parent emerged within 7 days (sd= 0.94) and DipC in 10.7 days (sd= 0.82),

classifying this trait into distinct classes. The field data also indicated a non-normal distribution of this trait (Table 4-2).

***Leaf area* (cm$^2$)**: *Leaf area* data using the method of Cornelissen *et al.* (2002), showed a non-normal distribution in the glasshouse, while the *leaf area* calculated in the field was revealed a normal distribution (Appendix 4). A Box-Cox transformation was applied to normalize the *leaf area* data of glasshouse (Figure 4-1).

**Table 4-1: Statistical analysis and the distribution of trait data in the F$_3$ population and the parents in the glasshouse.**

| Trait | Mean value | Variance[a] | Variance[b] | Standard deviation | Skewness | Kurtosis | P value | Individual no. | normality | parents DipC | parents Tiga necaru |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Days to emergence* | 7.9 | 1.3 | 1.5 | 2.80 | 3.0 | 13.5 | 0.00 | 73 | non-normal | 10.7 | 7.0 |
| *Days to flowering* | 36.4 | 9.1 | 11.1 | 12.28 | 2.3 | 8.2 | 0.00 | 64 | non-normal | 39.3 | 33.7 |
| *Flower no./plant* | 105.2 | 1187.2 | 1285.3 | 47.13 | 0.4 | -0.2 | 0.47 | 64 | normal | 116.7 | 82.0 |
| *Plant height* | 25.5 | 4.8 | 6.6 | 8.63 | -0.3 | -0.0 | 0.07 | 71 | normal | 27.3 | 21 |
| *Petiole length* | 19.1 | 2.9 | 4.5 | 6.47 | -0.8 | 0.4 | 0.01 | 71 | normal * | 20.6 | 14.1 |
| *Terminal leaflet length* | 8.2 | 0.7 | 1.0 | 2.81 | -0.5 | 0.0 | 0.22 | 64 | normal | 7.8 | 7.2 |
| *Terminal leaflet width* | 3.5 | 0.2 | 0.4 | 1.24 | -0.1 | -0.7 | 0.71 | 64 | normal | 3.6 | 2.4 |
| *Leaf area* | 3461.2 | 1925714 | 2153621 | 1719.54 | 1.8 | 5.6 | 0.00 | 70 | non-normal | 2678 | 2312.1 |
| *Plant spread* | 33.7 | 40.4 | 51.8 | 12.54 | 0.8 | 0.5 | 0.02 | 64 | normal* | 26.3 | 31 |
| *Stem no./plant* | 10.3 | 3.0 | 4.6 | 3.76 | 0.6 | 0.7 | 0.01 | 64 | normal* | 11.0 | 9.7 |
| *Node no./stem* | 8.7 | 4.7 | 6.6 | 3.49 | 0.6 | 1.1 | 0.08 | 64 | normal | 6.2 | 12.0 |
| *Internode length* | 2.0 | 0.9 | 1.1 | 1.12 | 0.6 | 0.1 | 0.17 | 64 | normal | 1.1 | 1.6 |
| *Growth habit* | 12.0 | 43.0 | 51.7 | 7.26 | 1.3 | 1.3 | 0.00 | 64 | non-normal | 13.5 | 8.6 |
| *Pod no./plant* | 47.6 | 256.0 | 319.2 | 21.57 | 0.6 | 0.2 | 0.08 | 64 | normal | 48.3 | 19.5 |
| *Double seeded pods/plant* | 5.4 | 16.9 | 24.3 | 4.22 | 1.3 | 1.6 | 0.00 | 64 | non-normal | 6.0 | 2.3 |
| *Peduncle length* | 2.9 | 1.1 | 1.5 | 1.35 | 0.2 | -0.7 | 0.24 | 64 | normal | 2.2 | 3.8 |
| *Pod weight/plant* | 36.8 | 216.4 | 287.6 | 18.26 | 0.7 | 0.1 | 0.03 | 64 | normal* | 40.0 | 35.0 |
| *Pod length/ plant* | 16.1 | 1.3 | 1.6 | 5.41 | -0.1 | -0.6 | 0.89 | 64 | normal | 17.4 | 14.8 |
| *Pod width/plant* | 13.0 | 0.7 | 1.0 | 4.33 | -0.2 | -0.4 | 0.59 | 64 | normal | 13.4 | 12.1 |
| *Double seeded pods length* | 24.8 | 6.7 | 8.9 | 9.19 | -0.5 | 1.3 | 0.04 | 62 | normal | 27.1 | 20.9 |
| *Double seeded pods width* | 13.3 | 0.6 | 0.8 | 4.82 | 0.1 | -0.5 | 0.58 | 62 | normal | 13.0 | 12.1 |
| *Seed length* | 11.1 | 0.7 | 0.9 | 3.73 | 0.1 | -0.4 | 0.42 | 64 | normal | 11.5 | 10.4 |
| *Seed width* | 9.2 | 0.3 | 0.4 | 3.08 | -0.8 | 1.7 | 0.08 | 64 | normal | 9.7 | 9.0 |
| *Seed no./plant* | 51.5 | 317.9 | 423.2 | 23.68 | 0.5 | -0.5 | 0.01 | 64 | normal* | 52.3 | 35.0 |
| *Seed weight* | 28.9 | 144.9 | 173.3 | 14.68 | 0.7 | -0.1 | 0.00 | 64 | non-normal | 33.4 | 14.7 |
| *Biomass dry weight* | 61.3 | 524.3 | 663.4 | 29.31 | 0.8 | 0.0 | 0.00 | 64 | non-normal | 60.8 | 49.8 |
| *Shelling%* | 79.1 | 40.8 | 48.3 | 26.68 | -0.2 | 0.1 | 0.37 | 64 | normal | 83.4 | 43.8 |
| *100-seed weight* | 55.6 | 75.6 | 97.8 | 19.98 | 0.2 | -0.3 | 0.65 | 64 | normal | 63.6 | 43.8 |
| *Eye pattern around hilum* | 1.45 | 0.20 | 0.3 | 0.65 | 0.25 | -1.76 | 0.00 | 64 | non-normal | 1 | 2 |

\* Normal distributed data at 99% of critical value

a: variance for the means of lines

b: variance of lines for all four replicates

***Internode length* (mm):** Wide variation was recorded for *internode length* counted as the average of the five longest stems per plant at the 4th internode, with the trait ranging from 0.33-4.72 cm in the glasshouse and 0.28-2.24cm in the field. An Anderson-Darling normality test indicated a normal distribution of this quantitative trait in both locations (Appendix 5 and Appendix 6).

**Table 4-2: Statistical analysis and distribution of the trait data in the $F_3$ population in the field**

| Trait | Mean value | Variance | Standard deviation | Skewness | Kurtosis | Individual No | *P* value | A-Squared | Distribution |
|---|---|---|---|---|---|---|---|---|---|
| *Days to emergence* | 6.5 | 0.4 | 1.41 | 1.3 | 2.7 | 70 | 0.00 | 9.57 | non-normal |
| *Days to flowering* | 42.9 | 4.7 | 17.83 | 1.2 | 1.1 | 65 | 0.00 | 4.78 | non-normal |
| *Plant height* | 18.5 | 11.1 | 6.57 | -0.5 | 1.4 | 65 | 0.08 | 0.67 | normal |
| *Petiole length* | 10.7 | 5.7 | 4.15 | -0.8 | 0.8 | 64 | 0.04 | 0.77 | normal |
| *Terminal leaflet length* | 4.8 | 0.7 | 1.70 | -1.3 | 1.8 | 65 | 0.00 | 1.96 | non-normal |
| *Terminal leaflet width* | 2.1 | 0.1 | 0.76 | -0.7 | 1.1 | 65 | 0.02 | 0.94 | normal* |
| *Leaf area* | 672.2 | 108700 | 377.69 | 0.1 | -0.7 | 64 | 0.51 | 0.33 | normal |
| *Plant spread* | 27.1 | 76.6 | 11.77 | -0.7 | 0.6 | 65 | 0.13 | 0.57 | normal |
| *Stem no./plant* | 4.0 | 0.5 | 1.54 | 0.8 | 3.9 | 63 | 0.02 | 0.90 | normal* |
| *Node no./stem* | 14.6 | 15.4 | 6.18 | 0.9 | 4.3 | 63 | 0.09 | 0.64 | normal |
| *Internode length* | 1.0 | 0.2 | 0.53 | 0.7 | 0.3 | 62 | 0.05 | 0.76 | normal |
| *Growth habit* | 12.8 | 40.1 | 7.51 | 1.6 | 3.2 | 60 | 0.00 | 2.18 | non-normal |
| *Pod no./plant* | 7.7 | 21.5 | 4.99 | 0.9 | 0.8 | 64 | 0.007 | 1.08 | non-normal |
| *Double seeded pods/plant* | 1.4 | 1.3 | 1.16 | 1.0 | 1.6 | 61 | 0.00 | 1.89 | non-normal |
| *Peduncle length* | 4.5 | 2.9 | 2.13 | 1.2 | 0.6 | 65 | 0.00 | 3.81 | non-normal |
| *Pod weight* | 4.0 | 6.4 | 2.69 | 0.7 | 0.0 | 64 | 0.02 | 0.89 | normal* |
| *Pod length* | 13.6 | 3.1 | 4.96 | 0.7 | 3.7 | 63 | 0.01 | 1.07 | normal* |
| *Pod width* | 9.5 | 0.9 | 3.39 | -0.6 | 0.9 | 63 | 0.19 | 0.51 | normal |
| *Seed length* | 8.1 | 1.3 | 3.44 | -1.8 | 6.6 | 58 | 0.00 | 1.23 | non-normal |
| *Seed width* | 6.1 | 0.6 | 2.54 | -1.9 | 5.2 | 58 | 0.00 | 2.53 | non-normal |
| *Seed no./plant* | 6.9 | 17.5 | 4.61 | 0.7 | 0.5 | 59 | 0.07 | 0.68 | normal |
| *Seed weight* | 2.7 | 3.4 | 1.97 | 0.7 | 0.2 | 59 | 0.06 | 0.72 | normal |
| *Biomass dry weight* | 8.5 | 13.8 | 5.06 | 0.4 | -0.4 | 63 | 0.40 | 0.38 | normal |
| *Shelling%* | 63.3 | 379.7 | 30.38 | -0.6 | 0.2 | 59 | 0.31 | 0.42 | normal |
| *100-seed weight* | 38.1 | 162.0 | 18.80 | -0.1 | 2.1 | 59 | 0.01 | 1.00 | normal* |

* Normal distributed data at 99% of critical value
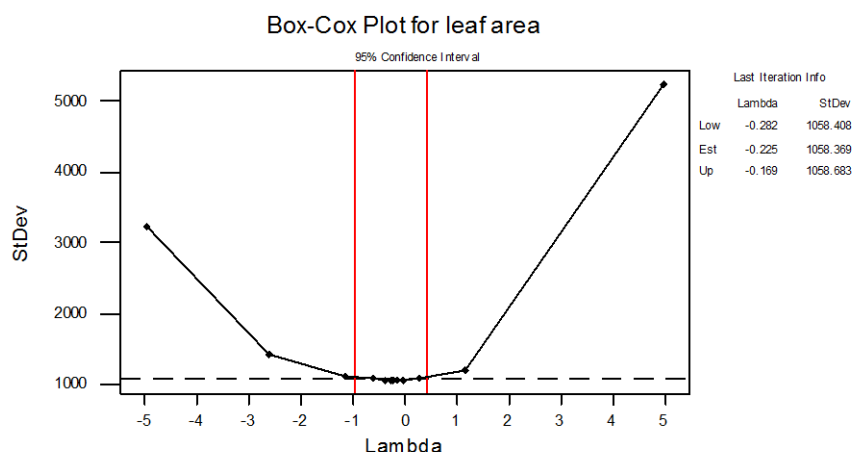
**Figure 4-1: Box-Cox plot for *leaf area* for the F₃ population grown in glasshouse.**

***Growth habit*:** Non-normal distribution of both glasshouse and field data was revealed (Appendix 7 and Appendix 8). Trait distribution of *growth habit* in the population under study showed a highly significant non-normal distribution (P-value= 0.000). Based on the parental values obtained an attempt was made to classify offspring into parental classes, bunched habit of 13.53 or more for DipC parent in contrast to Tiga necaru parent of semi bunched types of 8.567 or less (spreading types). A Chi-square test was carried out by grouping these trait data into three probable classes; beyond the parental classes and intermediate class. The results were consistent with segregation in a ratio of 3:2:3, with 21 plants representing the DipC parental class, 25 belonging to the Tiga necaru class and 19 classified as the intermediate type (21:19:25), out of 65 segregated F₃ lines. Chi-square tests gave a result of 0.95 at 2 df <5.99 of probability 0.05. The result of this hypothesis was consistent with co- dominance of *growth habit* in the present cross.

***Pod no./plant*:** In order to estimate the yield of bambara groundnut the number of *pods/plant* was recorded. A normal distribution was observed in the F₃ population

with a P-value of 0.08 (Appendix 9). While this trait was not normally distributed

having a P-value of 0.007) in the field. Box-Cox transformation was used to

normalize this trait to give a 0.25 P-value in an Anderson-Darling test (Appendix 10).

*Double seeded pods/plant* was recorded and found to be non-normally distributed in

both glasshouse and field experiments. The data distribution of *double seeded*

*pods/plant* in the glasshouse was normalized using a Box-Cox transformation

(Appendix 11).

**Seed weight** (**g/plant**): A wide range of trait segregation was observed for *seed*

*weight/plant*. Despite non–normal distribution of the data for this trait in the

glasshouse, normal distribution was revealed for the data recorded in the field (p =

0.056). Box-Cox transformation was applied for glasshouse data to maximize the

closeness to normality, the transformed data was revealed to be normally distributed

(p = 0.19) (Appendix 12).

**Biomass dry weight** (**g/plant**): Biomass dry weight reflects the amount of energy

stored in the plant through capturing carbon and sunlight. The data recorded for

biomass in the glasshouse, unlike the field data collected for this trait, did not show a

normal distribution pattern in the $F_3$ population lines. A Box-Cox transformation test

was used to normalize the data for the glasshouse a P-value of 0.06 (Figure 4-2)
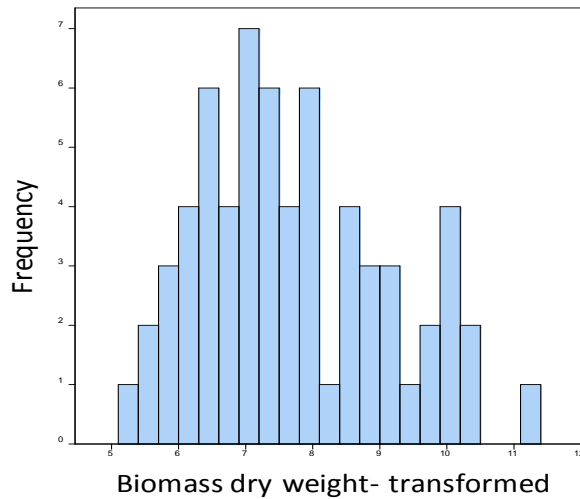
**Figure 4-2: Histogram of normal distribution for transformed *biomass dry weight* in the glasshouse for the F$_3$ population.**

*Eye pattern around hilum*: Since both parents have a creamy *testa color* and the *eye pattern* is present only in DipC maternal parent, the *eye pattern* segregation was studied in the glasshouse. Discontinuous distribution was revealed with the *eye pattern* trait (Figure 4-3). Chi-square tests were applied to the ratio of 40 presence: 24 absence. A hypothesis of complete dominance of *eye pattern* was proposed. 41 plants belonged to the maternal type (DipC) and 23 plants had no *eye pattern around hilum*, belonging to the Tiga necaru paternal type. The $X^2$ value at 1df was 0.06< 3.84 at 0.05 probability, indicating that this trait could be under the control of a single dominant gene for *eye pattern* presence.



**Figure 4-3: Presence versus absence of *eye pattern around hilum* for the seed in both parents of DipC and Tiga necaru landraces.**

133

## 4.2.2 Distribution of the $F_2$ population data

Previous data collected from the segregating $F_2$ population derived from the same intra-subspecific cross of DipC x Tiga necaru landraces was also used to study the inheritance of phenotypic traits of this crop before QTL analysis. The distribution of all these data except *seed weight* matched with data recorded for the traits of $F_3$ progenies of the glasshouse, although some of them were transformed to follow the continuous distribution (Table 4-3).

### *Trait data of the segregating $F_2$ population:*

As in the segregating $F_3$ data grown in glasshouse and field, *days to emergence* for the $F_2$ segregating population of the same cross grown in TCRU showed non-normal distribution pattern. It was not possible to partition the data into different categories depends on their parents' data as both DipC and Tiga necaru had no difference in the time spending to germinate.

**Table 4-3: Statistical analysis and the distribution of traits data of $F_2$ population and the parents of narrow cross (DipC x Tiga necaru).**

| Trait | Mean value | Variance | Standard deviation | Skewness | Kurtosis | Individual no. | *P* value | Normality | DipC | Tiga necaru |
|---|---|---|---|---|---|---|---|---|---|---|
| *Days to emergence* | 12.0 | 6.4 | 2.52 | 1.8 | 3.4 | 73 | 0.00 | non-normal | 10 | 10 |
| Days to flowering | 6.7 | 0.1 | 4.14 | 0.7 | 0.3 | 73 | 0.00 | non-normal | 43 | 40 |
| *Plant height* | 23.2 | 32.5 | 5.66 | -0.2 | -0.6 | 73 | 0.04 | normal* | 21 | 10 |
| *Petiole length* | 18.3 | 22.5 | 5.32 | 0.3 | 0.2 | 73 | 0.04 | normal* | 20.64 | 11.7 |
| *Terminal leaflet length* | 9.6 | 2.3 | 1.51 | 0.5 | 1.4 | 73 | 0.11 | normal | 7 | 7 |
| *Terminal leaflet width* | 4.1 | 0.7 | 0.82 | 0.0 | 0.2 | 73 | 0.16 | normal | 2.8 | 1.7 |
| *Plant spread* | 56.1 | 160.7 | 12.59 | 0.0 | -0.6 | 73 | 0.40 | normal | 53 | 29 |
| *Pod no./plant* | 120.9 | 6358.6 | 79.19 | 1.1 | 1.8 | 73 | 0.02 | normal* | 102 | 27 |
| *Double seeded pods* | 8.3 | 71.0 | 8.36 | 1.2 | 0.4 | 73 | 0.00 | non-normal | 3 | 1 |
| *Seed no./plant* | 115.4 | 6049.2 | 77.24 | 1.0 | 1.0 | 73 | 0.00 | non-normal | 70 | 22 |
| *Seed weight* | 39.6 | 784.0 | 27.81 | 0.4 | -0.8 | 73 | 0.00 | non-normal | 23.49 | 5.94 |
| *Biomass dry weight* | 99.2 | 4482.2 | 66.49 | 0.6 | 0.0 | 73 | 0.00 | non-normal | 66.12 | 14.56 |
| *Shelling%* | 73.1 | 46.0 | 6.73 | -1.2 | 2.2 | 73 | 0.01 | normal* | 77.77 | 74.32 |
| *100-seed weight* | 32.8 | 133.4 | 11.47 | 0.0 | -0.6 | 73 | 0.30 | normal | 33.57 | 27 |

\* Normal distributed data at 99% of critical value

Other traits including *terminal leaflet length, terminal leaflet width*, *plant spread*, *pod no./plant*, *shelling%* and *100-seed weight* (Figure 4-4 and Appendix 13) also tested as normally distributed with Anderson-Darling normality tests. While the traits of *double seeded pods/plant*, *seed no./plant* and *biomass dry weight* were transformed with a Box-Cox transformation to correct to a normal distribution (Appendix 14). *Seed dry weight* was not distributed normally even after transformation analysis.
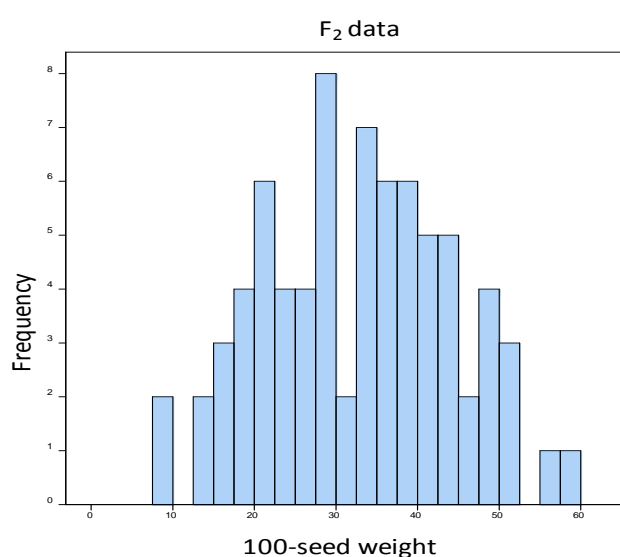


**Figure 4-4: Histogram of normal distribution of *100-seed weight* for F$_2$ progeny data**

### 4.2.3 Association of the traits

Phenotypic correlations were conducted between the evaluated vegetative and yield-related traits based on the means of F$_3$ segregating population derived from the cross between DipC x Tiga necaru using Pearson's Correlation Coefficient analysis. MINITAB (release 16) software was used to construct the regression plots and to detect the significant association between the traits data for all F$_2$ and F$_3$ populations evaluated under different environments.

**4.2.3.1 Association between traits for the F₃ generation in the glasshouse**

The correlation and adjusted regression values for the glasshouse trait data are presented in Table 4-4.

***Flower no./plant:*** was significantly associated with all vegetative growth traits except the traits *days to emergence*, *days to podding* and *growth habit*. This trait could be account for as much as 32% of the variation observed in *plant spread. Flower no./plant* also had significant negative correlation with *days to flowering* (Pearson's correlation coefficient, -0.48; p=0.000). Whereas it was found to be associated significantly with *seed weight* and a number of yield component traits such as *pod weight*, *seed no./plant, biomass dry weight and 100-seed weight*. It was accounted for 44% of the variation associated with *biomass dry weight* (Appendix 15).

***Plant spread***: A negative and significant correlation of this trait was found with *days to flowering* and *growth habit* (Pearson's correlation coefficient, -0.382; p= 0.002 and -0.615; p=0.000, respectively). *Plant spread* explained 56% of the variation in *biomass dry weight* of this population. Excluding the traits of *emergence*, *days to podding*, *stem no./plant*, *seed length* and *Shelling%, plant spread* showed a positive and significant associations with all the others. A regression analysis suggested *plant spread* accountable for 37% of the variation observed in *growth habit* (Appendix 16).

***Internode length***: Among the vegetative growth traits the highest positive correlation was observed between *internode length* and *peduncle length* at +0.799 (p= 0.000) followed by *plant spread*, whereas for yield related trait a higher Pearson's Correlation Coefficient value (r= +0.408; p= 0.001) was observed between *internode length* and *double seeded pods/plant*. Regression analysis suggested that 15% of the variation in *double seeded pods* could be accounted for by *internode length* (Figure 4-5). *Plant spread* had a significantly negative correlation with *growth habit*; -0.793 at

p= 0.000. The regression analysis suggested that *internode length* accounted for 62.4% of the variation in *growth habit* in the population under study.
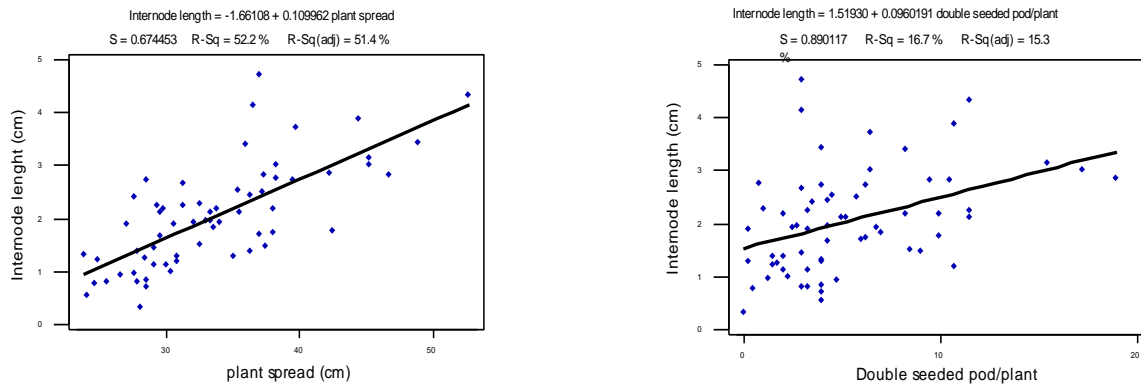


**Figure 4-5: A fitted line regression plot of *internode length* with both *plant spread* and *double seeded pods/plant* in F$_3$ population of glasshouse.**

***Pod no./plant*:** As a yield component trait, it had a significant positive association with all the vegetative growth traits, while it was in negative association with *days to flowering* (r= -385; p= 0.002). Regression analysis revealed that *pod no./plant* accounted for 51% of the variation in *node no./plant* of the population under study. *Pod no./plant,* beside its strong association with *biomass dry weight*, was found to have positive and significant correlation with *seed no.*, *seed weight* and *pod weight/plant* (r=+0.969, +0.881 and +0.875 at p= 0.000, respectively) but has no significant relationship with the dimensions of the pod and seed. *Seed yield* accounted for as much as 77% of the variation in *pod no./plant* in this population (Appendix 17). Similar pattern of the relationship for *double seeded pods/plant* with the other traits was scored with different significant level.

**Table 4-4: Relationship between traits in the F$_3$ population of the cross DipC x Tiga necaru for glasshouse data based on the Pearson's correlation analysis showing correlation and *P* values.**

Legend:
- r > 0.7 (yellow)
- r = 0.5-0.69 (green)
- r = 0.30-0.49 (blue)

Each cell shows the correlation coefficient (top) and the *P* value (in parentheses).

| Trait | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days to emergence | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Flower no./plant | 2 | -0.196 (0.121) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Leaf no./plant | 3 | -0.158 (0.212) | 0.553 (0.000) | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Days to flowering | 4 | 0.442 (0.000) | -0.478 (0.000) | -0.391 (0.001) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Days to podding | 5 | 0.418 (0.001) | 0.054 (0.672) | 0.105 (0.409) | 0.546 (0.000) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Plant height | 6 | -0.230 (0.068) | 0.435 (0.000) | 0.410 (0.001) | -0.426 (0.000) | -0.195 (0.122) | | | | | | | | | | | | | | | | | | | | | | | | | |
| Petiole length | 7 | -0.390 (0.001) | 0.442 (0.000) | 0.299 (0.017) | -0.477 (0.000) | -0.336 (0.007) | 0.859 (0.000) | | | | | | | | | | | | | | | | | | | | | | | | |
| Leaf area | 8 | -0.101 (0.426) | 0.624 (0.000) | 0.873 (0.000) | -0.454 (0.000) | 0.029 (0.818) | 0.456 (0.000) | 0.372 (0.002) | | | | | | | | | | | | | | | | | | | | | | | |
| Terminal leaflet length | 9 | -0.128 (0.313) | 0.445 (0.000) | 0.133 (0.293) | -0.489 (0.000) | -0.288 (0.021) | 0.341 (0.006) | 0.379 (0.002) | 0.485 (0.000) | | | | | | | | | | | | | | | | | | | | | | |
| Terminal leaflet width | 10 | -0.045 (0.721) | 0.344 (0.005) | 0.291 (0.020) | -0.308 (0.013) | -0.155 (0.222) | 0.333 (0.007) | 0.335 (0.007) | 0.645 (0.000) | 0.563 (0.000) | | | | | | | | | | | | | | | | | | | | | |
| Plant spread | 11 | -0.112 (0.379) | 0.576 (0.000) | 0.385 (0.002) | -0.382 (0.002) | -0.121 (0.339) | 0.254 (0.043) | 0.330 (0.008) | 0.525 (0.000) | 0.449 (0.000) | 0.447 (0.000) | | | | | | | | | | | | | | | | | | | | |
| Stem no./plant | 12 | -0.349 (0.005) | 0.352 (0.004) | 0.443 (0.000) | -0.381 (0.002) | -0.263 (0.036) | 0.583 (0.000) | 0.474 (0.000) | 0.416 (0.001) | 0.212 (0.093) | 0.270 (0.031) | 0.041 (0.750) | | | | | | | | | | | | | | | | | | | |
| Branch no./plant | 13 | 0.240 (0.056) | 0.441 (0.000) | 0.496 (0.000) | -0.041 (0.748) | 0.253 (0.044) | 0.174 (0.168) | 0.056 (0.662) | 0.466 (0.000) | 0.147 (0.246) | 0.093 (0.464) | 0.380 (0.002) | -0.046 (0.719) | | | | | | | | | | | | | | | | | | |
| Node no./stem | 14 | 0.103 (0.420) | 0.568 (0.000) | 0.620 (0.000) | -0.297 (0.017) | 0.126 (0.321) | 0.142 (0.264) | 0.108 (0.394) | 0.600 (0.000) | 0.257 (0.041) | 0.188 (0.138) | 0.612 (0.000) | -0.117 (0.358) | 0.772 (0.000) | | | | | | | | | | | | | | | | | |
| Internode length | 15 | -0.007 (0.958) | 0.395 (0.001) | -0.047 (0.713) | -0.321 (0.010) | -0.128 (0.312) | 0.051 (0.691) | 0.231 (0.066) | 0.139 (0.273) | 0.401 (0.001) | 0.290 (0.020) | 0.722 (0.000) | -0.132 (0.297) | 0.080 (0.528) | 0.343 (0.006) | | | | | | | | | | | | | | | | |
| Pod no./plant | 16 | -0.121 (0.342) | 0.684 (0.000) | 0.763 (0.000) | -0.385 (0.002) | -0.006 (0.964) | 0.448 (0.000) | 0.382 (0.002) | 0.772 (0.000) | 0.331 (0.008) | 0.411 (0.001) | 0.660 (0.000) | 0.262 (0.036) | 0.539 (0.000) | 0.720 (0.000) | 0.254 (0.043) | | | | | | | | | | | | | | | |
| Double seeded pods/plant | 17 | -0.219 (0.082) | 0.185 (0.143) | 0.346 (0.005) | -0.325 (0.009) | -0.202 (0.110) | 0.347 (0.005) | 0.379 (0.002) | 0.410 (0.001) | 0.238 (0.058) | 0.309 (0.013) | 0.641 (0.000) | -0.057 (0.653) | 0.182 (0.149) | 0.433 (0.000) | 0.408 (0.001) | 0.554 (0.000) | | | | | | | | | | | | | | |
| Peduncle length | 18 | -0.049 (0.703) | 0.370 (0.003) | 0.065 (0.609) | -0.315 (0.011) | -0.068 (0.592) | 0.035 (0.784) | 0.160 (0.208) | 0.211 (0.095) | 0.394 (0.001) | 0.243 (0.053) | 0.750 (0.000) | -0.172 (0.173) | 0.224 (0.076) | 0.472 (0.000) | 0.799 (0.000) | 0.268 (0.032) | 0.533 (0.000) | | | | | | | | | | | | | |
| Pod weight/plant | 19 | -0.160 (0.206) | 0.604 (0.000) | 0.619 (0.000) | -0.418 (0.000) | -0.138 (0.278) | 0.514 (0.000) | 0.456 (0.000) | 0.728 (0.000) | 0.431 (0.000) | 0.535 (0.000) | 0.701 (0.000) | 0.180 (0.156) | 0.489 (0.000) | 0.657 (0.000) | 0.287 (0.021) | 0.875 (0.000) | 0.691 (0.000) | 0.414 (0.001) | | | | | | | | | | | | |
| Pod length | 20 | -0.173 (0.171) | 0.233 (0.064) | 0.134 (0.292) | -0.189 (0.134) | -0.260 (0.038) | 0.463 (0.000) | 0.468 (0.000) | 0.303 (0.015) | 0.493 (0.000) | 0.416 (0.001) | 0.405 (0.001) | 0.291 (0.020) | 0.032 (0.800) | 0.086 (0.500) | 0.158 (0.213) | 0.171 (0.176) | 0.294 (0.018) | 0.342 (0.006) | 0.438 (0.000) | | | | | | | | | | | |
| Pod width | 21 | -0.176 (0.164) | 0.215 (0.089) | 0.121 (0.342) | -0.258 (0.040) | -0.286 (0.022) | 0.465 (0.000) | 0.411 (0.000) | 0.301 (0.015) | 0.466 (0.000) | 0.442 (0.000) | 0.350 (0.005) | 0.327 (0.008) | -0.025 (0.846) | 0.005 (0.968) | 0.134 (0.290) | 0.106 (0.405) | 0.219 (0.083) | 0.303 (0.015) | 0.387 (0.002) | 0.903 (0.000) | | | | | | | | | | |
| Pod length of double seeded | 22 | -0.178 (0.166) | 0.139 (0.281) | 0.071 (0.583) | -0.051 (0.694) | -0.116 (0.369) | 0.361 (0.004) | 0.366 (0.003) | 0.214 (0.096) | 0.356 (0.005) | 0.298 (0.018) | 0.424 (0.001) | 0.104 (0.423) | 0.010 (0.940) | 0.155 (0.229) | 0.261 (0.040) | 0.217 (0.090) | 0.481 (0.000) | 0.449 (0.000) | 0.453 (0.000) | 0.782 (0.000) | 0.621 (0.000) | | | | | | | | | |
| Pod width of double seeded | 23 | -0.032 (0.802) | 0.102 (0.428) | 0.136 (0.290) | 0.033 (0.800) | -0.224 (0.081) | 0.319 (0.011) | 0.296 (0.019) | 0.325 (0.010) | 0.355 (0.005) | 0.449 (0.000) | 0.320 (0.011) | 0.194 (0.131) | 0.092 (0.479) | 0.040 (0.758) | 0.033 (0.801) | 0.147 (0.253) | 0.264 (0.029) | 0.423 (0.038) | 0.840 (0.001) | 0.839 (0.000) | 0.661 (0.000) | (0.000) | | | | | | | | |
| Seed length | 24 | -0.189 (0.135) | 0.223 (0.076) | 0.056 (0.660) | -0.237 (0.060) | -0.278 (0.026) | 0.486 (0.000) | 0.485 (0.000) | 0.261 (0.037) | 0.475 (0.000) | 0.441 (0.000) | 0.237 (0.059) | 0.318 (0.011) | -0.022 (0.861) | 0.082 (0.681) | 0.086 (0.520) | 0.222 (0.498) | 0.236 (0.078) | 0.399 (0.061) | 0.803 (0.001) | 0.737 (0.000) | 0.644 (0.000) | 0.723 (0.000) | (0.000) | | | | | | | |
| Seed width | 25 | -0.389 (0.001) | 0.294 (0.018) | 0.051 (0.689) | -0.366 (0.003) | -0.437 (0.000) | 0.432 (0.000) | 0.513 (0.000) | 0.269 (0.031) | 0.536 (0.000) | 0.499 (0.000) | 0.312 (0.005) | 0.326 (0.000) | -0.003 (0.979) | 0.017 (0.894) | 0.193 (0.126) | 0.114 (0.371) | 0.241 (0.055) | 0.325 (0.009) | 0.423 (0.001) | 0.775 (0.000) | 0.750 (0.000) | 0.601 (0.000) | 0.734 (0.000) | 0.884 (0.000) | | | | | | |
| Seed no./plant | 26 | -0.145 (0.252) | 0.623 (0.000) | 0.703 (0.000) | -0.419 (0.001) | -0.077 (0.544) | 0.437 (0.000) | 0.405 (0.001) | 0.725 (0.000) | 0.343 (0.005) | 0.416 (0.001) | 0.718 (0.000) | 0.182 (0.149) | 0.519 (0.000) | 0.725 (0.000) | 0.332 (0.007) | 0.969 (0.000) | 0.698 (0.000) | 0.362 (0.003) | 0.910 (0.000) | 0.187 (0.139) | 0.109 (0.391) | 0.264 (0.038) | 0.153 (0.234) | 0.098 (0.441) | 0.146 (0.250) | | | | | |
| Seed weight/plant | 27 | -0.136 (0.285) | 0.645 (0.000) | 0.642 (0.000) | -0.446 (0.000) | -0.115 (0.367) | 0.485 (0.000) | 0.438 (0.000) | 0.745 (0.000) | 0.448 (0.000) | 0.523 (0.000) | 0.728 (0.000) | 0.149 (0.240) | 0.538 (0.000) | 0.713 (0.000) | 0.342 (0.006) | 0.881 (0.000) | 0.680 (0.000) | 0.449 (0.000) | 0.981 (0.000) | 0.382 (0.002) | 0.326 (0.008) | 0.389 (0.002) | 0.342 (0.007) | 0.374 (0.002) | 0.400 (0.001) | 0.920 (0.000) | | | | |
| Biomass dry weight | 28 | -0.099 (0.436) | 0.671 (0.000) | 0.667 (0.000) | -0.414 (0.000) | -0.010 (0.939) | 0.445 (0.000) | 0.394 (0.000) | 0.774 (0.000) | 0.441 (0.000) | 0.520 (0.000) | 0.754 (0.000) | 0.137 (0.279) | 0.565 (0.000) | 0.777 (0.000) | 0.371 (0.003) | 0.886 (0.000) | 0.644 (0.000) | 0.486 (0.000) | 0.948 (0.000) | 0.363 (0.002) | 0.310 (0.013) | 0.382 (0.002) | 0.308 (0.015) | 0.298 (0.009) | 0.323 (0.000) | 0.903 (0.000) | 0.956 (0.000) | | | |
| Shelling % | 29 | 0.052 (0.683) | 0.303 (0.015) | 0.232 (0.065) | -0.134 (0.291) | 0.115 (0.364) | -0.112 (0.380) | -0.030 (0.812) | 0.200 (0.114) | 0.064 (0.614) | -0.029 (0.821) | 0.184 (0.146) | -0.104 (0.412) | 0.244 (0.052) | 0.333 (0.007) | 0.279 (0.025) | 0.174 (0.169) | 0.049 (0.699) | 0.129 (0.309) | 0.066 (0.604) | -0.298 (0.017) | -0.334 (0.007) | -0.282 (0.026) | -0.392 (0.002) | -0.104 (0.412) | -0.125 (0.324) | 0.190 (0.133) | 0.234 (0.063) | 0.167 (0.187) | | |
| Growth habit | 30 | 0.083 (0.512) | -0.245 (0.051) | 0.051 (0.689) | 0.364 (0.003) | 0.295 (0.018) | 0.003 (0.978) | -0.140 (0.270) | -0.115 (0.366) | -0.427 (0.000) | -0.246 (0.050) | -0.615 (0.000) | 0.049 (0.702) | 0.067 (0.597) | -0.299 (0.017) | -0.793 (0.000) | -0.165 (0.191) | -0.349 (0.005) | -0.726 (0.000) | -0.216 (0.191) | -0.314 (0.026) | -0.245 (0.051) | -0.405 (0.001) | -0.155 (0.229) | -0.158 (0.214) | -0.253 (0.044) | -0.233 (0.064) | -0.243 (0.053) | -0.294 (0.018) | -0.100 (0.430) | |
| 100-seed weight | 31 | -0.060 (0.636) | 0.349 (0.005) | 0.129 (0.311) | -0.256 (0.041) | -0.148 (0.243) | 0.349 (0.005) | 0.317 (0.011) | 0.351 (0.004) | 0.444 (0.000) | 0.457 (0.000) | 0.287 (0.022) | 0.109 (0.390) | 0.188 (0.136) | 0.200 (0.113) | 0.175 (0.166) | 0.189 (0.135) | 0.224 (0.076) | 0.305 (0.014) | 0.524 (0.000) | 0.549 (0.000) | 0.564 (0.000) | 0.385 (0.002) | 0.503 (0.000) | 0.800 (0.000) | 0.726 (0.000) | 0.206 (0.102) | 0.550 (0.000) | 0.474 (0.000) | 0.242 (0.054) | -0.113 (0.375) |

***Biomass dry weight***: was found to be associated significantly with almost all other traits except the traits of *days to emergence* and *podding*, *stem no./plant*, shelling % and *growth habit*. Among the vegetative and growth stage traits *node no./stem* was correlated most significantly positively with biomass (r= 0.77; p= 0.000). A regression analysis suggested that 59.7% of the variation in *biomass dry weight* could be explained on the basis of *node no./stem* (Figure 4-6). Both *leaf area* and *plant spread* were followed *node no./stem* in their highest association with *biomass dry weight*, valued + 0.774 and +0.754, respectively (p= 0.000).

Pearson's correlation coefficient presented that the association of *biomass dry weight* with both *seed weight* and *pod weight* had recorded the highest positive value (r= +0.956 and 0.948; p= 0.000), respectively, among the yield component traits and explained 91.3% and 89.7% of the trait variation (Figure 4-6).
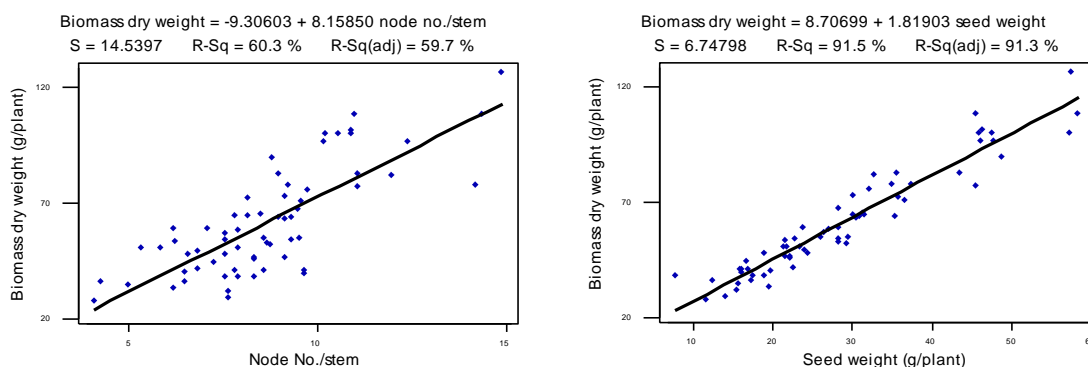


**Figure 4-6: A fitted line regression plot for *biomass dry weight* with both of *node no./stem* and *seed weight* in the F$_3$ population of glasshouse.**

## 4.2.3.2 Association of the traits for F$_3$ generation in field

Phenotypic relationship among the characters was studied and the coefficients are given in Table 4-5.

**Table 4-5: Relationship between traits in the F$_3$ population of the cross DipC x Tiga necaru for field data based on the Pearson's correlation analysis showing correlation and *P* values.**

Legend: r > 0.7 (yellow); r = 0.5-0.69 (green); r = 0.30 − 0.49 (blue)

| Traits | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days to emergence | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Leaf no./plant | 2 | -0.123 / 0.327 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Days to flowering | 3 | 0.284 / 0.034 | 0.025 / 0.855 | | | | | | | | | | | | | | | | | | | | | | | | |
| Plant height | 4 | -0.065 / 0.609 | 0.577 / 0.000 | -0.082 / 0.544 | | | | | | | | | | | | | | | | | | | | | | | |
| Petiole length | 5 | -0.136 / 0.288 | 0.734 / 0.000 | -0.028 / 0.837 | 0.818 / 0.000 | | | | | | | | | | | | | | | | | | | | | | |
| Leaf area | 6 | -0.038 / 0.767 | 0.909 / 0.000 | 0.006 / 0.964 | 0.724 / 0.000 | 0.858 / 0.000 | | | | | | | | | | | | | | | | | | | | | |
| Terminal leaflet length | 7 | -0.073 / 0.568 | 0.617 / 0.000 | -0.102 / 0.452 | 0.763 / 0.000 | 0.847 / 0.000 | 0.779 / 0.000 | | | | | | | | | | | | | | | | | | | | |
| Terminal leaflet width | 8 | -0.020 / 0.873 | 0.565 / 0.000 | 0.046 / 0.734 | 0.681 / 0.000 | 0.806 / 0.000 | 0.778 / 0.000 | 0.806 / 0.000 | | | | | | | | | | | | | | | | | | | |
| Plant spread | 9 | 0.016 / 0.898 | 0.808 / 0.000 | -0.020 / 0.885 | 0.691 / 0.000 | 0.766 / 0.000 | 0.879 / 0.000 | 0.770 / 0.000 | 0.678 / 0.000 | | | | | | | | | | | | | | | | | | |
| Stem no./plant | 10 | 0.005 / 0.967 | 0.019 / 0.880 | -0.406 / 0.002 | 0.072 / 0.580 | 0.100 / 0.445 | 0.015 / 0.909 | 0.144 / 0.265 | -0.040 / 0.755 | 0.032 / 0.807 | | | | | | | | | | | | | | | | | |
| Branch no./stem | 11 | -0.020 / 0.879 | 0.098 / 0.443 | -0.222 / 0.101 | 0.207 / 0.107 | 0.180 / 0.164 | 0.188 / 0.148 | 0.243 / 0.057 | 0.189 / 0.141 | 0.233 / 0.068 | 0.117 / 0.363 | | | | | | | | | | | | | | | | |
| Node no./plant | 12 | 0.179 / 0.165 | 0.146 / 0.253 | -0.249 / 0.064 | 0.277 / 0.029 | 0.221 / 0.086 | 0.214 / 0.098 | 0.267 / 0.036 | 0.262 / 0.040 | 0.148 / 0.250 | 0.318 / 0.011 | 0.421 / 0.001 | | | | | | | | | | | | | | | |
| Internode length | 13 | 0.068 / 0.604 | 0.079 / 0.541 | -0.308 / 0.022 | 0.099 / 0.446 | 0.107 / 0.416 | 0.085 / 0.516 | 0.265 / 0.039 | 0.161 / 0.215 | 0.171 / 0.188 | 0.228 / 0.074 | 0.154 / 0.232 | 0.519 / 0.000 | | | | | | | | | | | | | | |
| Growth habit | 14 | -0.083 / 0.533 | 0.259 / 0.045 | 0.350 / 0.010 | 0.161 / 0.218 | 0.241 / 0.064 | 0.212 / 0.107 | 0.010 / 0.937 | 0.040 / 0.763 | 0.133 / 0.312 | -0.155 / 0.236 | -0.293 / 0.023 | -0.416 / 0.001 | -0.797 / 0.000 | | | | | | | | | | | | | |
| Peduncle length | 15 | 0.245 / 0.051 | -0.051 / 0.687 | 0.173 / 0.197 | -0.129 / 0.305 | -0.234 / 0.063 | -0.058 / 0.650 | -0.153 / 0.225 | -0.142 / 0.259 | -0.120 / 0.342 | -0.160 / 0.216 | 0.200 / 0.119 | 0.049 / 0.704 | -0.046 / 0.724 | -0.080 / 0.541 | | | | | | | | | | | | |
| Pod no./plant | 16 | -0.254 / 0.044 | 0.059 / 0.645 | -0.339 / 0.010 | 0.040 / 0.755 | 0.113 / 0.382 | 0.103 / 0.424 | 0.099 / 0.438 | 0.208 / 0.101 | 0.099 / 0.440 | -0.049 / 0.702 | 0.244 / 0.054 | 0.061 / 0.633 | 0.331 / 0.009 | -0.336 / 0.009 | -0.116 / 0.365 | | | | | | | | | | | |
| Double seeded pods/plant | 17 | 0.017 / 0.899 | -0.106 / 0.415 | -0.253 / 0.063 | -0.051 / 0.697 | -0.130 / 0.328 | -0.042 / 0.751 | -0.129 / 0.326 | -0.154 / 0.241 | -0.025 / 0.849 | 0.067 / 0.608 | 0.276 / 0.031 | 0.105 / 0.421 | 0.066 / 0.617 | -0.268 / 0.042 | 0.111 / 0.400 | 0.315 / 0.013 | | | | | | | | | | |
| Pod weight/plant | 18 | -0.184 / 0.149 | 0.030 / 0.814 | -0.292 / 0.028 | 0.097 / 0.448 | 0.169 / 0.188 | 0.153 / 0.234 | 0.162 / 0.204 | 0.278 / 0.027 | 0.174 / 0.173 | -0.067 / 0.601 | 0.246 / 0.052 | 0.061 / 0.632 | 0.268 / 0.035 | -0.299 / 0.020 | -0.104 / 0.419 | 0.913 / 0.000 | 0.384 / 0.002 | | | | | | | | | |
| Pod length | 19 | 0.407 / 0.001 | 0.022 / 0.864 | 0.102 / 0.454 | 0.328 / 0.009 | 0.427 / 0.001 | 0.284 / 0.025 | 0.243 / 0.058 | 0.346 / 0.006 | 0.237 / 0.064 | -0.047 / 0.718 | 0.072 / 0.577 | 0.218 / 0.089 | -0.046 / 0.726 | -0.124 / 0.350 | -0.088 / 0.495 | 0.166 / 0.194 | 0.333 / 0.009 | 0.376 / 0.003 | | | | | | | | |
| Pod width | 20 | 0.147 / 0.255 | 0.086 / 0.503 | -0.105 / 0.439 | 0.301 / 0.017 | 0.339 / 0.008 | 0.219 / 0.087 | 0.229 / 0.073 | 0.369 / 0.003 | 0.262 / 0.040 | -0.042 / 0.746 | 0.039 / 0.765 | 0.216 / 0.092 | 0.281 / 0.028 | -0.243 / 0.064 | -0.152 / 0.238 | 0.464 / 0.000 | 0.083 / 0.531 | 0.580 / 0.000 | 0.676 / 0.000 | | | | | | | |
| Seed length | 21 | -0.102 / 0.451 | 0.268 / 0.042 | 0.214 / 0.123 | 0.320 / 0.014 | 0.401 / 0.002 | 0.348 / 0.008 | 0.173 / 0.193 | 0.384 / 0.003 | 0.362 / 0.005 | -0.093 / 0.491 | 0.110 / 0.416 | -0.042 / 0.759 | -0.132 / 0.332 | 0.192 / 0.156 | -0.092 / 0.491 | 0.284 / 0.030 | 0.089 / 0.520 | 0.477 / 0.000 | 0.539 / 0.000 | 0.553 / 0.000 | | | | | | |
| Seed width | 22 | -0.238 / 0.075 | 0.131 / 0.328 | 0.105 / 0.455 | 0.192 / 0.148 | 0.275 / 0.037 | 0.199 / 0.138 | 0.109 / 0.416 | 0.288 / 0.028 | 0.218 / 0.101 | -0.275 / 0.038 | 0.044 / 0.747 | -0.123 / 0.364 | 0.014 / 0.917 | 0.062 / 0.648 | -0.128 / 0.340 | 0.439 / 0.001 | 0.090 / 0.515 | 0.587 / 0.000 | 0.396 / 0.002 | 0.601 / 0.000 | 0.879 / 0.000 | | | | | |
| Seed no./plant | 23 | -0.077 / 0.564 | 0.065 / 0.625 | -0.367 / 0.007 | -0.006 / 0.967 | 0.075 / 0.577 | 0.059 / 0.665 | 0.076 / 0.569 | 0.204 / 0.124 | 0.116 / 0.385 | -0.079 / 0.558 | 0.177 / 0.183 | 0.112 / 0.402 | 0.418 / 0.001 | -0.376 / 0.004 | -0.064 / 0.633 | 0.858 / 0.000 | 0.312 / 0.019 | 0.831 / 0.000 | 0.264 / 0.046 | 0.537 / 0.000 | 0.304 / 0.022 | 0.474 / 0.000 | | | | |
| Seed weight | 24 | -0.081 / 0.544 | 0.084 / 0.526 | -0.331 / 0.016 | 0.051 / 0.706 | 0.154 / 0.249 | 0.126 / 0.349 | 0.171 / 0.200 | 0.282 / 0.032 | 0.187 / 0.160 | -0.083 / 0.535 | 0.186 / 0.163 | 0.116 / 0.387 | 0.392 / 0.003 | -0.339 / 0.011 | -0.069 / 0.605 | 0.830 / 0.000 | 0.318 / 0.017 | 0.894 / 0.000 | 0.411 / 0.001 | 0.635 / 0.000 | 0.460 / 0.000 | 0.601 / 0.000 | 0.947 / 0.000 | | | |
| Biomass dry weight | 25 | 0.051 / 0.680 | 0.392 / 0.001 | -0.041 / 0.763 | 0.758 / 0.000 | 0.817 / 0.000 | 0.769 / 0.000 | 0.947 / 0.000 | 0.761 / 0.000 | 0.746 / 0.000 | 0.096 / 0.460 | 0.249 / 0.051 | 0.166 / 0.196 | 0.138 / 0.289 | 0.045 / 0.732 | -0.124 / 0.327 | 0.041 / 0.751 | -0.021 / 0.876 | 0.148 / 0.246 | 0.358 / 0.004 | 0.207 / 0.107 | 0.202 / 0.129 | 0.127 / 0.341 | -0.030 / 0.824 | 0.077 / 0.560 | | |
| Shelling% | 26 | 0.052 / 0.696 | 0.233 / 0.075 | -0.192 / 0.169 | -0.098 / 0.463 | 0.025 / 0.851 | 0.089 / 0.512 | -0.017 / 0.899 | 0.161 / 0.229 | 0.190 / 0.154 | -0.169 / 0.203 | -0.014 / 0.917 | -0.133 / 0.320 | 0.167 / 0.216 | -0.073 / 0.592 | 0.075 / 0.576 | 0.153 / 0.247 | 0.057 / 0.674 | 0.164 / 0.216 | 0.023 / 0.862 | 0.280 / 0.033 | 0.385 / 0.003 | 0.464 / 0.000 | 0.504 / 0.000 | 0.515 / 0.000 | -0.131 / 0.322 | |
| 100-seed weight | 27 | -0.008 / 0.952 | 0.090 / 0.497 | -0.008 / 0.952 | 0.227 / 0.087 | 0.351 / 0.007 | 0.350 / 0.008 | 0.237 / 0.074 | 0.410 / 0.001 | 0.381 / 0.003 | -0.101 / 0.452 | 0.065 / 0.627 | -0.077 / 0.567 | -0.096 / 0.476 | 0.048 / 0.723 | -0.096 / 0.471 | 0.249 / 0.057 | 0.236 / 0.080 | 0.487 / 0.000 | 0.696 / 0.000 | 0.644 / 0.000 | 0.782 / 0.000 | 0.736 / 0.000 | 0.195 / 0.139 | 0.410 / 0.001 | 0.339 / 0.009 | 0.266 / 0.042 |

***Plant height*** **and** ***plant spread***: A significant positive correlation existed between them (r= +0.691; p=0.000) and both showed significant positive correlation with each of *petiole length*, *leaf area*, *terminal leaflet length* and *width*, *leaf no./plant pod width* and *seed length*. *Plant height* itself presented a very high positive correlation with *biomass dry weight* (Appendix 18). Most of these associations reflected results from data collected from the glasshouse.

***Internode length***: like the data from the glasshouse it had positive and significant relationship with *node no./plant, pod no./plant*, *pod weight*, seed no and *seed weight*. It also had significant and negative correlation with *growth habit* valued -0.797 at p= 0.000, while *internode length* explained 63% of the variation in the *growth habit* among the individuals of this population (Appendix 19). In addition, this trait was in a negative association with *days to flowering* (-0.308; p<0.05).

***Seed weight***: was positively and significantly associated with most of the other yield component traits and with *internode length* and *leaflet width* (Appendix 20). It also showed a significant negative correlation with *growth habit* (r= -339; p<0.05).

***Biomass dry weight***: As with the glasshouse experiment the coefficient analysis showed that biomass had a strong and positive correlation with some of vegetative growth traits included *leaf no./plant*, *plant height*, *plant spread*, *leaflet length, leaflet width, leaf area* and *petiole length*, among them a higher value of +0.947 was recorded in the association of *terminal leaflet length* with biomass at p= 0.000 level of significant. Regression analysis made for *biomass dry weight* accountable for 89.6% and 55% of the variations observed in *leaflet length* and *plant spread*, respectively (Figure 4-7). Among the yield-related traits biomass was only showed a significant

correlation with both of *pod length* and *100-seed weight* (r= +0.358 and +0.339; p= 0.004 and 0.009, respectively).
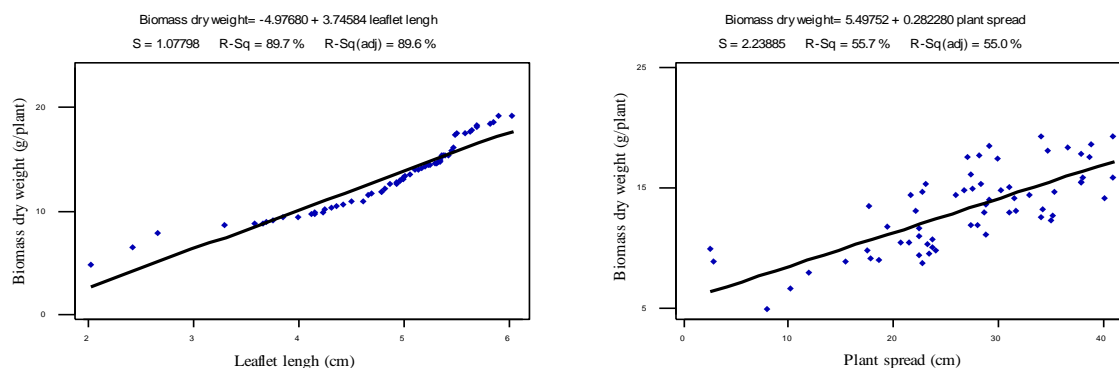


**Figure 4-7: A fitted line regression plot for *biomass dry weight* with both of *leaflet length* and *plant spread* in the F$_3$ population of the field.**

**The correlation between the same traits of different trials:**

There was a positive significant correlation between the data measured for the same traits of both glasshouse and the field experiments of the F$_3$ population derived from the narrow cross of DipC and Tiga necaru landraces (Table 4-6).

**Table 4-6: Correlation coefficient values for the same traits values of both glasshouse and the field trial of the F$_3$ population derived from the cross of DipC x Tiga necaru**

| No | Trait | Correlation value | P value |
|----|-------|-------------------|---------|
| 1 | Days to emergence | 0.410 | 0.001 |
| 2 | Leaf area | 0.283 | 0.030 |
| 3 | Terminal leaflet length | 0.349 | 0.006 |
| 4 | Terminal leaflet width | 0.328 | 0.011 |
| 5 | Plant spread | 0.285 | 0.030 |
| 6 | Internode length | 0.611 | 0.000 |
| 7 | Double seeded pods | 0.291 | 0.021 |
| 8 | Biomass dry weight | 0.331 | 0.009 |
| 9 | Seed weight | 0.320 | 0.016 |
| 10 | 100-seed weight | 0.304 | 0.024 |

### 4.2.3.3 Association of the traits for $F_2$ generation in TCRU

The result of associations between the traits of this generation can be found in Table 4-7.

**Table 4-7: Relationship between traits in the $F_2$ population of the cross DipC x Tiga necaru for TCRU data based on the Pearson's correlation analysis showing correlation and *P* values.**

| Traits | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Days to emergence* | 1 | | | | | | | | | | | | | | |
| *Leaf no./plant* | 2 | -0.107 0.366 | | | | | | | | | | | | | |
| *Days to flowering* | 3 | 0.609 0.000 | -0.199 0.091 | | | | | | | | | | | | |
| *Plant height* | 4 | 0.275 0.018 | 0.575 0.000 | 0.217 0.065 | | | | | | | | | | | |
| *Petiole length* | 5 | 0.125 0.301 | 0.673 0.000 | 0.200 .097 | 0.749 0.000 | | | | | | | | | | |
| *Terminal leaflet length* | 6 | -0.029 0.810 | 0.183 0.129 | -0.192 0.104 | 0.213 0.071 | 0.312 0.009 | | | | | | | | | |
| *Terminal leaflet width* | 7 | -0.064 0.593 | 0.495 0.000 | -0.219 0.063 | 0.350 0.002 | 0.447 0.000 | 0.752 0.000 | | | | | | | | |
| *Plant spread* | 8 | 0.055 0.641 | 0.635 0.000 | 0.049 0.681 | 0.615 0.000 | 0.832 0.000 | 0.547 0.000 | 0.590 0.000 | | | | | | | |
| *Pod no./plant* | 9 | 0.079 0.507 | 0.530 0.000 | 0.133 0.263 | 0.545 0.000 | 0.768 0.000 | 0.388 0.001 | 0.456 0.000 | 0.728 0.000 | | | | | | |
| *Double seeded pods/plant* | 10 | 0.032 0.790 | 0.407 0.000 | 0.099 0.412 | 0.329 0.005 | 0.508 0.000 | 0.328 0.005 | 0.332 0.005 | 0.653 0.000 | 0.615 0.000 | | | | | |
| *Pod weight/plant* | 11 | 0.126 0.288 | 0.632 0.000 | 0.198 0.094 | 0.655 0.000 | 0.836 0.000 | 0.276 0.018 | 0.410 0.000 | 0.800 0.000 | 0.909 0.000 | 0.702 0.000 | | | | |
| *Seed weight* | 12 | 0.166 0.159 | 0.648 0.000 | 0.228 0.053 | 0.677 0.000 | 0.840 0.000 | 0.264 0.024 | 0.396 0.001 | 0.788 0.000 | 0.888 0.000 | 0.690 0.000 | 0.990 0.000 | | | |
| *Biomass dry weight* | 13 | 0.114 0.336 | 0.636 0.000 | 0.203 0.083 | 0.665 0.019 | 0.854 0.000 | 0.273 0.000 | 0.411 0.000 | 0.807 0.000 | 0.918 0.000 | 0.666 0.000 | 0.987 0.000 | 0.973 0.000 | | |
| *Shelling%* | 14 | 0.304 0.009 | -.001 0.994 | 0.294 0.011 | 0.117 0.323 | 0.107 0.379 | -0.147 0.216 | -0.165 0.162 | -0.062 0.604 | -0.081 0.497 | -0.074 0.542 | 0.064 0.588 | 0.156 0.186 | 0.025 0.832 | |
| *100-seed weight* | 15 | 0.306 0.009 | 0.445 0.000 | 0.296 0.011 | 0.481 0.000 | 0.456 0.000 | -0.004 0.975 | 0.176 0.137 | 0.370 0.001 | 0.156 0.188 | 0.234 0.050 | 0.447 0.000 | 0.507 0.000 | 0.418 0.000 | 0.628 0.000 |

Legend:
r > 0.7 (yellow)
r = 0.5-0.69 (green)
r = 0.30 – 0.49 (blue)

***Days to emergence***: It was behaved as like as $F_3$ generation data grown in the glasshouse being in a positive and significant association with shelling percent and *100-seed weight*, however the significant associations with other traits was not recorded significantly.

***Plant spread***: Found to be associated significantly with other traits of *terminal leaflet length* and *width*, and with some yield related traits such as *pod no./plant, growth habit*, *biomass dry weight* and *100-seed weight*. The highest association of this trait

was recorded with *biomass dry weight* (r= +0.81 at p= 0.000) and the regression analysis between these two traits revealed an R-Sq (adj) value of 64.6% (Appendix 21). This result supported those for the $F_3$ populations in glasshouse and field trial.

***Seed weight*:** Was found to be correlated with *leaf no./plant*, *plant height*, *plant spread*, *pod no./plant*, *pod weight* and *biomass dry weight* (Appendix 22).

**4.3 Discussion**

The inheritance and the segregation patterns of morphological traits were investigated in the segregating $F_3$ population lines derived from an intra-subspecific cross between two domesticated landraces. DipC from Botswana was crossed with the wild accession (VSSP11). This investigation into the genetic relationships between wild and domesticated landraces is likely to reflect the genetic changes which occurred during domestication. The same landrace (DipC) was chosen as the female parent and crossed with the Tiga necaru landrace to construct a segregating population from a fully domesticated background. Tiga necaru from Mali was found to be from a different localized cluster using DArT markers in a genetic diversity analysis of bambara groundnut. This analysis suggested that the major molecular variation division was between West African and South/East African accessions (Bamlink annual report, 2008). The wide (although intra-subspecific) genetic differences between the parents of cross was also supported by another investigation studying the diversity in bambara groundnut using 201 DArT and 65 SSR markers (Molosiwa, 2012).

 The current study with a series of statistical investigations has described the nature of trait variation observed for a wide range of agronomical important traits within controlled crosses, as a preliminary to a full QTL analysis. This population was evaluated in glasshouse and field conditions to expose the population to a wider range of genotype x environment effects.

Reasonable numbers of morphological traits were evaluated in this study for the vegetative growth stage, flowering and yield development stages. The distribution of

different traits was tested and a Box-Cox transformation used to convert non-normally

distributed traits into normal distributions for QTL analysis.

### 4.3.1 Distribution of the traits

For all agro-morphological traits a distribution analysis was made using an Anderson-

Darling test. Most traits were shown to be normally distributed and consistent over

different environments for both $F_2$ and $F_3$ generations ( *flower no./plant*, *plant height*,

*petiole length*, *terminal leaflet length* and *width*, *plant spread*, *stem no./plant*, *node*

*no./plant*, *internode length*, *pod weight/plant*, *pod width*, *double seeded pods length*

and *width*, *shelling%* and *100-seed weight*). Distribution of these characters normally

indicates that they are more likely to be controlled by additive gene effect. Additive

gene effects were also reported earlier for *100-seed weight* and *shelling percentage* in

bambara groundnut (Karikari, 2000). Other traits (*days to emergence*, *days to*

*flowering*, *growth habit* and *double seeded pods/plant)* were non-normally distributed.

However, some of these traits (*leaf area*, *peduncle length*, *pod no./plant, double*

*seeded pods/plant, pod length*, *seed length* and *biomass dry weight)* did not show the

same distribution pattern over different trials or across different generations, Box-Cox

transformation pulled them back to a continuous and normally distribution. This

might be due to differences in inter-planting distances in the glasshouse and field

experiments or the effect of genotype x environment interaction. Jonah *et al.* (2012)

also reported that *seed length, pod length* and *pod width* are largely under additive

genetic control recording 100% broad sense heritability and predicted high genetic

advance of these traits by evaluating twelve cultivars of bambara groundnut sourced

from farmers collection in the north eastern Nigeria.

*Days to emergence* for both generations of $F_3$ and $F_2$ was characterized as a discontinuous trait, although it was not possible to determine a single gene effect as the trait did not match the expected pattern of Mendelian segregation ratio in both populations. In the glasshouse ($F_3$) data recorded for *days to emergence* only one individual line took longer to emerge than the DipC parent, however the difference in time to emergence between the parents of the cross did not exceed 4 days. In the $F_2$ population no difference was recorded for *days to emergence* of both parents. The result of this trait agrees with the findings of Basu *et al.* (2007c) studying segregation of this trait in an $F_2$ population derived from the cross of DipC x VSSP11.

Characterizing the number of flowers produced by the plant is laborious work, having to be done each 2-3 days for 3-4 months from the first day of flowering, this trait could be an important one in breeding terms. It was recorded for the $F_3$ population grown in glasshouse, and was found to follow a normal distribution, suggesting a multigenic trait.

Among the traits not distributed normally *growth habit* and *eye pattern around the hilum* were shown to follow a segregation pattern consistent with Mendelian inheritance. Incomplete dominance was hypothesised to control *growth habit* in the $F_3$ progenies in the glasshouse. *Growth habit* is one of the important morphological/agricultural traits and selection criteria derived in domestication and adaptation studies of bambara groundnut landraces have emphasized the study of this trait (Massawe *et al.*, 2005). In the analysis of the wide (inter-sub-specific cross; DipC x VSSP11; Basu *et al.*, 2007c), the F1 hybrid between the spreading spontenea type and the bunched *subterranea* type shows a morphology consistent with the wild spreading parent, suggesting that 'bunching' is a recessive mutation that has occurred during domestication.

147

*The presence of eye pattern around the hilum* was consistent with the hypothesis of control by complete dominance in the segregating $F_3$ population in the glasshouse. The monogenic inheritance of this trait was in agreement with what was detected previously in the segregation $F_2$ population of the wide cross (DipC x VSSP11) (Basu, 2005), however in the current study on the association of *eye pattern* with *testa color* was not confirmed as both parents of the current population possess creamy *testa color*, which made the following of the segregation for *testa colour* impossible.

### 4.3.2 Association of the traits

The characters studied here are most likely to have a complex relationship with each other. They mostly demonstrate substantial phenotypic variability and these characters could be used for breeding of this crop. A strong correlation was observed between yield and other traits studied here. The result was in agreement with previous reports (Ntundu *et al.*, 2006; Ouedraogo *et al.*, 2008). These correlations could serve as important criteria in breeding and selection of genotypes for this crop, using simpler or early traits which are correlated to later or more difficult to assess traits. Such correlations would need to be validated before use.

***Days to emergence***: Time required to initiate the first real leaf was positively correlated to the time taken for flowering in the $F_2$ and $F_3$ populations under study. In previous investigations it was also reported to show a positive but not significant correlation with *days to 50% flowering* (Nawab *et al.*, 2008). A negligible negative correlation was observed for *days to emergence* with *pod* and *seed yield*. A non-significant correlation of *days to emergence* with *seed weight* was recorded in the early and late planting date of bambara groundnut by Benedict and Michael (2011). However *days to emergence* had no direct correlation with yield in bambara

148

groundnut in the current study, despite the significant and positive correlation of *days to 50% emergence* with *pod yield* reported by researchers in Zimbabwe (Makanda *et al.*, 2009).

***Days to flowering*:** This trait had negative and significant association with most of the vegetative growth traits and also with *seed yield* and *biomass dry weight* in both glasshouse and field trial for $F_3$ population. These relationships suggest that spending more time in the vegetative stage for this material will reduce time spent in the reproductive phase, resulting in lower yields. A similar result of lower yield in semi arid-conditions due to a longer vegetative stage was reported in the evaluation of 310 bambara groundnut accessions in the Northern area of Burkina Faso (Ouedraogo *et al.*, 2008). It has been identified by Makanda *et al.* (2009) that poor yield in bambara groundnut is a function of delayed flowering. This association was not consistent in the $F_2$ population derived from the same cross as *days to flowering* was in a positive but not significant association with yield and its components.

***Flower no./plant*:** Among vegetative growth traits, there was a significant association of *flower number* with *plant spread* accounted for 32.1% [R-Sq(adj) = 32.1 %] of the trait variation observed in *plant spread*. Furthermore, the negative and significant correlation detected between *flower no./plant* and *days to flowering* support a hypothesis; early flowering gives the plant time to deliver the stored energy into the seeds instead of partitioning to other organs. This possibility was also supported through the significant and positive correlation of *flower no./plant* with *yield* and its components. A high positive correlation of *flower no./plant* and *seed weight* was reported previously (Benedict and Michael, 2011).

**Leaf dimensions:** The measurements of *terminal leaflet length* and *width* were found to be correlated positively to *seed yield* and its components for the $F_2$ and $F_3$ generations, at different levels of significance, under glasshouse and field conditions. Increased photosynthetic area resulting from greater leaf size may have contributed to increase agronomical performance. *Leaf length* and *width* participate in determining the spread of the plant canopy, so selection for these traits could be used to control soil evaporation by optimal canopy spread. Another investigation also suggests longer leaved plants in bambara groundnut produce more pods and *seed yield* (Misangu *et al.*, 2007).

*Plant spread***:** Had significant and positive associations with other vegetative traits, especially with leaf dimensions and this was consistent for both segregating populations over all three experiments. A positive correlation of *plant spread* with *biomass dry weight* and *seed weight* was observed for all the experiments, although the association with *seed weight* did not reach significance in the field trial data. This combination of associations makes plant spread to be a good indicator for yield of bambara groundnut, in this material at least. Based on a strong positive correlation between *plant spread* and *double seeded pods/plant* in the glasshouse and TCRU experiments, increasing the yield of bambara groundnut in this material could be achieved by selection of plants having greater canopy spread, with double seeded pods providing a boost to yield. A positive correlation of canopy spread with pod yield was also reported in adaptation studies of bambara groundnut conducted in Owerri southeastern Nigeria (Onwubiko *et al.*, 2011).

*Internode length***:** This trait is a major component of plant architecture as it accounts for the major differences between bunch and spreading types. It is one of the important traits responsible for the variability of landraces (Siise and Massawe, 2012).

150

*Internode length* was the main variable in an investigation assessing the level of diversity among 101 accessions to define collection and conservation strategies in bambara groundnut (Sévérin and Yao, 2011). In the current study correlations between *internode length* and other traits were consistent in both glasshouse and field trials for the $F_3$ population, *internode length* was also in a positive and significant association with other traits such as *node no.* and *pod no./plant*, *pod weight*, *seed no.* and *seed weight*. *Internode length* and *days to flowering* were negatively correlated. Thus higher yield could be obtained from plants having longer *internode length,* however selection efforts have worked to produce shorter *internode length* leading to a bunched habit (Smartt, 1985). It can be concluded that selection for this trait could result in better yield in this crop.

***Biomass dry weight***: Any improvement program for bambara groundnut should include high biomass production to emphasise bio-productivity and high harvest index to ensure partitioning biomass accumulation to the seed. *Biomass dry weight* had positive correlations with *plant spread* and the leaf measurements for the $F_3$ generation (glasshouse and the field) and the $F_2$ population grown in TCRU. The associations of other yield component traits with biomass dry weight were significant in the glasshouse and TCRU experiments, regression analysis suggested that nearly 90% of the variation in *pod* and *seed weight* could be explained on the basis of *biomass dry weight* for both experiments (glasshouse and TCRU).

***Seed yield***: In the current investigation the correlation of *100-seed weight* to the *seed yield* was consistent for both segregating populations over the three experiments. Ofori (1996) in a correlation analysis in bambara groundnut also identified positive correlation of *100-seed weight* with *seed yield*. Therefore, studying the trait of *100-seed weight* was considered valuable in the breeding program as it was found to be

among the important characters to be considered during selection of bambara groundnut (Karikari, 2000). *Pod no./plant* is an important component of *seed yield* in bambara groundnut, not surprisingly, and this has been noted through highly significant correlation between both traits in different environments. It has been concluded that the yield in bambara groundnut could be improved through selection for the *pod no./plant* (Adu-Dapaah and Sangwan, 2004).

# Chapter 5: CONSTRUCTION OF A GENETIC LINKAGE MAP IN BAMBARA GROUNDNUT

## 5.1 Introduction

A genetic linkage map represents the relative order of genetic markers along a chromosome and the relative distance between them determined by recombination frequency. Based on this relationship, the markers in a genetic map are placed into linkage groupings which describe the gene order and high density maps can often assign individual linkage groups a chromosomal identity (Yeboah *et al.*, 2007).

The construction of genetic linkage maps relies on the choice of parental lines to develop a controlled cross. The cross should be segregating for traits and reasonable levels of marker polymorphism are needed, which can be scored across the individuals of the population. It is possible to construct a genetic linkage map based on a number of different family structures; $F_2$, $F_3$, backcrosses, double haploid or Recombined Inbred Lines (RILs). Recombinational distances can be estimated through comparison of the number of recombinant genotypes for a pair of markers with the total number of observations (Wu *et al.*, 2008).

## 5.2 Genetic mapping in related legume crops

Bambara groundnut belongs to the Phaseoloid-Millettioid clade, which contains most legume crops, such as pea, alfalfa (*Medicago spp.*), chickpea, soybean and common bean (Choi *et al.*, 2004b). Investigating the genetic linkage map in related legume crops and emphasizing shared synteny among these species might help facilitate the identification of markers closely linked to traits of interest in bambara groundnut.

### 5.2.1 Soybean

The first soybean (*Glycine max* L. Merr.) ($2n=2x=40$) genetic linkage map of molecular markers was developed using 150 restriction fragment length polymorphism (RFLP) loci in an $F_2$ population derived from an inter-specific cross of *G. max* (A81-356022) x *G. soja* (PI468916) (Keim *et al.*, 1990).

Five soybean maps constructed from the crosses of 'Minsoy' x 'Noir 1', 'Minsoy' x 'Archer', 'Archer' x 'Noir 1', 'Clark' x 'Harosoy', and A81-356022 x PI468916 were combined into an integrated genetic map spanning 2,523.6 cM (Kosambi) map distance across 20 linkage groups. A total of 1,849 markers were used, including 1,015 SSRs, 709 RFLPs, 73 RAPDs, 24 classical traits, six AFLPs, ten isozymes, and 12 others. On average, 51 SSR markers mapped per linkage group, ranging from 35 to 64. The average length of the interval between any two adjacent SSR markers was 2.5 cM. It was suggested that more SSRs than expected were closely linked and clustering of SSR markers on the soybean map was observed, suggesting an association of genes and SSRs. The primer sequences for all SSR loci, as well as the genetic maps of each of the 20 consensus linkage groups can be found on the SoyBase Web site of the USDA, ARS Soybean Genome Database (http://soybase.agron.iastate.edu/) (Song *et al.*, 2004).

### 5.2.2 Cowpea

Omo-Ikerodah *et al.*, (2008) reported construction of a linkage map to screen for QTL to resistance to flower bud thrips (*Megalurothrips sjostedti*) in cowpea (*Vigna unguiculata* (L.) Walp) using 145 RIL individuals ($F_{10}$). The population was developed from a cross between two cowpea lines, 'Sanzi' (resistant to FTh) and 'VITA7' (susceptible to FTh) and developed by the single seed descent method. A

total of 134 AFLP and the 5 SSR markers were scored in the developed mapping population to facilitate QTL analysis, covering 1620.1 cM of the cowpea genome.

A consensus genetic linkage map was constructed for cowpea [*Vigna unguiculata* (L.) Walp.] ($2n=2x=22$) using a high-throughput EST-derived SNP assay. The markers were applied to 741 recombinant inbred lines from six mapping populations. Approximately 90% of the SNPs were technically successful, providing 1,375 dependable markers. Of these, 928 were integrated into a consensus genetic map spanning 680 cM with 11 linkage groups and an average marker distance of 0.73 cM. Recently a new consensus map containing 1107 EST-derived SNP markers (856 bins) was reported  by Lucas *et al.*, 2011, which was developed through integrating 13 population-specific maps, segregating for the most important traits in this crop. Eleven of these populations were $F_8$ to $F_{10}$ RIL populations developed by inbreeding and single seed descent while the remaining two were $F_3$–derived $F_4$ families. A total of 179 SNP markers were added to the previous mapping data set used by Muchero *et al.* (2009) and Xu *et al.* (2011). The improved methods of data analysis are realized in map characteristics when surveying synteny of cowpea with soybean, *Medicago truncatula* Gaertn, and *Arabidopsis thaliana* (L.) Heynh. Using HarvEST:Cowpea 1.27 (Wanamaker and Close, 2011) and Circos (Krzywinski *et al.*, 2009).

In the soybean and *M. truncatula* genomes, homologous genes were identified for 85% and 80% of the SNPs mapped in cowpea, respectively Development of such a highly robust genetic map is of value to develop projects, including genome assembly, marker assisted breeding, quantitative trait loci (QTL) analysis, map-based cloning, and comparative genomics. These results support the evolutionary close relationship between cowpea and soybean and identify regions for synteny-based functional genomics studies in legumes (Lucas *et al.*, 2011; Muchero *et al.*, 2009).

Another investigation was conducted by Pottorff *et al.*, 2012 to identify the candidate gene for leaf morphology in Cowpea [*Vigna unguiculata* (L.) Walp]. A RIL population of Sanzi (sub-globose leaf shape) x Vita 7 (hastate leaf shape) was used, showing considerable variation in leaf shape. A QTL for leaf shape, *Hls* (hastate leaf shape), was identified on the Sanzi x Vita 7 genetic map spanning from 56.54 cM to 67.54 cM on linkage group 15. The corresponding *Hls* locus was positioned on the cowpea consensus genetic map on linkage group 4, spanning from 25.57 to 35.96 cM. Synteny was examined using EST-derived SNP markers previously BLASTed and aligned to the soybean and *Medicago* genomes which are housed in the publicly available data set of HarvEST:Cowpea (http://harvest.ucr.edu). High co-linearity was observed for the syntenic *Hls* region in *Medicago truncatula* and *Glycine max*. One syntenic locus for *Hls* was identified on *Medicago* chromosome 7 Identifying 8 *Medicago* genes orthologous to cowpea SNP markers in that syntenic region of *Medicago* chromosome. While syntenic regions for *Hls* were identified on two soybean chromosomes, 3 and 19 (Figure 5-1). The *Hls* locus was identified on the cowpea physical map via SNP markers 1_0910, 1_1013 and 1_0992 which were identified in three BAC contigs; contig926, contig821 and contig25. The conserved gene order within the same legume family, cowpea, *Medicago* and soybean, enabled the identification of a candidate gene for the *Hls* locus (Pottorff *et al.*, 2012).
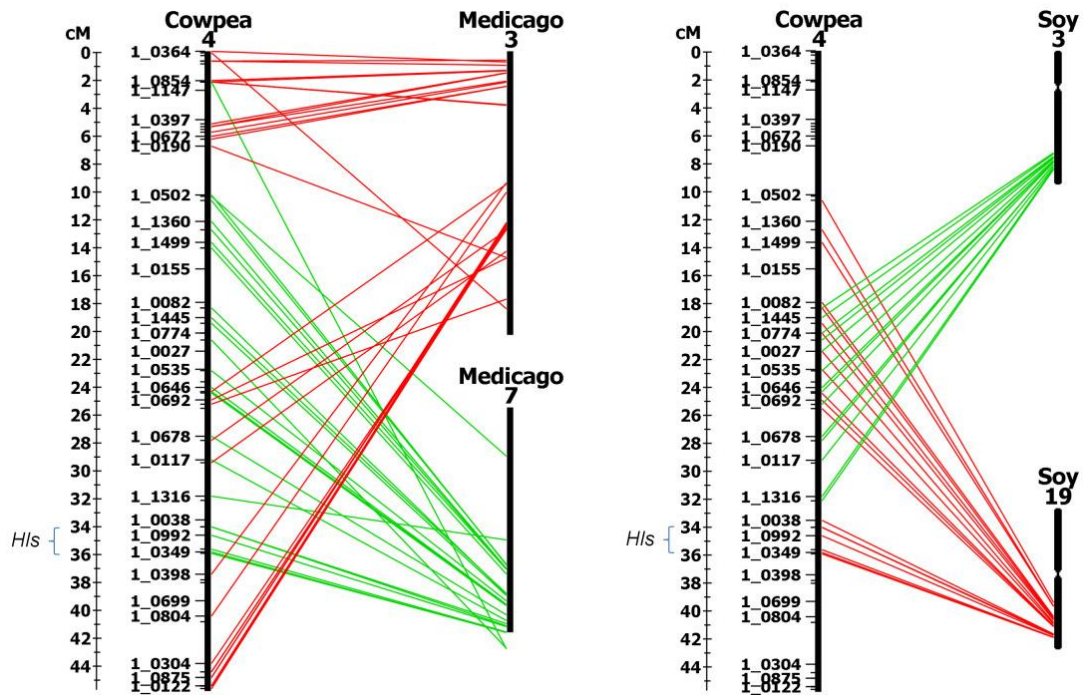
**Figure 5-1: Synteny of the *Hls* locus for Cowpea with *Medicago truncatula* and *Glycine max*, using EST-derived SNP markers previously BLASTed and aligned to the sequenced genomes. The *Hls* locus which is on linkage group 4 of the cowpea consensus genetic map was syntenic with *Medicago* chromosomes 3 and 7. Two syntenic loci were identified for the *Hls* locus in soybean chromosomes 3 and 19. The syntenic map was constructed using the HarvEST: Cowpea database (http://harvest.ucr.edu) with a cut off e-score value of −10 and a minimum number of 10 lines drawn per linkage group. Colored lines indicate cowpea genes orthologous to genes on *M. truncatula* and *G. max* chromosomes (Pottorff *et al.*, 2012).**

### 5.2.3 Chickpea

Chickpea (*Cicer arietinum* L.) is a self-pollinated, diploid ($2n=2x=16$), grain legume crop with a genome size of 740 Mb (Arumuganathan and Earle, 1991). A comprehensive genetic map was constructed in chickpea using a recombinant inbred line (RIL) population of 131 individuals derived from the cross of ICC 4958 (*C. arietinum*)×PI 489777 (*C. reticulatum*). A total of 253 bacterial artificial chromosome BES-SSR and 675 DArT polymorphic markers between the parental genotypes were used along with genotyping data published recently (192 genic molecular markers

(GMMs) including 83 conserved orthologous sequences (COS)-based SNPs, 54 cleaved amplified polymorphic sequences (CAPS), 35 conserved intron spanning region (CISR) and 20 EST-derived SSR) (Gujaria *et al.*, 2011). The map comprises 1,291 markers on eight linkage groups (LGs) spanning a total distance of 845.56 cM with an average inter-marker distance of 0.65 cM. This comprehensive genetic map with integrated BES-SSR markers will facilitate its anchoring to the physical map to accelerate map-based cloning of genes in chickpea and comparative genome evolution studies in legumes (Thudi *et al.*, 2011).

### 5.2.4 Lentil

Lentil is self-pollinated, diploid ($2n=2x=14$) with a large genome size of approximately 4Gb (Arumuganathan and Earle, 1991). A linkage map was constructed to improve understanding of the genetic structure. The linkage map of lentil (*Lens culinaris* L.) was constructed for a 94 RIL population (developed from the parents Precoz x WA8949041) using 166 markers. The map consisted of 11 linkage groups spanning 1396.3 cM of the lentil genome at a minimum LOD score of 3. The length of LGs varied from 16 cM to 436.8 cM with average marker density of 8.4 cM. The number of marker loci per LG ranged from 6 to 41. LG1 was the group containing the most markers with an average marker density of 10.6 cM in length. There was great variability in these densities as most of the RAPD and ISSR markers were evenly distributed across the genome while the majority AFLP markers were clustered on LG1. The authors suggested that the map could be used to detect agronomically important genes and in marker assisted selection of this crop. (Tanyolac *et al.*, 2010).

**5.2.5 Alfalfa**

The first genetic map was established for the model legume species of *Medicago truncatula* using an $F_2$ segregating population of 124 individuals. This map spanned 1225 cM with an average distance between adjacent markers of 470 kb/cM. It comprised 289 markers including RAPD, AFLP, known genes and isozymes arranged in 8 linkage groups. Markers were uniformly distributed throughout the map and segregation distortion was limited to only three linkage groups. Locating a number of common markers on the map, the eight linkage groups were shown to be homologous to those of diploid alfalfa (*Medicago sativa*), implying a good level of macrosynteny between the two genomes (Thoquet *et al.*, 2002).

Comparative genome analysis has been performed between two closely related tribes (Trifoliae and Viciae) of the subfamily Papilionoideae with different basic chromosome numbers. The linkage map of diploid alfalfa (*Medicago sativa*) derived from $F_2$ segregating population was compared to those of homologous loci on the combined genetic map of pea (*Pisum sativum*) derived from RIL populations. The analysis aimed to analyze the degree of colinearity between their linkage groups. The linkage groups of *Medicago* and pea showed a high degree of colinearity as was also demonstrated by other researchers (Choi *et al.*, 2004a). It was concluded that the difference in genome size between the two species (the pea genome is 5- to 10-fold larger than that of alfalfa) is not a consequence of genome duplication in the *Pisum* lineage. The high degree of synteny observed between pea and *Medicago* loci indicates that further map-based cloning of pea genes based on the available genome resources for *Medicago truncatula* and other related species will be productive which will enhance the strategy of transferring map information from this model legume to other related (Kalo *et al.*, 2004).

### 5.2.6 Mungbean

A linkage map was constructed in mungbean (*Vigna radiate* L.) ($2n=2x=22$) using an F$_2$ population (186 individuals) derived from an intra-specific cross between the wild accession 'W021' (*Vigna radiata* var. *sublobata*) and cultivated line 'KUML29-1-3' (*Vigna radiata* var. *radiata*). JoinMap 3.0 program was used to develop the linkage map. A minimum LOD score of 3.0 was used as a threshold value for grouping the markers and mapping used the Kosambi map function. One hundred and fifty SSR markers were assigned into 11 linkage groups with 5 markers at least for each. The map covered 1,174.2 cM of the mungbean genome having an average marker density of 7.8 cM. Extensive genomic conservation has been revealed by comparing this map with azuki bean (*Vigna angularis*) and blackgram (*Vigna mungo*) linkage maps based on azuki common markers (Kajonphol *et al.*, 2012).

### 5.2.7 Pigeon pea

The first genetic linkage map was developed for pigeon pea (*Cajanus cajan* spp. (L.) Mill sp.) ($2n=2x=22$) using a total of 554 diversity arrays technology (DArT) markers. The MAPMAKER/EXP version3 software was used to generate the map. The mapping population was an F$_2$ of 72 progenies derived from a cross between cultivated (*C. cajan*) genotype and its wild relative (*C. scarabaeoides*). Segregation distortion was 13%. Two groups of paternal and maternal genetic maps were generated with 172 and 122 unique marker loci, covering 451.6 cM and 270.0 cM, respectively. The authors state that those markers developed would be a useful genomic resource for inter-specific genetic analysis and the linkage map will allow map-based cloning of genes for biotic and abiotic stress from wild relatives of pigeon pea in the future (Yang *et al.*, 2011).

160

The genetic linkage maps presented above were derived from different inter and intra-specific crosses using different kinds of molecular markers in legume crops to identify QTLs for the traits interested. Some of the maps constructed were proposed to be used in comparative genome analysis with the other non model legumes crops. The results also suggested close relationships between some of these legumes by identify a high degree of synteny. This could allows to identify markers and cloning genes and facilitate the strategy of transferring genomic information from this model legume such as soybean (*Glycine max*) and *Medicago truncatula* to other non-model related species such as bambara groundnut.

The aims of the research reported in this chapter were firstly to construct 'narrow' and 'wide' genetic linkage maps of bambara groundnut (*Vigna subterranea* (L.) Verdc.) to clarify the underlying genetic mechanisms controlling agronomic traits and facilitate MAS in future breeding programs, then combining these two maps to identify common markers in an attempt to build a consensus map. Two inter and intra-subspecific mapping populations were used. The first one was an $F_3$ segregating population derived from the cross of two domesticated landraces (DipC x Tiga necaru) based on SSR and DArT markers, while the second was a segregating $F_2$ population derived by crossing *V. subterranea* var. *subterranea* x *V. subterranea* var. *spontanea*. Three marker types of SSR, DArT and AFLP were used in constructing the second map.

**5.3 Result**

In this chapter the analysis of polymorphic markers, their segregation in the population under study and linkage map construction are presented for both 'narrow' and 'wide' maps. In addition, an attempt to combine the maps resulting from these two constructed maps is also presented here.

**5.3.1 Polymorphism and markers for mapping**

*Polymorphic markers in a segregating $F_3$ population of the 'narrow cross'*

Out of 124 SSR markers developed from the genome and transcriptome of bambara groundnut and from soybean (one marker) 45 primers were polymorphic (36.3%) and discriminated between the parental alleles of the cross. From this number 33 primers were used to screen all individuals in this population and in the construction of the genetic map. As there was not enough time or money, the other 12 polymorphic markers were not mapped in the population. Of the 7680 features present in the DArT slide array, 236 (3.1%) were identified as polymorphic markers in the cross and scored in the narrow cross population.

In total 269 polymorphic SSR and DArT markers were used to construct the initial map and showed linkage to at least one other marker using the segregating $F_3$ population derived from narrow cross of DipC x Tiga necaru landraces.

*Polymorphic markers in the segregating $F_2$ population of 'wide cross'*

The same primer sets of 124 SSR markers were tested for polymorphism in this population and 41 polymorphic primers were detected (33.1%). Parental alleles of this cross were scored for all the individuals using only 22 polymorphic SSR primers due

to shortage of time and money. These primers were combined with other sets previously developed for this population (Basu, 2005). Previously developed primers included 136 DArT, 141 AFLP and 2 SSR primers. In total, 301 polymorphic SSR, DArT and AFLP were used in linkage analysis and map construction using a segregating $F_2$ population derived from the cross of cultivated landrace (DipC) and wild accession (VSSP11).

## 5.3.2 Inheritance and the segregation distortion of markers

Marker segregation patterns and their potential distortion was detected automatically by JoinMap4, performing a Chi-square test against expected segregation patterns (p<0.05 for significance). In the narrow cross the segregation of markers was estimated from the second map run (RIL3 model) for phased markers (239). The locus genotype frequency table (Appendix 23) suggested 163 markers segregated in the expected Mendelian ratio of 3:2:3 and 3:5 for both marker types SSR and DArT, respectively, in the $F_3$ population. Seventy-six markers (32%) showed segregation distortion and did not segregate according to expected Mendelian inheritance. The segregation distortion in the DArT markers was found to be greater compared to SSRs (Table 5-1).

In the wide cross map 73% (218) of the markers evaluated segregated in the expected Mendelian ratio (Appendix 24). Eighty three markers from all marker types (SSRs, DArTs and AFLPs) were distorted (P≤ 0.05, chi-square test). The highest distortion ratio (36%) was among dominant AFLP markers for this map followed by SSR markers (29%) and the lowest distortion (19%) was observed for DArT markers (Table 5-1). The results of linkage analysis for both maps revealed that markers with distorted segregation were distributed throughout the genome.

**Table 5-1: Markers and their segregation patterns in both population types of narrow and wide cross of bambara groundnut.**

| Type of marker | $F_3$ population of narrow cross | | | $F_2$ population of wide cross | | |
|---|---|---|---|---|---|---|
| | Mendelian Segregation ratio ($x^2$ at p<0.05) | Polymorphic marker | Distortion | Mendelian Segregation ratio ($x^2$ at p<0.05) | Polymorphic marker | Distortion |
| SSR | 3:2:3 | 29 | 7 (24%) | 1:2:1 | 24 | 7 (29%) |
| DArT | 3:5 | 210 | 69 (33%) | 1:3 | 136 | 26 (19%) |
| AFLP | - | - | - | 1:3 | 139 | 50 (36%) |
| Co-dominant AFLP | - | - | - | 1:2:1 | 2 | 0 |
| Total markers | - | 239 | 76 (32%) | - | 301 | 83 (27%) |

## 5.3.4 Linkage phase determination

With no information available on the parents' genotypes (DipC and Tiga necaru) in segregating $F_3$ population for DArT markers, phase was determined through initial analysis of the population as a Cross Pollinator (CP). Two hundred and thirty-six DArT markers were combined with 33 SSR markers, grouped and initial linkage maps determined. Out of 269 markers 239 grouped into 28 linkage groups, with their phases determined by JoinMap4. CP segregation type coding of <hkxhk> for all loci was converted into the RIL appropriate code, dependant on the marker type and the phase (a,h,b) = codominant markers , (a,c) = dominant coding where phase was determined as (0,0) and (b,d) = dominant coding where phase was determined to be (1,1). The .loc file recoded as a, b, c, d, h was used for linkage determination using an RIL3 model, producing 21 linkage groups of 238 markers. The other 31 unmapped markers were presented in Appendix 25.

## 5.3.5 Calculation of the map

The JoinMap4 program builds the map by adding loci one by one, starting with the pair for which there is the strongest linkage evidence. For each added locus the best

position is determined by comparing the goodness-of-fit of the resulting map for each tested position. When at the best position, the overall goodness-of-fit (measured as a cumulative chi-square score) should be lower when compared with other possible positions. This position is adopted, although the final position may still lead to an increase in overall group cumulative chi-sqr score which leads to an initial rejection of the new marker (the default '*jump*' threshold is set at an increase in overall cumulative chi-sqr of 5). The locus is removed if the jump score increases above the default value of +5 (suggesting a conflict with the markers already mapped) or if strong negative distances are observed. The process is continued until all loci are tested once. This is the end of the first round resulting in Map1. Subsequently, a second attempt is made to add the previously removed loci to the map. This can be successful since the map will contain more loci than at the first attempt because of more pairwise data being used and testing local order may lead to changes in marker order. If high jump values or negative distances are still not resolved for the problematic loci in Map2, they will be removed again. This is the end of the second round resulting in Map 2. After that, all loci previously removed are added back to the map in the best possible positions, ignoring both jump and negative distance thresholds. This results in a final or third round of mapping giving rise to Map3. In this process each map is calculated using the pairwise data of loci present in the map, but only those pairs of loci that have a recombination frequency smaller than the REC (Recombination) threshold (0.4 default) and a LOD value larger than the LOD threshold (1.0 default) are used.

Constructing the linkage map using the narrow cross population resulted in performing more than one round of mapping (Map1) to build the map for the groups of 1, 5, 6 and 11. All of these groups except 11 had no significant conflicts and the

'jumped' markers from round 1 could be added in the second round (Map2) without a need to relax criteria. For group 11, because the third map for this group had shown a significant conflict between existing marker data from round two and the additional 'jumped' markers, the first round map was adopted. Note, the second round map did not lead to additional markers being added to Map2.

In the map constructed from the wide cross population the maps of linkage groups 1, 6, 7, 13, 14, 15, and 16 had three rounds of map, but the round one (Map1) was followed from these groups to assemble the linkage map in the segregating $F_2$ population. The large threshold value for *jump* in the second (Map2) and the third maps (Map3) of these groups indicated the poor fit of the added markers and all added markers (10) for these rounds were removed from the map (Table 5-2). These removed markers and those unmapped on the linkage maps of 'wide' map were present in Appendix 26.

**Table 5-2: Map rounds for the linkage groups of 'narrow ' and 'wide' cross maps with the identified adopted rounds in the genetic maps.**

| Linkage groups | 'Narrow' cross map | | 'Wide' cross map | |
|---|---|---|---|---|
| | No of round (Map) | Map used | No of round (Map) | Map used |
| 1 | 2 | 2 | 3 | 1 |
| 2 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 |
| 5 | 2 | 2 | 1 | 1 |
| 6 | 2 | 2 | 3 | 1 |
| 7 | 1 | 1 | 3 | 1 |
| 8 | 1 | 1 | 1 | 1 |
| 9 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 |
| 11 | 3 | 1 | 1 | 1 |
| 12 | 1 | 1 | 1 | 1 |
| 13 | 1 | 1 | 3 | 1 |
| 14 | 1 | 1 | 3 | 1 |
| 15 | 1 | 1 | 3 | 1 |
| 16 | 1 | 1 | 3 | 1 |
| 17 | 1 | 1 | 1 | 1 |
| 18 | 1 | 1 | 1 | 1 |
| 19 | 1 | 1 | 1 | 1 |
| 20 | 1 | 1 | 1 | 1 |
| 21 | 1 | 1 | - | - |

### 5.3.6 Linkage map and marker distribution

Map construction was conducted with JoinMap4. It started with the calculation of LOD groupings (tree) and manually selection using a LOD score of 5.0 coming down to 3.0. Two hundred and sixty nine markers of SSR and DArT were employed in the linkage analysis of narrow cross map and the Haldane mapping function was used.

Initially map construction was setup to analyse the population under study as a Cross Pollinator (CP) to determine linkage phase for these markers, as the parental data was not available by the time of running slide-based DArT markers. Out of 269 markers (SSRs and DArTs) 239 were assigned to 28 linkage groups. The genotypes of DArT markers for the individuals were converted to different phases for the markers in linkage groups. Different classifications was arranged depends on their phases; a or c gave to {0,0} phase and b or d for {1,1} phase in an 'RIL3' population type.

All arranged markers (29 SSRs and 210 converted DArTs) were grouped again with JoinMap4 using 'independence LOD'. All markers except one were assigned to 21 linkage groups (Figure 5-2).

The 21 linkage groups spanned 608.6 cM of bambara groundnut genome in this map. The distance between two consecutive markers varied from 0-10.1 cM, with a mean of 3 cM. The number of markers per linkage group ranged from 2 to 46 markers with an average length of 28.98 cM. The 21 linkage groups spanned 608.6 cM of bambara groundnut genome in this map. The distance between two consecutive markers varied from 0-10.1 cM, with a mean of 3 cM. The number of markers per linkage group ranged from 2 to 46 markers with an average length of 28.98 cM. The longest group with 23 markers covered a distance of 76.4 cM.

**LG1**

| Position | Marker |
|---|---|
| 0.0 | bgPabg-597086 |
| 1.7 | bgPt-596563 |
| 3.2 | bgPabg-596075 |
| 6.6 | bgPabg-593892 |
| 7.9 | bgPabg-423334 |
| 8.1 | bgPabg-423430 |
| 8.4 | bgPabg-423067 |
| 8.6 | bgPabg-423727 |
| 8.8 | bgPabg-423390 |
| 9.1 | bgPabg-423053 |
| 9.2 | bgPabg-422590 |
| 9.3 | bgPabg-423378 bgPabg-423556 |
| | bgPabg-422473 |
| | bgPabg-422586 bgPabg-423147 |
| 9.5 | bgPabg-422294 |
| 9.6 | bgPabg-423618 bgPabg-423248 |
| 9.7 | bgPabg-422657 bgPabg-422618 |
| 9.8 | bgPabg-423122 bgPabg-422379 |
| 9.9 | bgPabg-423557 |
| 10.0 | bgPabg-423395 bgPabg-422619 |
| 10.2 | bgPabg-423420 bgPabg-423121 |
| 10.3 | bgPabg-422656 |
| 10.6 | bgPabg-423708 |
| 10.7 | bgPabg-422458 |
| 10.8 | bgPabg-422461 |
| 11.2 | bgPabg-422515 |
| 15.2 | bgPabg-423185 |
| 25.0 | bgPabg-600828 |
| 28.3 | bgPt-600790 |
| 28.9 | bgPt-602039 |
| 33.0 | bgPabg-596774 |
| 39.1 | bgPabg-593965 |
| 44.8 | bgPt-595005 |
| 52.7 | GH-19-B2-D9 |
| 54.8 | PRIMER26 |
| 57.4 | bgPabg-595666 |
| 59.6 | bgPabg-596618 |
| 61.0 | mBam3co7 |
| 72.6 | bgPt-601022 |

**LG2**

| Position | Marker |
|---|---|
| 0.0 | bgPt-602268 |
| 6.3 | bgPt-422567 |
| 10.2 | bgPt-597731 |
| 13.3 | bgPt-598594 |
| 18.8 | bgPt-600827 |
| 21.4 | bgPt-600818 |
| 23.2 | bgPt-598598 |
| 25.2 | bgPt-598513 |
| 27.5 | bgPt-600914 |
| 34.2 | bgPt-597457 |
| 36.6 | bgPt-598621 |

**LG3**

| Position | Marker |
|---|---|
| 0.0 | bgPabg-596210 |
| 1.9 | bgPabg-595315 |
| 2.4 | bgPabg-595273 |
| 15.6 | bgPt-597113 |
| 19.7 | bgPt-600935 |
| | bgPabg-597705 bgPabg-597860 |
| 30.2 | bgPabg-597746 bgPabg-595707 |
| 30.4 | bgPabg-594814 |
| | bgPabg-594981 |
| 37.5 | bgPabg-422614 |
| 37.6 | bgPabg-422333 |
| 38.0 | bgPabg-596646 |
| 38.1 | bgPabg-423257 |

**LG4**

| Position | Marker |
|---|---|
| 0.0 | BN6b |
| 2.4 | bgPt-423527 |
| 3.0 | bgPabg-596988 bgPt-600823 |
| 3.5 | bgPt-598611 |
| 3.7 | bgPabg-598611 |
| 3.9 | bgPabg-423527 |
| 5.8 | BN259 |
| 8.6 | bgPt-596916 |
| 8.9 | bgPabg-593893 |
| 11.2 | bgPt-600898 |
| 22.7 | PRIMER19 |
| 33.5 | bgPabg-597624 |

**LG5**

| Position | Marker |
|---|---|
| 0.0 | bgPt-596933 |
| 0.3 | bgPt-423050 |
| 0.4 | bgPt-601086 |
| 0.9 | bgPabg-601086 |
| 1.6 | bgPabg-423807 |
| 2.5 | bgPabg-595038 |
| 2.6 | bgPabg-596655 |
| 3.3 | bgPabg-594508 |
| 3.4 | bgPabg-596933 |
| 36.9 | PRIMER2 |
| 40.1 | PRIMER95 |
| 42.5 | PRIMER45 |
| 44.8 | bgPt-601401 |
| 45.4 | PRIMER98 |
| 45.6 | bgPt-598647 |
| 46.8 | bgPt-597667 |
| 46.9 | bgPt-423805 |
| 47.4 | bgPt-601656 |
| 49.2 | bgPabg-422621 |
| 59.2 | bgPt-597491 |
| 65.2 | bgPt-601091 |
| 65.7 | bgPt-596460 |
| 65.8 | bgPt-593939 bgPt-598122 |
| 66.0 | bgPt-596596 |
| 66.1 | bgPt-602174 |
| 67.0 | PRIMER38 |
| 74.2 | bgPt-595387 |

**LG6**

| Position | Marker |
|---|---|
| 0.0 | bgPt-598407 |
| 0.3 | bgPt-598212 |
| 1.0 | bgPt-598809 |
| 9.0 | bgPt-598428 |
| 10.1 | bgPt-598184 |
| 11.1 | bgPt-597632 |
| 11.2 | bgPabg-597632 bgPabg-598184 |
| | bgPabg-598428 |
| 11.4 | bgPabg-601173 |
| 17.4 | bgPt-595129 |
| 18.4 | bgPabg-597436 |
| 20.9 | bgPabg-594142 |

**LG7**

| Position | Marker |
|---|---|
| 0.0 | PRIMER103 |
| 4.2 | bgPt-598507 bgPt-601098 |
| 4.4 | bgPt-598641 |
| 4.5 | bgPt-598456 |
| 4.6 | bgPabg-594467 |
| 5.1 | bgPabg-596040 |
| 8.4 | BamcoL17 |
| 10.3 | bgPt-594335 |
| 10.5 | bgPt-601852 |
| 12.3 | bgPabg-594537 |
| 13.3 | bgPabg-594335 |

**LG8**

| Position | Marker |
|---|---|
| 0.0 | Bam2coL63 |
| 5.5 | PRIMER37 |
| 10.6 | bgPabg-595799 |
| 10.8 | bgPabg-595118 |
| 16.8 | bgPabg-597688 |

**LG9**

| Position | Marker |
|---|---|
| 0.0 | bgPt-596705 |
| 1.5 | bgPt-597688 |
| 1.6 | bgPt-602071 |

**LG10**

| Position | Marker |
|---|---|
| 0.0 | bgPabg-594877 |
| 0.3 | bgPt-594877 |
| 10.8 | PRIMER66 |
| 19.9 | bgPt-601402 |
| 24.9 | bgPt-596392 |
| 30.4 | bgPabg-595854 |
| 33.8 | bgPabg-596250 |
| 35.8 | bgPt-596528 |
| 36.3 | bgPt-595011 |
| 36.4 | bgPt-594258 |
| 49.1 | bgPabg-593983 |
| 49.7 | bgPabg-596546 |
| 56.2 | bgPt-594658 |
| 58.6 | bgPabg-597858 |
| 59.3 | PRIMER48 |
| 59.6 | bgPt-597858 |
| 59.7 | bgPt-598651 |
| 60.0 | bgPt-596444 |
| 60.4 | PRIMER32 |
| 61.3 | bgPt-594435 |
| 63.1 | BN145 |
| 70.1 | bgPabg-596205 |
| 76.4 | bgPabg-423332 |

**Figure 5-2: A genetic linkage map of 21 linkage groups. This was constructed in 73 F₃ individuals derived from the cross between individuals of the DipC and Tiga necaru landraces. The locations of 29 SSR and 209 DArT markers are given. Positions are given in centimorgan (Haldane units) to the right of the linkage groups and the name of markers to the left. Microsatellite markers were highlighted.**

The final coverage of 608.6 cM was expected to equate to about 54 % of bambara groundnut genome, based on a default 100cM per chromosome and 11 chromosomes which the genome size (*G*) has been estimated from the partial linkage data according to the method of Hulbert *et al.* (1988):

$$G = \frac{n(n-1)}{2} \cdot \frac{2x}{yx}$$

Where *n* is the number of mapped markers, *yx* is the number of two-point linkages at

a distance equal to *x* cM. This estimation applies directly when the markers are all

informative in the same number of meioses in inbreeding species. Table 5-3 presents a

summary of these results, with all groups adopting Map1 unless otherwise indicated.

**Table 5-3: Distribution of the markers, linkage group size and marker density in the genetic map constructed in a $F_3$ population of DipC x Tiga necaru cross.**

| Linkage groups | Length (cM) | No. Of markers mapped in the groups | | | Average marker interval (cM) |
|---|---|---|---|---|---|
| | | Total marker | SSR marker | DArT marker | |
| 1 (Map2) | 72.6 | 46 | 3 | 43 | 1.58 |
| 2 | 36.6 | 11 | 0 | 11 | 3.33 |
| 3 | 38.1 | 15 | 0 | 15 | 2.54 |
| 4 | 33.5 | 13 | 3 | 10 | 2.58 |
| 5 (map2) | 74.2 | 28 | 5 | 23 | 2.65 |
| 6 (map2) | 20.9 | 13 | 0 | 13 | 1.61 |
| 7 | 13.3 | 12 | 2 | 10 | 1.11 |
| 8 | 16.8 | 5 | 2 | 3 | 3.36 |
| 9 | 1.6 | 3 | 0 | 3 | 0.52 |
| 10 | 76.4 | 23 | 4 | 19 | 3.32 |
| 11 | 15.3 | 14 | 1 | 13 | 1.10 |
| 12 | 47.5 | 10 | 1 | 9 | 4.75 |
| 13 | 18.0 | 9 | 1 | 8 | 2.01 |
| 14 | 3.1 | 9 | 1 | 8 | 0.35 |
| 15 | 42.9 | 7 | 1 | 6 | 6.12 |
| 16 | 8.0 | 5 | 1 | 4 | 1.60 |
| 17 | 23.5 | 4 | 2 | 2 | 5.87 |
| 18 | 34.4 | 4 | 1 | 3 | 8.60 |
| 19 | 30.2 | 3 | 1 | 2 | 10.07 |
| 20 | 0.0 | 2 | 0 | 2 | 0.00 |
| 21 | 1.4 | 2 | 0 | 2 | 0.71 |
| Total | 608.61 | 238 | 29 | 209 | - |
| Range | 0- 77.43 | 2-46 | 1-5 | 2-45 | 0-10.1 |

The wide cross map was constructed using three marker types; SSR, DArT and AFLP.

A total of 301 markers were subjected to linkage analysis, from which 194 markers

were assigned to 20 linkage groups (all groups adopting Map1). The mean marker-marker distance was 5.9cM. LG1 included the largest number of DArT markers (43). The greatest distance (109 cM) was observed for LG15 with 15 markers (Figure 5-3).
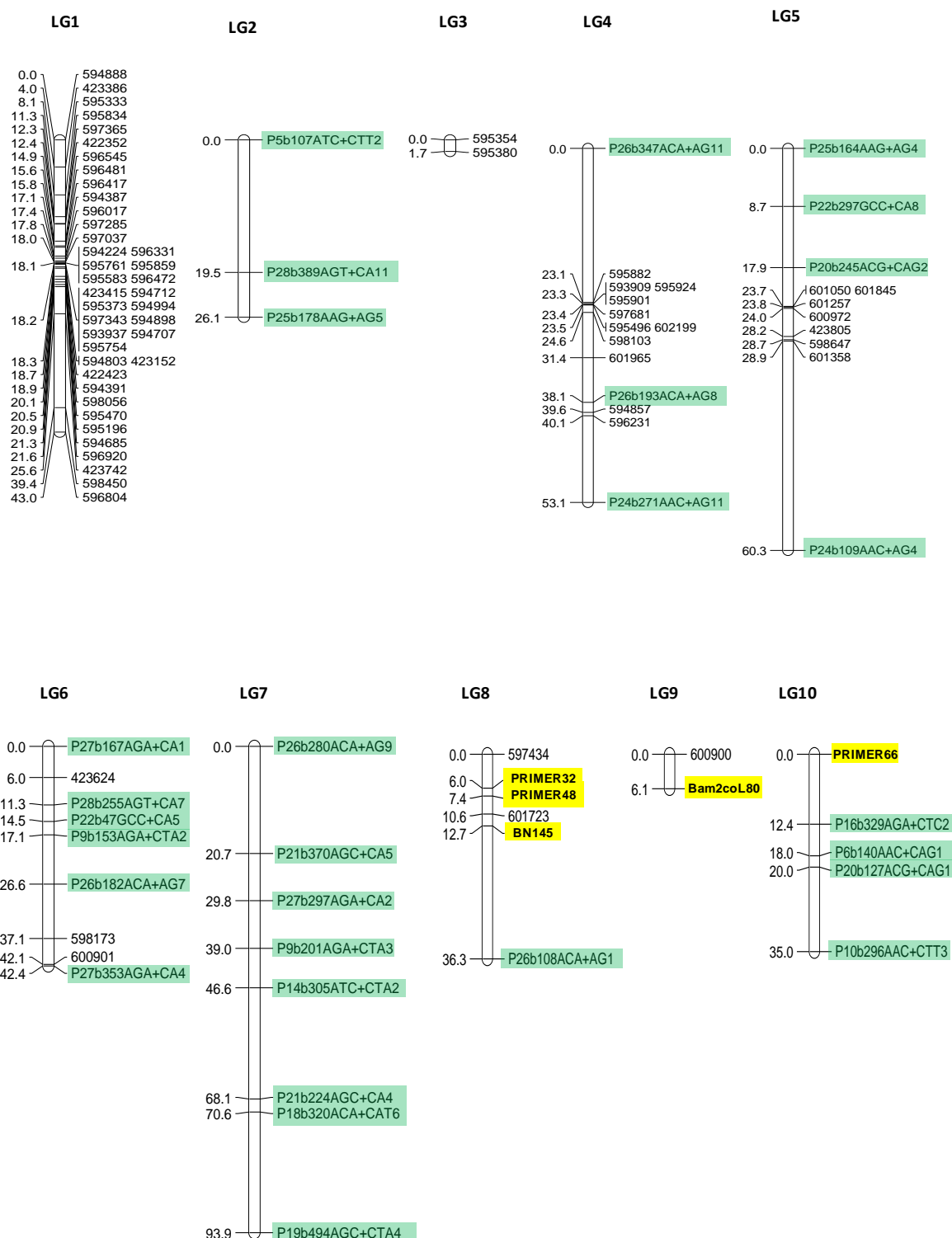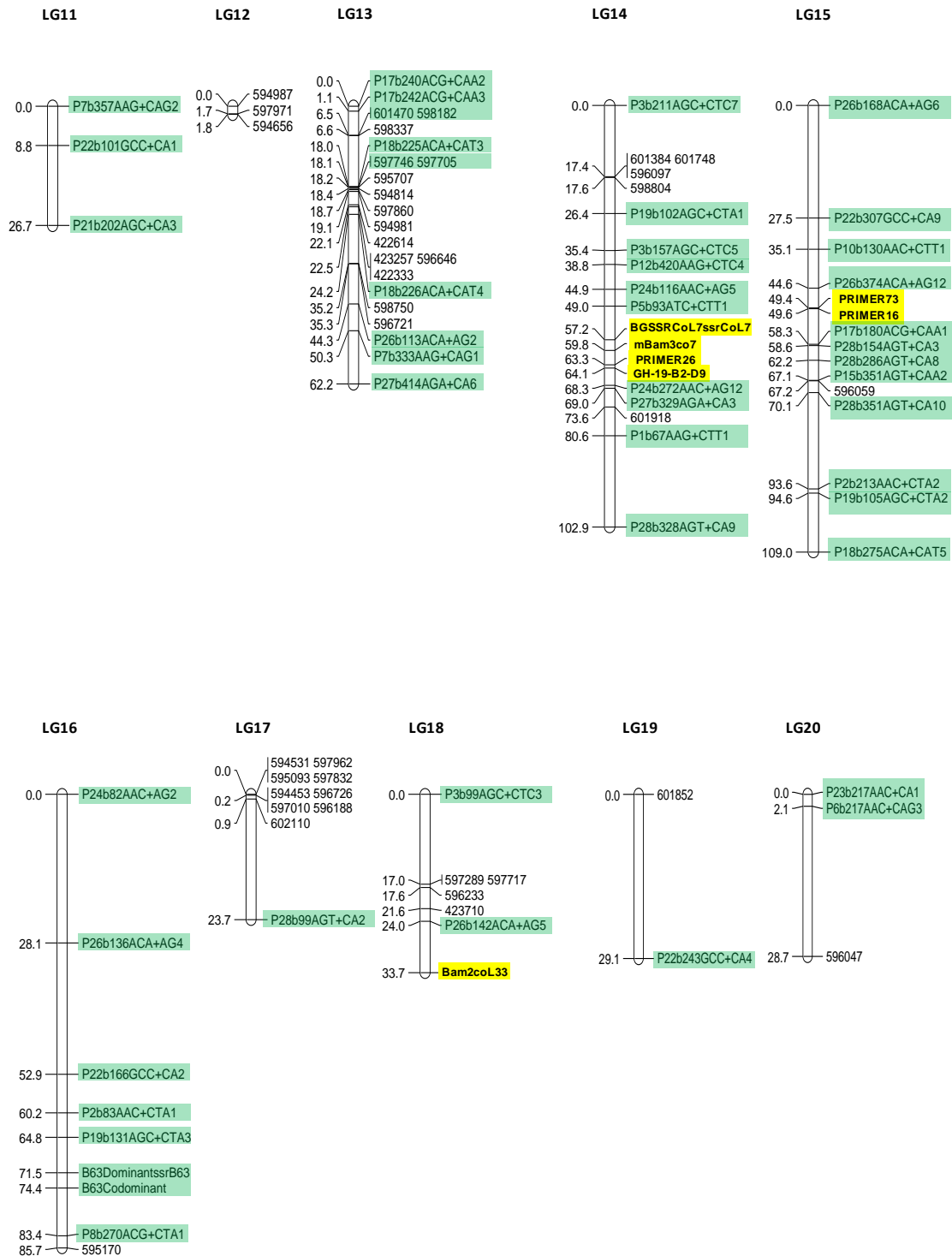
**Figure 5-3: genetic linkage map of 20 linkage groups. This was constructed in 98 F$_2$ individuals derived from the cross between individuals of the DipC and VSSP11 landraces. The map represents the locations of 12 SSR, 110 DArT and 82 AFLP markers. Positions are given in centimorgan (Haldane units) to the right of the linkage groups and the name of markers to the left. Microsatellite and AFLP markers were highlighted on the linkage groups with yellow and green color, respectively.**

The linkage map of the $F_2$ population derived from the wide cross spans a total of 901 cM, giving genome coverage of 79.6%, compared with the 67% of marker-marker linkage compared to the narrow cross map (Table 5-4).

**Table 5-4: Distribution of the markers, linkage group size and marker density in the genetic map constructed in an $F_2$ population of DipC x VSSP11**

| Linkage groups | Length (cM) | No. of markers mapped in the groups | | | | Average marker interval (cM) |
|---|---|---|---|---|---|---|
| | | Total marker | SSR | DArT | AFLP | |
| 1 | 43.0 | 41 | - | 41 | - | 1.05 |
| 2 | 26.1 | 3 | - | - | 3 | 8.69 |
| 3 | 1.7 | 2 | - | 2 | - | 0.83 |
| 4 | 53.1 | 14 | - | 11 | 3 | 3.79 |
| 5 | 60.3 | 11 | - | 7 | 4 | 5.48 |
| 6 | 42.4 | 9 | - | 3 | 6 | 4.71 |
| 7 | 93.9 | 8 | - | - | 8 | 11.74 |
| 8 | 36.3 | 6 | 3 | 2 | 1 | 6.05 |
| 9 | 6.1 | 2 | 1 | 1 | 0 | 3.05 |
| 10 | 35.0 | 5 | 1 | - | 4 | 7.00 |
| 11 | 26.7 | 3 | - | - | 3 | 8.91 |
| 12 | 1.8 | 3 | - | 3 | - | 0.58 |
| 13 | 62.2 | 22 | - | 15 | 7 | 2.83 |
| 14 | 102.9 | 19 | 4 | 4 | 11 | 5.41 |
| 15 | 109.0 | 15 | 2 | 1 | 12 | 7.27 |
| 16 | 85.7 | 9 | - | 1 | 8 | 9.52 |
| 17 | 23.7 | 10 | - | 9 | 1 | 2.37 |
| 18 | 33.7 | 7 | 1 | 4 | 2 | 4.81 |
| 19 | 29.1 | 2 | - | 1 | 1 | 14.55 |
| 20 | 28.7 | 3 | - | 1 | 2 | 9.57 |
| Total | 901.2 | 194 | 12 | 106 | 76 | - |
| Range | 1.7-105.7 | 2-45 | 1-4 | 1-45 | 1-12 | 0.83-14.55 |

### 5.3.7 Combining 'Narrow' and 'Wide' cross maps

JoinMap4 was used to combine linkage groups from both maps. ''Combine Maps'' command was used to align genetic maps obtained in different populations for a visual inspection of the marker order. The 'narrow' and 'wide' crosses carried a total of 8 common SSR and 24 DArT markers (Table 5-5). These were used to combine linkage groups from both crosses, where two or more common markers existed. The JoinMap4 function 'combine groups for map integration' was applied for the pairwise data, followed by regression mapping under default conditions. Two sets of linkage groups could only be linked through the existence of one common marker each and they were also shown in Figure 5-4. While it was possible to confirm common groups through one common marker on the same chromosome, their relative orientations were not determined.

**Table 5-5: Number and the type of common markers in both 'narrow' and 'wide' cross population used to combine linkage groups**

| Linkage group | Linkage groups | | Common marker | |
|---|---|---|---|---|
| | 'narrow' cross map | 'wide' cross map | number of markers | type of markers |
| combined group-1 | 1 | 14 | 3 | SSR |
| combined group-2 | 10 | 8 | 3 | SSR |
| combined group-3 | 3 | 13 | 10 | DArT |
| combined group-4 | 5 | 5 | 2 | DArT |
| combined group-5 | 14 | 17 | 8 | DArT |
| combined group-6 | 13 | 14 | 4 | DArT |
| non grouped a | 17 | 15 | 1 | SSR |
| non grouped b | 10 | 10 | 1 | SSR |
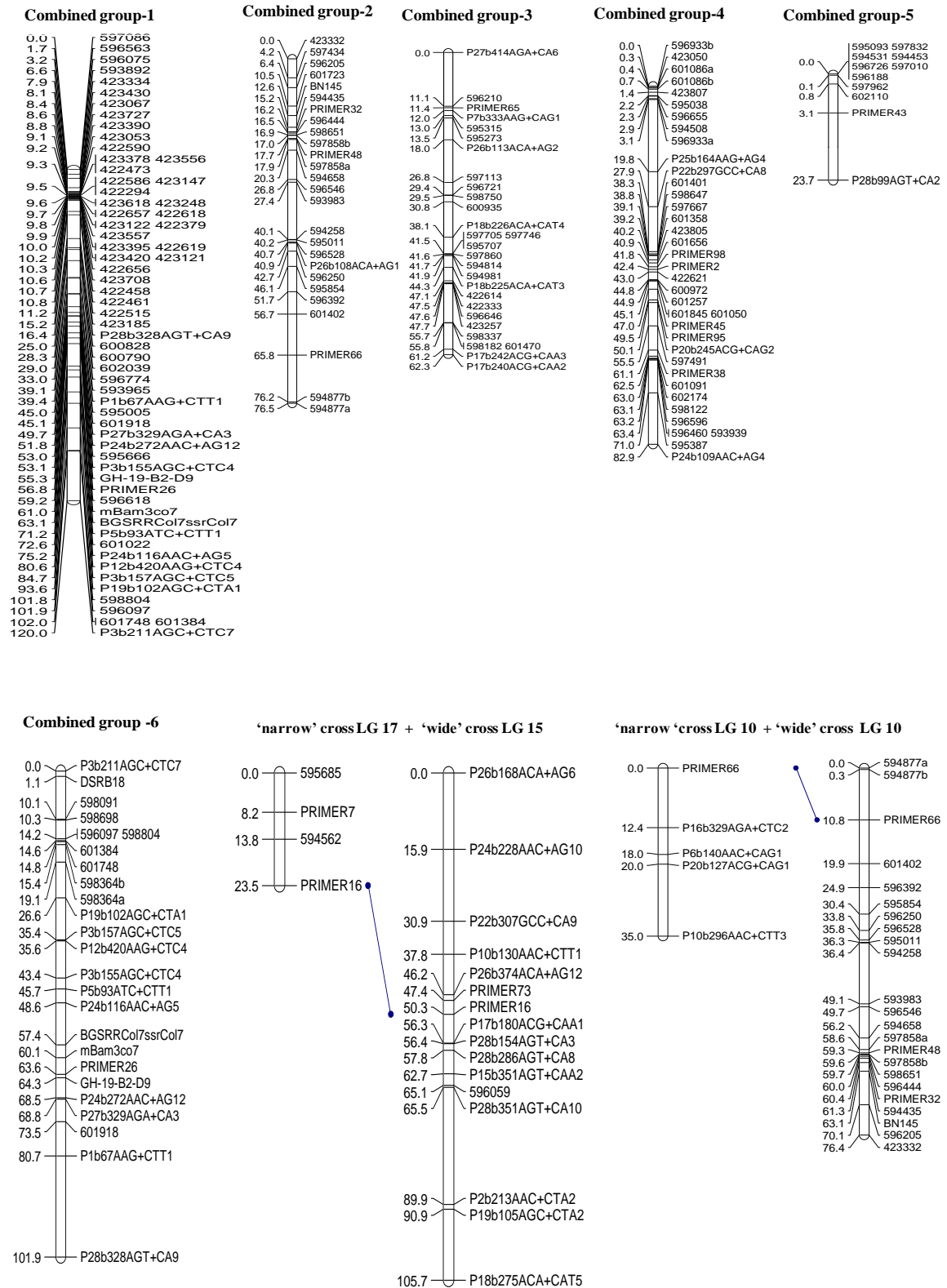
**Figure 5-4: Combined linkage groups from the 'narrow' and 'wide' maps where two or more common markers exist; and the 'combined' linkage groups where a single common marker exists and relative orientation cannot be determined.**

**5.4 Discussion**

**5.4.1 Mapping populations**

The same basic approaches were used to construct the first genetic linkage map of bambara groundnut, based on a population derived from a cross of individuals of a wild accession and of a cultivated landrace (Basu, 2005; Mayes *et al.*, 2009) and an $F_3$ population derived from the cross between two individuals of domesticated landraces (DipC and Tiga necaru). Mapping in this cross aimed to further our understanding of the genetic basis of agronomically important traits in bambara groundnut. The genetic analysis of the previous cross had identified a number of genes important for the domestication process for bambara groundnut (Basu *et al.*, 2007c). Construction of a comprehensive genetic linkage map relies upon polymorphism between the progenitors of the mapping population to study segregation patterns in the offspring and the existence of sufficient molecular markers. Although, a relatively low level of genetic polymorphism in the specific gene pool of bambara groundnut had been reported previously (Basu *et al.*, 2007a), a polymorphism rate of 36.3% and 33.1% for SSRs was identified in narrow and wide cross populations of bambara groundnut, respectively. This level of polymorphism is higher than reports in a number of other legume crops. A polymorphism level of 16.1% was identified for the parental alleles using 945 SSR markers in 186 $F_2$ plants derived from a cross between an annual cultivated mungbean line 'KUML29-1-3' (*Vigna radiata* var. *radiata*) and an Australian wild perennial mungbean accession 'W021' (*Vigna radiata* var. *sublobata*) (Kajonphol *et al.*, 2012). The polymorphism level remains reasonable even for transferred SSR markers from non-target legume species, for instance 109 transferable SSR markers from soybean were amplified successfully in peanut and

they detected 28% polymorphism among 4 cultivated peanut (*Arachis hypogaea* L)

lines (He *et al.*, 2006).

## 5.4.2 Population size

Population sizes used in constructing genetic maps generally ranges from 50 to 250

individuals, however larger populations are required to determine the marker order for

high-resolution mapping (Mohan *et al.*, 1997). The population size in most cases is

technically limited mainly by the amount of seed available from a single F1 cross.

This is clearly a species dependent issue. Population size is one of the factors that

affect the power of a QTL experiment to be able to identify the loci that underlie

phenotypic traits (Erickson *et al.*, 2004). It has been demonstrated that the confidence

level generated for a QTL based on 100 individuals in genetic map drops from 90% to

60% when the population size is decreased to 50 individuals, affecting the power of

detecting the linkage as well as the estimate and accuracy of the recombination

frequency (Liu, 1998).

In order to construct the genetic map in the present study 73 $F_3$ progeny were derived

from a controlled cross between two single genotype accessions both with a

domesticated background. A population size of 75 recombinant inbred lines was used

to construct a linkage map in common bean (Freyre *et al.*, 1998). Construction of the

wide cross genetic linkage map combining SSR, DArT and AFLP markers was based

on 98 $F_2$ progeny derived from a controlled cross between the domesticated landrace

(DipC) and the wild accession (VSSP11).

**5.4.3 Phase issues with the dominant markers**

Lack of available parental samples (DipC and Tiga necaru) for the segregating $F_3$ population at the time of the generation of the DArT array data resulted in a lack of phase information for these dominant markers. By the time the SSR data was generated, the parental genotypes had been recovered and confirmed by genetic fingerprinting. The linkage phases were determined prior to the final mapping. The genotype data was run in JoinMap4 as a cross pollinator (unknown linkage phase), after which the right phase could be determined for the initially grouped loci. Determining the correct linkage phase led to 30 markers which had weak association not being linked into the groups. Only markers showing linkage to other markers were used to develop the RIL3 .loc file. This decreased the number of markers to 239 for linkage analysis using the RIL3 model.

Linkage phase problems associated with some DArT and AFLP markers for the segregating $F_2$ population initially prevented map calculations in certain groups (Basu, 2005). Reversing the genotype codes for the population from (a, c) to (b, d) or *vice versa*, has solved the problem of suspect linkage for those markers and parental genotype data is also available for most markers, but not AFLPs.

**5.4.4 Segregation distortion**

Marker segregation distortion is commonly encountered in actual genetic mapping populations, leading to a skew in the frequency of genotypes from the expected Mendelian ratio (Lu *et al.*, 2002) due to differential inheritance of alleles. The genotype frequencies for each locus were calculated by the JoinMap4 software to study segregation distortion. The segregation pattern was tested against normal

Mendelian expectation ratios through normal classification of genotypes using chi-square at p<0.05 based on a RIL3 model for the 'narrow' map and $F_2$ model for the 'wide' map.

Markers (SSR, DArT and AFLP) revealed high proportions of distorted loci (32% and 27%) from expected Mendelian ratios (*P*<0.05) in both narrow and wide cross maps, respectively. The distorted markers were found to be distributed across the linkage groups of both maps, rather than clustered on specific groups.

Skewed segregation ratios of the markers were reported to be highly variable in other studies. An investigation has reported 40.6% marker segregation distortion in an $F_2$ intra-specific population of *Medicago tornata* (Jenczewski *et al.*, 1997). One of the highest frequencies of marker distortion of 73% was reported in an inter-specific recombinant inbred of tomato (Paran *et al.*, 1995). In mapping and QTL analysis soybean (*Glycine max*) distortion of 19.93% of SSR markers on 14 linkage groups out of 20 was reported. The markers were distorted from the ratio of 1:1 in 106 RILs ($F_9$) derived from the cross of two soybean genotypes (BD2 and BX10) (Liang *et al.*, 2010). Marker deviation from the expected ratio was thought to be a result of small population sizes, genotyping score errors, or the consequence of missing data (Millan *et al.*, 2010). Other factors apart from the mapping population could affect the segregation of markers. (Zhang *et al.* (2006) reported that segregation distortion could be partially caused by gametophytic and sterility factors. Additionally, abortion of the parental gametes or zygotes, non-homologous recombination, transposable element and environmental agents would be counted among the factors involved (Knox and Ellis, 2002; Yamagishi *et al.*, 2010). Among the 65 polymorphic SSR loci used in constructing a genetic map in barley , 22 loci (33.8%) showed genetic distortion (*P*<0.05), using 260 $F_2$ individuals derived from a male sterile line and elite cultivar.

It has been identified that all of the distorted markers were deviated toward male sterile parent (Liu *et al.*, 2011).

## 5.4.5 Map construction and marker distribution

The genetic mapping of SSR and DArT markers was conducted in a population derived from the narrow cross. Twenty one linkage groups were identified, with a total 608.6 cM length to cover approximately 54% of bambara groundnut genome (based on 11 Morgans for 11 chromosomes as a crude assumption), although the high marker-marker linkage (238 out of 269 tested in the CP population) at 89% suggests more comprehensive coverage. Parental dissimilarity could be suppressing recombination, or potentially the developed markers could be clustered to particular regions of the bambara groundnut genome. The wide cross genetic linkage map constructed using an interspecific $F_2$ progeny had a total length of 901 cM with greater genome coverage (79.6%) compared to the narrow cross linkage map, however it had far lower marker-marker linkage of 67% (194 markers out of 301) resulted from 12 SSRs, 106 DArTs and 76 AFLPs. A possible explanation would be that the marker data for wide cross is less accurate than the data for the narrow cross, especially for AFLP markers (represented 141 out of 194 markers) which are largely responsible for the wider genome coverage of the map derived from the cross between DipC and VSSP11. The combination of three marker types in the wide cross map might also have a positive effect on coverage. It has been reported that different marker systems differ in the mechanism of detecting polymorphism, their genome coverage or location and the use of multiple marker types could be complemently and increase coverage overall (Adawy *et al.*, 2005).

In the narrow cross map both kind of markers were evenly distributed across the linkage map, although the largest number of 43 and 41 DArT markers were located in linkage group 1 for both 'wide' and 'narrow' maps. The observed cluster of DArT markers is consistent with the earlier investigation of two major distinct DArT genotype groups (Stadler, 2009), including DipC in the smaller group containing the unusual repeat cluster As DipC was the maternal parent of both crosses here, this clustering effect is perhaps not unexpected. The persistency of DArT clustering seems to be reflected also in the wide cross map (LG1). Clustering of DArT markers in a particular genome region has been reported in chick pea, barley and wheat (Akbari *et al.*, 2006; Semagn *et al.*, 2006c; Thudi *et al.*, 2011).

In the narrow cross, the markers in LG11 were only combined into a single map after three rounds of mapping. Because this included markers with negative distances, Map1, from the first round was used (and no further markers were added to Map2). While the second round (Map2) was used for the linkage groups 1, 5 and 6 as they allowed to map the optimum number of loci into the genetic map with no exceed of default jump (threshold value < 5) or negative map distances.

Three rounds were used to build the maps of 7 linkage groups for the wide cross map. A large jump in the goodness-of-fit of the second to third round indicated the poor fit of these additional loci. It was decided to remove 10 markers and adopt the map from round 1 (Map1) for the groups 1, 6, 7, 13, 14, 15, and 16.

Overall, the loss of 10 markers due to exceeding the default *jump* (5) or being associated with negative distances suggests that the map generated is relatively free from noisy data.

To our knowledge this is the first linkage map based on a combination of SSR and DArT markers in bambara groundnut. This map along with the improved wide cross map using a combination of SSR, DArT and AFLP markers, represents an important step toward genetic analysis of interesting agronomical traits in bambara groundnut.

The number of linkage groups in both narrow and wide cross map exceeded the expected number of 11 linkage groups for a comprehensive map in bambara groundnut ($2n=2x=22$). Since some of these groups had only a few markers, it can be concluded that the apparent excess of linkage groups might be due to incomplete coverage of the genome with the marker loci. We predict that the smaller groups will be brought together by adding more markers into the map and data has been generated from a sequence-based DArT approach which should allow this to be done. Unfortunately, time constraints prevented the use of this data in the current analysis

The SSR and DArT makers developed will be useful for comparative genomic analysis between the mapping populations of bambara groundnut and between this crop and other relative genomes. They are a pre-requisite for the application of more efficient breeding technologies such as marker-assistant selection and will give a better understanding of agronomical complex traits and phylogenetic analysis in bambara groundnut.

## 8.4.6 Attempt to combine the linkage maps

The 'narrow' and 'wide' linkage groups in bambara groundnut were combined where possible through a number of SSR and DArT markers. Combining linkage groups with at least two common markers was possible for 6 groups, however due to the lack

of a physical map for bambara groundnut the orientation of two other groups with single linkages was not.

The combined map will assist breeders to accurately select tightly-linked markers for agronomically important genomic regions for marker-assisted selection, will allow the positions of QTL for similar traits derived from the two crosses to be compared (where linkage exists) and will facilitate comparative mapping with other related legumes in the future (Millan *et al.*, 2010).

# Chapter 6:  QUANTITATIVE TRAIT LOCI (QTL) MAPPING OF AGRONOMIC TRAITS

## 6.1 Introduction

QTL analysis is the study of genetic variation resulting from many genes with small effects to attempt to locate the genes responsible in complex traits and explore their effects and interactions (Kearsey, 1998). In order to identify molecular makers associated with QTLs, members of a random segregating population have to be scored first for a quantitative trait. The molecular genotype of each member of the population is then determined and a thorough search will be made for associations that might exist between the markers and the quantitative trait.

Since there has only been one attempt to date to identify molecular markers responsible for agronomic traits in bambara groundnut (Basu, 2005) a  review of the QTL mapping for some agronomically important traits in related legume crops might enhance the understanding of these traits in bambara groundnut.

### QTL mapping of agronomic traits in legume crops

In soybean (*Glycine max* L. Merr.) QTL mapping was conducted for agronomic traits across two years in a study by Wenxin *et al.* (2008) . A total of 136 SSR markers segregated in the population of 126 RIL ($F_5$) derived from a cross between two lines of late maturity (PI 171451) and early maturity (Hwaeomputkong). The markers were distributed over 20 linkage groups (LGs), covering 1073.9 cM of the soybean genome. A total of 15 QTLs were detected with LOD scores >3 for six major agronomic traits. QTLs of two traits (*days to flowering* and *days to maturity*) were

located at the same position in two linkage groups and reflected significant correlations among corresponding traits based on field data.

A SNP-based genetic linkage map was generated by Bobby *et al.* (2012) in soybean (*Glycine max* L. Merr.), using a RIL population derived from the cross between cultivars PI 438489B (resistant to sudden death syndrome) and Hamilton (high yield); *n*=50. A total of 18 QTLs were detected for *days to germination, days to flowering, plant height, pod number, seed number, 100-seed weight,* and *total seed weight* on 10 different chromosomes. The clusters of QTL for these traits were identified containing QTL controlling other important traits such as sudden death syndrome resistance and soybean cyst nematode resistance. The author emphasised the importance of these QTLs and identified that they may be introduced into breeding programs to develop soybean cultivars with high yield potential which have good resistance to these diseases.

In the model legume *Lotus japonicus* the first QTL analysis was performed by Gondo *et al.* (2007) for agronomic traits. A RIL population derived from the cross between the accessions Miyakojima MG-20 and Gifu B-129 was used in 2 year trial. The traits of yield and its components were studied. Some QTLs detected were co-located, especially those for *pod length, pod width, seeds/pod*, and *seed mass. Seed mass* QTLs were located at 5 locations that mapped to the corresponding genomic positions of equivalent QTLs in soybean, pea, chickpea, and mungbean. It was concluded that this study could provide key information on the traits studied for marker-assisted breeding of the important legume crops (Gondo *et al.*, 2007).

Kajonphol *et al.*, (2012) conducted a QTL analysis for mungbean (*Vigna radiata L.*) ($2n=2x=22$) to identify chromosome regions controlling agronomic traits, using an $F_2$ mapping population derived from an inter-specific cross between the wild and

cultivated mungbean. Twenty QTLs were identified controlling major agronomic characters for *days to first flower*, *days to first pod maturity*, *days to harvest*, *100 seed weight*, number of *seed per pod* and *pod length*. Most of these QTLs were located on only two linkage groups. Extensive genome conservation between the mungbean map and maps for azuki bean (*Vigna angularis*) and blackgram (*Vigna mungo*) reveal the potential for cross-species genetic markers to detect marker-trait association among *Vigna* species.

In garden pea (*Pisum sativum* L.) Irzykowska and Wolko (2004) conducted interval mapping of QTLs for the agronomic traits of *seed number, pod number, 1000-seed weight, seed yield,* and *seed protein content*. Traits were measured in $F_2$ (114 plants) and in $F_4$ (104 RILs) plant populations derived from a cross of a large-seeded line (Wt10245) with a small-seeded line (Wt11238). The map consisted of a combination of 204 markers of different types (140 AFLPs, 24 RAPDs, 10 ISSRs, 5 CAPSs, 1 STS, 11 isozymes and 13 morphological markers). Thirty seven QTLs were detected across both populations on the seven pea chromosomes. A number of these QTLs for different traits were localized to the same map interval, possibly indicating pleiotropy or the actions of a common gene effecting an underlying trait. The authors postulate that these QTLs have an important role in controlling the yield and seed protein content in garden pea.

The aim of the current study was to carry out QTL analysis of bambara groundnut to identify regions of the genome that are contributing to variation in traits of agronomic importance. Detection, localization and estimating the effects of potential QTL could clarify the genetic mechanisms underlying agronomic traits to facilitate MAS in future breeding programs of bambara groundnut (*Vigna subterranea* (L.) Verdc.; $2n=2x=22$). QTL analysis was performed for agronomic traits that were either quantitatively distributed or did not follow Mendelian segregation in the $F_2$ and $F_3$ progenies of the crosses. Where links between the 'wide' and 'narrow' maps could be identified, the potential presence of the same QTL in both maps was examined.

## 6.2 Result

QTL analysis was conducted for the two constructed linkage maps of 'narrow' and 'wide' crosses in bambara groundnut. The first analysis of the 'narrow' cross was carried out for agronomically important traits in an $F_3$ population derived from the cross between individual genotypes of the DipC and Tiga necaru landraces. The trait data came from the controlled environment tropical glasshouses based at Sutton Bonington ($F_3$; FutureCrop) and field experiments ($F_3$; Indonesia). Trait data for the segregating $F_2$ population from the same cross was also used to support the $F_3$ QTL analysis and was derived from single plants in a controlled environment glasshouse.

In the second analysis ('wide' cross), the QTL analysis was conducted for an updated map derived from the controlled cross between individuals of the DipC and VSSP11 landraces. Previous data from the segregating $F_2$ population of this cross was used in marker-trait analysis and was derived from controlled environment glasshouse measurements on single plants.

### 6.2.1 QTL analysis in 'narrow' cross map

### 6.2.1.1 Trait distribution

As a prerequisite for QTL analysis the distribution patterns of data were tested for the entire range of agronomic traits. Data came from the segregating $F_3$ population evaluated in FutureCrop glasshouse and Indonesian field and the $F_2$ population of the same cross evaluated in Tropical Crops Research Unit. The traits distribution was tested for different environments and generations using Anderson darling normality

tests. The description of the data distribution and the correlation between trait values was presented in Chapter 4. Non-normally distributed data were subjected to Box-Cox transformations prior to the QTL analysis. Transformation was used to improve the normality of traits for both $F_2$ and $F_3$ populations as shown in Table 6-1.

The initial tests suggested a non-normal distribution in these traits. Data from only two traits - *growth habit* and *eye pattern around hilum* - followed known inheritance patterns of Mendelian incomplete and complete dominance in the glasshouse, respectively. Traits that showed continuous distribution (in Chapter 4) and those transformed (Table 6-1) which tested as normally distributed after transformation were subjected to quantitative trait loci analysis using interval mapping. Non normal distributions were observed for *days to emergence* in both trials of $F_3$ progenies (glasshouse and field) and in the $F_2$ population. *Growth habit* (in glasshouse and field) and *eye pattern around hilum* (recorded in glasshouse only) were also tested to be non-normally distributed, despite attempts to transform the data to achieve normality. Detecting markers linked to QTLs for non-normally distributed data was performed only with Kruskal-Wallis analysis.

### 6.2.1.2 Marker and trait associations

Maximum LOD scores were determined by interval mapping of traits, with the most likely location for the QTL given at maximum LOD. LOD peaks $\geq 2$ are presented as a potential/indicative QTL in the current investigation. Significant LOD thresholds for QTL determination were estimated for each trait after 10000 permutation tests. These ranged from LOD = 2.2 to LOD = 3.1 for different traits studied in the segregating $F_2$ and $F_3$ populations. A LOD score $\geq$ the significant threshold value was used to declare a QTL as significant.

**Table 6-1: Statistical analysis of traits showing non-normal distributions, before and after transformation, for $F_2$ and $F_3$ progenies data derived from the narrow cross between DipC and Tiga necaru evaluated under different environments.**

| Trait | Data type | Min. value | Max. value | Median | Mean value | Variance | Skewness | Kurtosis | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $F_3$ population evaluated in glasshouse | | | | | |
| *Leaf no./plant* | Non normal | 29.50 | 148.25 | 60.75 | 62.49 | 325.58 | 2.07 | 7.89 | 0.00 |
| | Transformed | 5.43 | 12.18 | 7.79 | 7.83 | 1.12 | 1.22 | 4.16 | 0.01 |
| *Leaf area* | Non normal | 1402.34 | 9562.37 | 3200.75 | 3461.23 | 1925713.87 | 1.80 | 5.55 | 0.00 |
| | Transformed | 37.45 | 97.79 | 56.58 | 57.83 | 119.29 | 0.94 | 2.21 | 0.05 |
| *Double seeded pods/plant* | Non normal | 0.00 | 19.00 | 4.00 | 5.42 | 16.94 | 1.26 | 1.57 | 0.00 |
| | Transformed | 3.41 | 3.90 | 3.65 | 3.65 | 0.01 | 0.05 | -0.50 | 0.64 |
| *Seed weight* | Non normal | 7.79 | 58.42 | 27.67 | 28.93 | 144.90 | 0.72 | -0.11 | 0.00 |
| | Transformed | 2.79 | 7.64 | 5.26 | 5.27 | 1.22 | 0.29 | -0.42 | 0.19 |
| *Biomass dry weight* | Non normal | 27.82 | 126.53 | 54.86 | 61.34 | 524.26 | 0.81 | -0.02 | 0.00 |
| | Transformed | 5.27 | 11.25 | 7.41 | 7.70 | 2.01 | 0.49 | -0.52 | 0.06 |
| | | | | $F_3$ population evaluated in the field | | | | | |
| *Pod no./plant* | Non normal | 1.00 | 23.40 | 6.83 | 7.70 | 21.48 | 0.89 | 0.81 | 0.01 |
| | Transformed | 1.00 | 4.84 | 2.61 | 2.65 | 0.70 | 0.22 | -0.51 | 0.25 |
| *Pod length/ plant* | Non normal | 8.00 | 20.00 | 13.51 | 13.61 | 3.14 | 0.66 | 3.75 | 0.01 |
| | Transformed | 2.83 | 4.47 | 3.68 | 3.68 | 0.06 | 0.16 | 3.66 | 0.02 |
| | | | | $F_2$ population evaluated in TRRU | | | | | |
| *Double seeded pods/plant* | Non normal | 0.00 | 34.00 | 6.00 | 8.32 | 71.02 | 1.16 | 0.36 | 0.00 |
| | Transformed | 0.00 | 5.83 | 2.24 | 2.38 | 2.45 | 0.22 | -0.72 | 0.01 |
| *Seed no./plant* | Non normal | 19.00 | 355.00 | 104.00 | 115.37 | 6049.15 | 1.03 | 0.99 | 0.00 |
| | Transformed | 4.36 | 18.84 | 10.20 | 10.13 | 12.83 | 0.30 | -0.52 | 0.24 |
| *Biomass dry weight* | Non normal | 11.00 | 306.30 | 99.99 | 99.17 | 4482.23 | 0.64 | -0.05 | 0.00 |
| | Transformed | 3.32 | 17.50 | 10.00 | 9.33 | 12.34 | 0.02 | -0.96 | 0.02 |

The QTL results for individual traits here are presented separately and the trait data evaluated in all three experiments of FutureCrop glasshouse, Indonesian field and in TCRU (when presented) are introduced together for each trait.

***Days to emergence*:** Kruskal-Wallis analysis revealed three potential QTL for *days to emergence* associated with DArT and SSR markers (bgPabg-601086, PRIMER66 and mBam3co33) on different linkage groups of 5, 10 and 11, respectively, valued K* >6 at p = 0.05 in the glasshouse (Table 6-2). Based on the field dataset, the marker bgPabg-423556 was identified on LG1 to be in association with this trait (K* = 7.3 at p = 0.01), in addition to two other associations with markers bgPt-598091 and bgPabg-601086 on LGs 13 and 5, respectively (Table 6-3). Data recorded for the $F_2$ progeny in the TCRU indicated a strong association of *days to emergence* with SSR marker PRIMER16 (K* = 12.3 at p = 0.005) on LG17 (Table 6-4).

***Flower no./plant*:** A putative QTL was found for glasshouse data on LG8 at the 2.0 cM position mapped by interval mapping at a LOD score of 2.4, explaining 16% of the phenotypic variation. Marker bam2coL63 at 0.0 cM position (K* = 9.9 at p = 0.01) was the nearest to this locus.

***Terminal leaflet length* (cm):** A significant QTL was detected on LG8 at the 2.0 cM position mapped by interval mapping at a LOD score of 3.2 (Figure 6-1). This genomic region explained 20.3% of the phenotypic variation and it is close to the marker bam2coL63 at 0.0 cM position (K* = 11.2 at p = 0.005). This marker also recorded a LOD score of 2, showing a just-putative QTL for this trait in the $F_2$ progenies grown in TCRU. Marker bam2coL63 was also found to be linked to *flower no./plant*.

**Table 6-2: QTL mapping in bambara groundnut using interval mapping and Kruskal-Wallis analysis for the F₃ generation data of the 'narrow' cross in the FutureCrop glasshouses.**

| Traits | Linkage group | Position (cM) | Locus | Interval mapping | | | | Kruskal-Wallis analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | LOD | PT[a] | PVE[b] | Additive | K* | Df | Signifi. level |
| *Days to emergence* (DE) | 11 | 15.0 | mBam3co33 | | | | | 6.9 | 2 | ** |
| | 5 | 0.9 | bgPabg-601086 | | | | | 6.6 | 1 | ** |
| | 10 | 10.8 | PRIMER66 | | | | | 6.5 | 2 | ** |
| | 12 | 12.9 | bgPabg-594494 | | | | | 6.3 | 1 | ** |
| *Flower no./plant* (FN) | 8 | 2.0 | | 2.4 | 2.9 | 15.8 | -16.7 | | | |
| | 8 | 0.0 | Bam2coL63 | 2.3 | | 15.3 | -15.6 | 9.9 | 2 | *** |
| *Terminal leaflet length* (TLL) | 8 | 2.0 | | 3.2 | 2.6 | 20.3 | -0.4 | | | |
| | 8 | 0.0 | Bam2coL63 | 3.1 | | 19.7 | -0.4 | 11.2 | 2 | **** |
| *Terminal leaflet width* (TLW) | 3 | 19.7 | bgPt-600935 | 3.2 | 2.6 | 20.4 | 0.2 | 13.1 | 1 | ****** |
| *Leaf area* (LA) | 3 | 16.6 | | 2.3 | 2.8 | 15.5 | 1.8 | | | |
| | 3 | 15.6 | bgPabg-597113 | 2.3 | | 15.3 | 1.4 | 6.5 | 1 | ** |
| *Plant spread* (PS) | 4 | 0.0 | BN6b | 3.9 | 2.7 | 24.6 | 3.7 | 16.5 | 2 | ****** |
| *Stem no./plant* (STN) | 4 | 14.2 | | 2.4 | 2.6 | 16.1 | -0.9 | | | |
| | 4 | 11.2 | bgPt-600898 | 2.3 | | 15.4 | -0.8 | 4.8 | 1 | ** |
| *Node no./stem* (NN) | 1 | 33.0 | bgPabg-596774 | 3.3 | 2.6 | 21.1 | -1.1 | 8.8 | 1 | **** |
| | 4 | 11.2 | bgPt-600898 | 2.7 | | 17.9 | 1.1 | 9.7 | 1 | **** |
| *Internode length* (IL) | 4 | 3.0 | bgPabg-596988 | 7.9 | 2.6 | 43.5 | 0.7 | 18.9 | 1 | ******* |
| *Growth habit* (GH) | 4 | 0.0 | BN6b | | | | | 23.4 | 2 | ******* |
| | 10 | 70.1 | bgPabg-596205 | | | | | 8.3 | 1 | **** |
| | 18 | 5.1 | PRIMER10 | | | | | 7.8 | 2 | ** |
| *Pod no./plant* (PN) | 1 | 33.0 | bgPabg-596774 | 2.3 | 2.7 | 15.4 | -7.0 | 8.8 | 1 | **** |
| *Double seeded pods/plant* (DPN) | 4 | 1.0 | | 3.4 | 2.9 | 22.0 | 0.5 | | | |
| | 4 | 0.0 | BN6b | 3.3 | | 21.7 | 0.5 | | | |
| *Peduncle length* (PEL) | 4 | 1.0 | | 9.7 | 2.7 | 50.3 | 0.9 | | | |
| | 4 | 2.4 | bgPt-423527 | 9.6 | | 49.9 | 0.9 | 8.1 | 1 | **** |
| *Pod weight* (PWE) | 1 | 33.0 | bgPabg-596774 | 2.6 | 2.5 | 17.0 | -6.8 | 6.4 | 1 | ** |
| *Pod length* (PLE) | 12 | 15.1 | | 4.6 | 2.7 | 28.4 | 0.8 | | | |
| | 12 | 12.9 | bgPt-598767 | 4.6 | | 28.0 | 0.8 | 9.6 | 1 | **** |
| *Pod width* (PWD) | 12 | 20.1 | | 5.7 | 2.4 | 33.4 | 0.6 | | | |
| | 12 | 22.5 | bgPabg-595682 | 5.5 | | 32.7 | 0.5 | 17.7 | 1 | ******* |
| *Pod length of double seeded* (DPL) | 1 | 0.0 | bgPabg-597086 | 3.8 | 2.7 | 24.5 | -1.5 | 14.7 | 1 | ****** |
| | 12 | 10.5 | | 3.3 | | 21.7 | 1.6 | | | |
| *Pod width of double seeded* (DPW) | 12 | 17.1 | | 4.0 | 2.8 | 25.7 | 0.5 | | | |
| | 12 | 12.9 | bgPt-598767 | 3.7 | | 24.0 | 0.5 | 9.1 | 1 | **** |
| *Seed length* (SEL) | 10 | 49.1 | bgPabg-593983 | 2.5 | 2.7 | 16.4 | 0.4 | 9.6 | 1 | **** |
| *Seed width* (SEW) | 12 | 15.1 | | 2.0 | 2.7 | 13.6 | 0.3 | | | |
| | 12 | 12.9 | bgPt-598767 | 2.0 | | 13.5 | 0.3 | 4.6 | 1 | ** |
| *Seed no./plant* (SEN) | 1 | 34.0 | | 2.3 | 2.5 | 15.1 | -7.9 | | | |
| | 1 | 33.0 | bgPabg-596774 | 2.3 | | 15.0 | -7.6 | 7.6 | 1 | *** |
| *Seed weight* (SWT) | 1 | 33.0 | bgPabg-596774 | 2.7 | 2.6 | 17.8 | -0.5 | 6.0 | 1 | ** |
| *Biomass dry weight* (BDW) | 1 | 33.0 | bgPabg-596774 | 3.5 | 3.0 | 22.4 | -11.6 | 10.0 | 1 | **** |
| *Shelling%* (SH%) | 7 | 13.3 | bgPabg-594335 | 3.0 | 2.9 | 19.4 | 3.4 | 13.6 | 1 | ****** |
| *100-seed weight* (HSW) | 7 | 9.4 | | 2.7 | 2.5 | 17.4 | 4.3 | | | |
| | 7 | 10.5 | bgPt-601852 | 2.6 | | 17.3 | 4.3 | 11.6 | 1 | ***** |
| *Eye pattern around hilum* (EP) | 12 | 22.5 | bgPabg-594999 | | | | | 29.68 | 1 | ******* |
| | 18 | 0.0 | bgPabg-594261 | | | | | 9.304 | 1 | **** |

a : permutation-10000 times test

b : percentage of total phenotypic variation explained by the QTL

Significant level of K * values: *: 0.10, **: 0.05, ***: 0.01, ****: 0.005, *****: 0.001, ******: 0.0005, *******: 0.0001

**Table 6-3: QTL mapping in bambara groundnut using interval mapping and Kruskal-Wallis analysis for F₃ generation data of 'narrow' cross in the field.**

| Traits | Linkage group | Position (cM) | Locus | Interval mapping | | | | Kruskal-Wallis analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | LOD | PT[a] | PVE[b] | Additive | K* | Df | Signifi. level |
| *Days to emergence* (DE) | 1 | 9.3 | bgPabg-423556 | | | | | 7.3 | 1 | *** |
| | 13 | 9.0 | bgPt-598091 | | | | | 6.9 | 1 | *** |
| | 5 | 0.9 | bgPabg-601086 | | | | | 5.8 | 1 | ** |
| *Node no./stem* (NN) | 3 | 30.2 | bgPabg-595707 | 2.8 | 2.7 | 18.4 | 1.0 | 10.8 | 1 | **** |
| *Internode length* (IL) | 4 | 3.0 | bgPabg-596988 | 7.1 | 2.7 | 40.9 | 0.3 | 20.9 | 1 | ******* |
| *Growth habit* (GH) | 4 | 3.0 | bgPabg-596988 | | | | | 18.1 | 1 | ******* |
| | 4 | 0.0 | BN6b | | | | | 17.6 | 2 | ****** |
| | 18 | 5.1 | PRIMER10 | | | | | 9.7 | 2 | *** |
| | 14 | 0.0 | bgPt-597832 | | | | | 7.4 | 1 | *** |
| *Pod no./plant* (PN) | 18 | 3.0 | | 2.5 | 3.1 | 16.3 | -0.4 | | | |
| | 18 | 5.1 | PRIMER10 | 2.4 | | 15.8 | -0.4 | 9.8 | 2 | *** |
| *pod length* (PLE) | 11 | 3.0 | | 3.2 | 2.5 | 20.9 | 0.1 | | | |
| | 11 | 0.0 | bgPabg-595822 | 3.0 | | 19.9 | 0.1 | 14.8 | 1 | ****** |
| *Biomass dry weight* (BDW) | 1 | 28.9 | bgPt-602039 | 2.9 | 2.9 | 17.6 | -1.8 | 17.9 | 1 | ******* |

a : permutation-10000 times test

b : percentage of total phenotypic variation explained by the QTL

Significant level of K * values: *: 0.10, **: 0.05, ***: 0.01, ****: 0.005, *****: 0.001, ******: 0.0005, *******: 0.0001

**Table 6-4: QTL mapping in bambara groundnut using interval mapping and Kruskal-Wallis analysis for F₂ generation data of narrow cross in TCRU**

| Traits | Linkage group | Position (cM) | Locus | Interval mapping | | | | Kruskal-Wallis analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | LOD | PT[a] | PVE[b] | Additive | K* | Df | Signifi. level |
| *Days to emergence* (DE) | 17 | 23.5 | PRIMER16 | | | | | 12.3 | 2 | **** |
| | 5 | 74.2 | bgPt-595387 | | | | | 7.8 | 1 | *** |
| | 2 | 6.3 | bgPt-422567 | | | | | 6.6 | 1 | ** |
| | 4 | 3.7 | bgPabg-598611 | | | | | 6.4 | 1 | ** |
| *Terminal leaflet length* (TLL) | 8 | 0.0 | Bam2coL63 | 2.0 | 2.6 | 11.6 | -0.6 | 8.6 | 2 | ** |
| *Terminal leaflet width* (TLW) | 5 | 74.2 | bgPt-595387 | 2.6 | 2.5 | 15.3 | 0.3 | 10.4 | 1 | **** |
| *Plant spread* (PS) | 4 | 33.5 | bgPabg-597624 | 3.2 | 2.7 | 18.0 | 5.5 | 14.8 | 1 | ****** |
| *Pod no./plant* (PN) | 1 | 68.0 | | 2.0 | 2.7 | 11.8 | -35.3 | | | |
| | 1 | 72.7 | bgPt-601022 | 1.9 | | 11.0 | -31.5 | 8.0 | 1 | **** |
| *Double seeded pods/plant* (DPN) | 4 | 33.5 | bgPabg-597624 | 3.2 | 2.8 | 19.2 | 0.7 | 13.1 | 1 | ****** |
| *Seed no./plant* (SEN) | 1 | 72.6 | bgPt-601022 | 2.5 | 2.8 | 14.4 | -30.7 | 7.65 | 1 | *** |
| *Biomass dry weight* (BDW) | 1 | 67.0 | | 2.0 | 2.9 | 11.5 | -27.8 | | | |
| | 1 | 59.6 | bgPabg-596618 | 2.0 | | 11.3 | -26.1 | 10.5 | 1 | **** |
| *Shelling%* (SH%) | 12 | 47.5 | bgPt-595486 | 4.8 | 2.6 | 26.3 | -4.0 | 15.3 | 1 | ******* |
| *100-seed weight* (HSW) | 11 | 0 | bgPabg-595822 | 2.1 | 2.5 | 12.2 | 4.6 | 10.6 | 1 | **** |
| | 12 | 47.5 | bgPt-595486 | 2.1 | 2.5 | 12.1 | -4.6 | 9.8 | 1 | **** |

a : permutation-10000 times test

b : percentage of total phenotypic variation explained by the QTL

Significant level of K * values: *: 0.10, **: 0.05, ***: 0.01, ****: 0.005, *****: 0.001, ******: 0.0005, *******: 0.0001

***Terminal leaflet width* (cm):** A single significant QTL was detected with LOD score 3.2 at 19.7 cM on LG 3. The locus bgPt-600935 was linked to this position and it was supported by Kruskal–Wallis output analysis at the same position with a high K* value (K* = 13.1 at p = 0.0005). Based on the data from the TCRU a significant QTL for this trait was detected on LG5 and linked with marker bgPt-595387 at 74.2 cM, having a LOD value of 2.6. Marker-trait linkages for *terminal leaflet width* were also revealed using Kruskal-Wallis analysis.
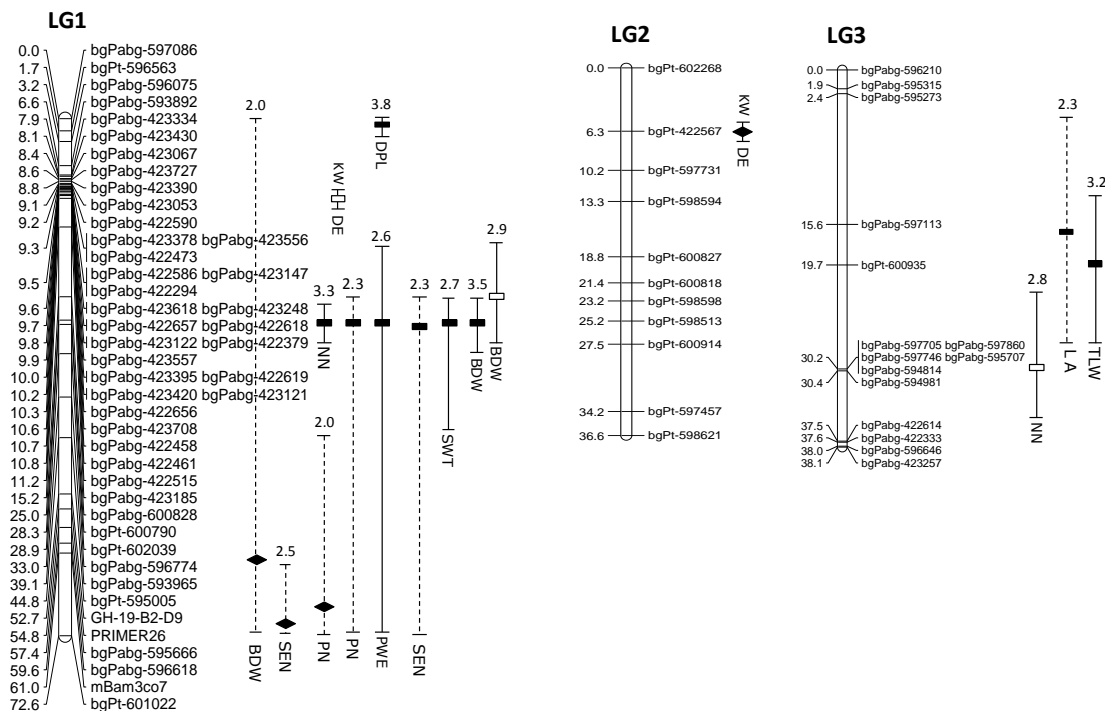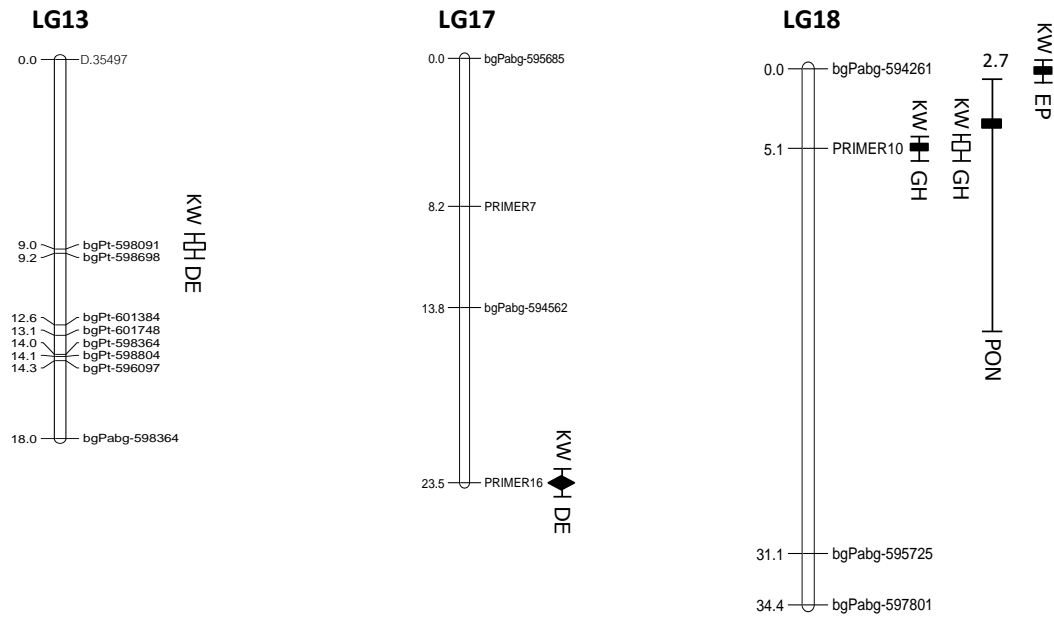
**Figure 6-1: Map positions of the QTL on the 'narrow' cross (DipC x Tiga necaru) linkage map. Only linkage groups for which 'putative' or 'significant' QTL were found are shown. Positions are given in cM (Haldane units), to the right of the linkage groups. QTL terminology is described in the text. The position of the maximum LOD value of a particular QTL is written at the top of QTL pointer and indicated by a rectangle, in black for glasshouse experiment, white for field data and in black four pointed star for the $F_2$ progeny. QTL confidence intervals (1 LOD drop-off) are represented by plain lines with LOD score $\geq$ significant threshold value, determined by 10,000 permutation tests and by dotted lined for putative QTL below the significant threshold LOD value, but with LOD scores >2. QTLs detected with Kruskal-Wallis analysis are discontinuous with no confidence interval and the rectangular or four pointed star only represents the entire position of QTL.**

*Leaf area* (cm$^2$): Based on glasshouse data a putative QTL for leaf area was identified on LG3 at 16.6 cM. This accounted for15.5% of the phenotypic variation of leaf area. The marker bgPabg-597113 at 15.6 cM is close to this QTL (K*= 6.5 at p =0.05).

*Plant spread* (cm): Interval mapping revealed a single QTL controlling *plant spread* on LG4 for the glasshouse data. This QTL was associated with the marker BN6b and had a LOD score of 3.9 and explained 24.6% of total phenotypic variation. A strong QTL-marker association was confirmed using non-parametric mapping (K* = 16.5 at

196

p = 0.0005). In the $F_2$ progeny dataset the marker locus BN6b had a LOD score of 2.1 at the same position for *plant spread*, whereas the highest detected LOD value of 3.2 was at 33.5 cM, explaining 18% of the phenotypic variation and associated with marker bgPabg-597624 (K* = 14.8 at p = 0.0005).

**Stem no./plant**: A putative QTL was mapped on the LG4 for *stem no./plant*, located at 14.2 cM for the glasshouse dataset. A peak LOD value of 2.4 was slightly lower than the genome-wide permutation test threshold (2.6; 10000 permutation). The closest marker to this genomic region was bgPt-600898 at 11.2 cM position (K* = 4.8 at p = 0.05).

**Node no./stem**: Glasshouse data analysis detected two QTLs for this trait on LG1 and LG4. Both QTLs accounted for about 39% of total phenotypic variation. While $F_3$ progenies evaluated in the field suggested a major QTL on LG3 associated with *node no./plant*. This QTL was linked to marker bgPabg-595707 at a LOD value of 2.8 at 30.2 cM and accounted for 18.4% of the total phenotypic variance of *node no./stem*. The trait-marker association was also revealed through non-parametric mapping to be significant (K* = 10.8 at p = 0.005).

**Internode length** (**cm**): Data analysis for the FutureCrop glasshouse and the field detected a major QTL for *internode length* mapped on LG4, with the highest LOD values of 7.9 and 7.1, respectively. This significant QTL was located at 3.0 cM and was associated with marker bgPabg-596988 and accounted for the greatest effects observed of 43.5% and 41% of total phenotypic variations in both datasets, respectively. KW analysis also identified the strongest association seen with *internode length* at this locus (at p = 0.0001).

***Growth habit*:** Non-parametric mapping of the $F_3$ generation in FutureCrop glasshouses and the field showed a very strong association between this trait and marker BN6b on LG4. The association between this SSR marker and two other traits of *plant spread* and *double seeded pods/plant* was identified in the interval mapping analysis. Two other markers were linked to growth habit on LG10 and LG18 at different level of significant p value, whereas the field evaluation revealed that marker bgPabg-596988 was identified as the most strongly associated with this trait. This marker was potentially linked to *internode length*, being a major QTL in both glasshouse and the Indonesian field. Two other markers associated with *growth habit* were identified on LG18 (as detected also in the glasshouse) and LG14.

***Pod no./plant*:** Two putative QTLs for *pod no./plant* were detected on LG1 at different locations for $F_3$ and $F_2$ progeny data analysed in FutureCrop glasshouse and TCRU, respectively. Interval mapping for the field dataset indicated another putative QTL for this trait on LG18 at 3.0 cM with a LOD value of 2.5. This locus explained 15.8% of phenotypic variation and was close to an SSR marker (primer10) (K* = 9.8 at p = 0.01). The locus of primer10 was also associated with *growth habit* in nonparametric mapping of both glasshouse and field data.

***Double seeded pods/plant*:** A QTL was detected for *double seeded pods* at 1.0 cM on the LG4, with a LOD score of 3.4 for the glasshouse $F_3$ dataset. This locus explains 22% of phenotypic variation. Marker BN6b was the nearest to this QTL and has been also associated with *plant spread*. Analysing $F_2$ progeny data detected a QTL for *double seeded pods/plant* on the same LG at a different location with a LOD score of 3.2 and an additive mode of action.

***Peduncle length* (cm):** A major QTL was detected for *peduncle length* analysing the data of the $F_3$ progeny recorded in the glasshouse. It was located on LG4 with a maximum LOD score of 9.7 at 1.0 cM. This QTL explained 50.3% of the phenotypic variation in the glasshouse dataset. The marker bgPt-423527 at 2.4 cM was the nearest to this QTL (K* = 8.1 at p = 0.005).

***Pod weight* (g/plant):** Interval mapping located a single QTL for *pod weight* in the $F_3$ population grown in the FutureCrop glasshouse with LOD 2.6 at 33.0 cM on LG1. This genomic region explained 17% of the phenotypic variation and was associated with marker bgPabg-596774. This marker was also detected as associated with a putative QTL for *pod no./plant*.

***Pod and seed dimensions*:** Analyzing the segregating $F_3$ population grown in the glasshouse detected four significant QTLs for *pod length*, *pod width*, *pod length with double seeded* and *pod width with double seeded* positioned close to each other. Marker bgPt-598767 was identified to be the nearest to the maximum LOD score of *pod length*, *pod width of double seeded* and *seed width* at 12.9 cM, whereas two other QTLs for *pod length with double seeded* and *seed length* were also mapped on LG1 and LG10, respectively. Based on the field dataset, another significant QTL for *pod length* was detected on LG11. The phenotypic variance explain by these QTL ranged from 13.5-33.4%.

***Seed no.* and *seed weight* (g/plant):** On LG1 putative QTLs were detected for *seed no./plant* in the $F_3$ population of glasshouse and $F_2$ progenies of TCRU at 34.0 cM and 72.6 cM, respectively, although both had LOD scores under the significance threshold. Marker bgPabg-596774 at 33.0 cM was found to be the nearest to this locus in the $F_3$ generation and also detected as a significant QTL for *seed weight*.

***Biomass dry weight* (g/plant):** Interval mapping analysis of the $F_3$ progeny grown in the glasshouse identified a major QTL for total biomass located on LG1. It was tightly linked with the marker bgPabg-596774 located at 33.0 cM with a LOD score of 3.5 explained 22.4% of the phenotypic variation. This genomic region was identified as containing 5 QTL for other traits (as a major QTL for *node no./stem*, *pod weight* and *seed weight* and as a putative QTL for *pod no./plant* and *seed no./plant*. Kruskal-Wallis analysis for non-parametric mapping also confirmed the significant association between this trait and the marker reported in the interval mapping.

A single QTL for *biomass dry weight* in the field trial was also detected on the same linkage group (LG1) at different positions; this region was linked to marker bgPt-602039, while analysis of data for the $F_2$ population detected a putative QTL for this trait at a different position on the same linkage group.

***Shelling percentage*:** Interval mapping analysis of segregation in the $F_3$ and $F_2$ populations evaluated in glasshouse and TCRU indicated that *shelling percent* was effected by major QTLs. These two QTL mapped on LG7 and LG12 with the LOD score of 3.0 and 4.8, for both glasshouse and TCRU datasets, respectively. The high significant trait-marker association was confirmed through non-parametric mapping for both of QTLs.

***100-seed weight* (g):** One significant QTL was detected for *100-seed weight* in the $F_3$ generation in the glasshouse dataset. It was located on LG7 at 9.4 cM with a LOD score of 2.7. Analyzing the $F_2$ generation data suggested a putative QTL for this trait which mapped on LG11 and LG12 at a LOD peak recorded under the significant threshold value genome-wide (2.5).

***Eye pattern around hilum*:** Non-parametric analysis revealed significant association between this trait and two DArT markers on LG12 and LG18. Marker bgPabg-594999 was found to be the most significantly linked (K* = 29.7 at p = 0.0001) on LG12 at 22.5 cM.

### 6.2.2. QTL analysis in the 'wide' cross map

A number of traits evaluated previously for the $F_2$ population of the inter-specific cross between DipC x VSSP11 (Basu, 2005) were analysed using the improved map which combined three marker types (SSR, DArT and AFLP). These traits included *days to emergence*, *days to flowering* from emergence, *stems/plant, internode length*, *leaf area*, *specific leaf area* (SLA), *carbon isotope discrimination* (CID), and *100-seed weight.*

*Days to emergence* and *days to flowering* were both non-normally distributed and analysied with nonparametric mapping. *Internode length* data was used directly without transformation as the normality thresholds was *P* > 0.01 and *A-square* (0.89) was smaller than the 99% critical value (1.09). For *stem no./plant* transformed data was distributed normally and analyzed with parametric mapping, other traits were distributed normally and they were used directly for parametric analysis.

### Marker and trait association

The LOD threshold for detection of putative QTLs was peak $\geq$ 2 for parametric results. Significance thresholds were determined by permutation testing of 10000 replications. Thresholds ranged from 3.1 to 3.6.

***Days to emergence***: An association between emergence and AFLP marker P19b494AGC+CTA4 was identified on LG7 (K\* = 7.1 at p = 0.01), using non-parametric mapping. Other associations between marker P3b211AGC+CTC7 and this trait was detected on LG14, but scored a very low K\* = 4; p = 0.05 (Table 6-5).

**Table 6-5: QTL mapping in bambara groundnut using interval mapping and Kruskal-Wallis analysis for the $F_2$ generation data derived from the 'wide' cross of DipC x VSSP11**

| Trait | Linkage group | Position (cM) | Locus | Interval mapping LOD | PT[a] | PVE[b] | Additive | Kruskal-Wallis analysis K\* | Df | Signifi. level |
|---|---|---|---|---|---|---|---|---|---|---|
| *Days to emergence* | 7 | 93.9 | P19b494AGC+CTA4 | | | | | 7.1 | 1 | *** |
| (DE) | 14 | 0.0 | P3b211AGC+CTC7 | | | | | 4.0 | 1 | ** |
| | 10 | 18.0 | P6b140AAC+CAG1 | | | | | 6.8 | 1 | *** |
| *Days to flowering* | 7 | 70.6 | P18b320ACA+CAT6 | | | | | 5.7 | 1 | ** |
| (DF) | 15 | 67.1 | P15b351AGT+CAA2 | | | | | 4.9 | 1 | ** |
| | 14 | 64.1 | GH-19-B2-D9 | | | | | 4.8 | 2 | ** |
| *Leaf area* (LA) | 15 | 83.1 | | 3.0 | 3.4 | 14.1 | -17.6 | | | |
| | 15 | 94.6 | P19b105AGC+CTA2 | 2.3 | | 11.2 | -22.5 | 5.27 | 1 | ** |
| *Specific leaf area* | 10 | 7.0 | | 3.0 | 3.2 | 14.6 | -9.3 | | | |
| (SLA) | 10 | 12.4 | P16b329AGA+CTC2 | 2.6 | | 12.7 | -7.8 | 9.9 | 1 | **** |
| *Stem no./plant* (STN) | 13 | 1.1 | P17b242ACG+CAA3 | 4.9 | 3.6 | 22.2 | -2.6 | 15.7 | 1 | ******* |
| | 1 | 20.9 | 595196 | 2.4 | | 11.8 | -2.1 | | | |
| | 9 | 0.0 | 600900 | 2.9 | 3.1 | 14 | 1.8 | | | |
| *Internode length* (IL) | 14 | 26.5 | P19b102AGC+CTA1 | 2.5 | | 12 | 1.6 | 10.1 | 1 | **** |
| | 13 | 0.0 | P17b240ACG+CAA2 | 2.4 | | 11.8 | 1.5 | | | |
| *Carbon isotope discrimination CID* | 7 | 93.9 | P19b494AGC+CTA4 | 4.7 | 3.3 | 22.1 | 0.0 | 17.4 | 1 | ******* |
| *100-seed weight* | 7 | 87.6 | | 3.3 | 3.6 | 16.9 | -2.6 | | | |
| (HSW) | 7 | 93.9 | P19b494AGC+CTA4 | 3.0 | | 15.7 | -4.0 | 12.0 | 1 | ***** |

a : permutation-10000 times test

b : percentage of total phenotypic variation explained by the QTL

Significant level of K * values: *: 0.10, **: 0.05, ***: 0.01, ****: 0.005, *****: 0.001, ******: 0.0005, *******: 0.0001

***Days to flowering***: Kruskal-Wallis mapping indicated that *days to flowering* was associated with marker P6b140AAC+CAG1 on LG10 (K\* = 6.8 at p = 0.01). Three other associations with both AFLP and SSR markers on three different linkage groups of 7, 14 and 15 were also identified (Figure 6-2).

***Leaf area* ($cm^2$):** QTL analysis identified a putative QTL for leaf area on LG15 with a LOD score of 3.0 at 83.1 cM and this explained 14.1% of the total phenotypic

variance for this trait. Marker P19b105AGC+CTA2 was detected to be the nearest to this locus.

*Specific leaf area* (SLA): Interval mapping analysis identified a putative QTL for SLA located on LG10. The LOD score of 3.0 was under the genome-wide significance threshold value of 3.2, determined by permutation testing. The nearest marker to this genomic region was P16b329AGA+CTC2 which showed an association with the trait studied in nonparametric mapping (K* = 9.9; p = 0.005).
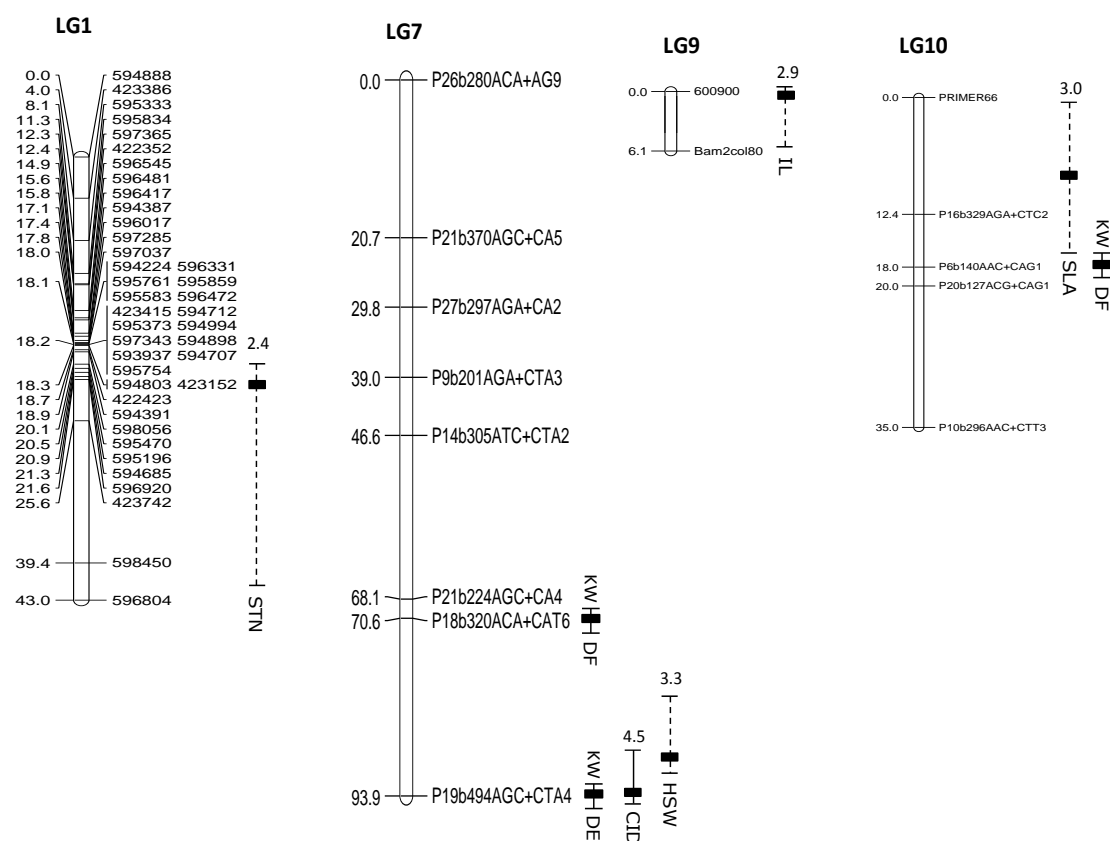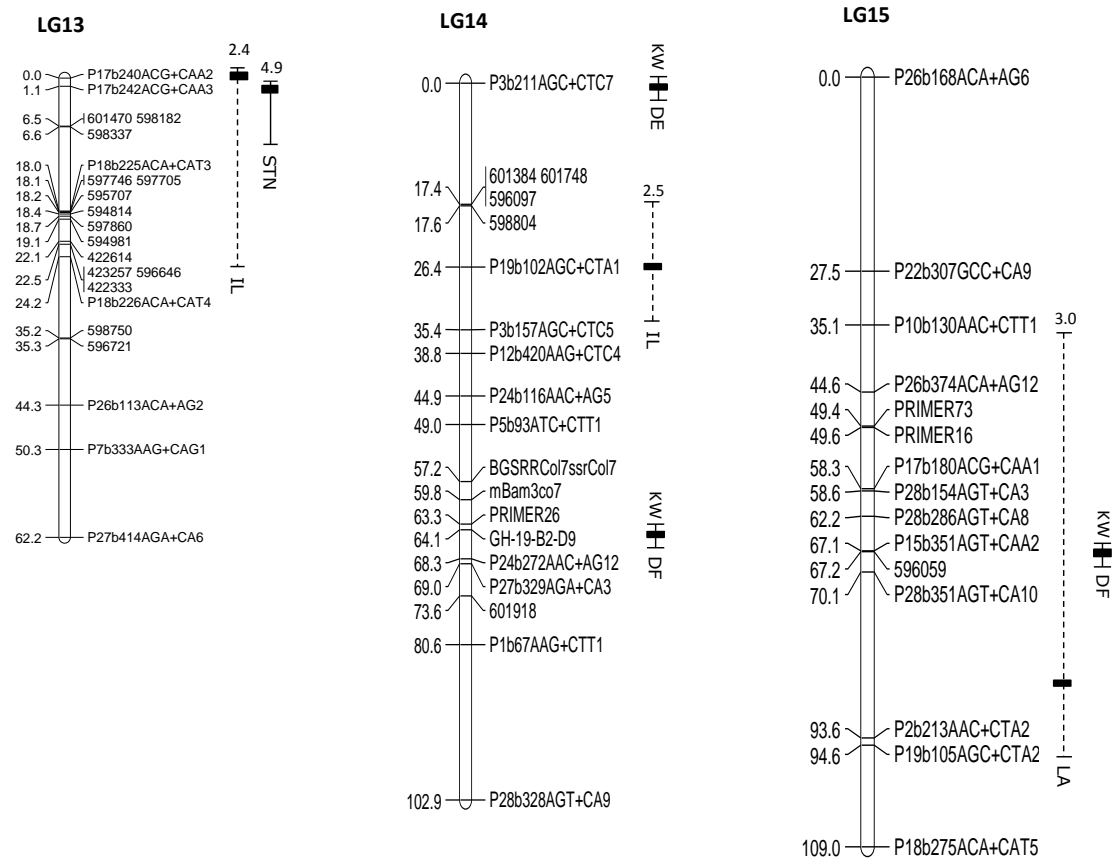
**Figure 6-2: Map positions of the QTL on the wide cross (DipC x VSSP11) linkage map. Only linkage groups for which QTL were found are shown. Positions are given in cM (Haldane units) to the right of the linkage groups. QTL terminology is described in the text. The position of the maximum LOD value is written at the top of the QTL pointer and indicated by a rectangle in black. The maximum confidence LOD values of a particular QTL (1 LOD drop-off) are represented by plain lines with LOD score ≥ significant threshold value, as determined by 10,000 permutation tests and by dotted lined for putative QTL below the significant threshold LOD value but ≥ 2 LOD score. QTLs detected with Kruskal-Wallis analysis are discontinuous with no confidence interval and the rectangular only represents the entire position of QTL.**

*Stem no./plant*: A significant QTL on LG13 was mapped using interval mapping at a LOD score of 4.9. The marker P17b242ACG+CAA3 was found to be the most significantly linked (K* = 15.7; p = 0.0001) locus to this QTL for *stem no./plant* which was mapped at 1.7 cM and explained 22.2% of total phenotypic variance.

Another putative QTL was detected for this trait on LG1 at LOD = 2.4. DArT marker 595196 was linked to this trait at 20.9 cM.

***Internode length*** (cm): Interval mapping for the inter-specific wide cross population derived from DipC and VSSP11 detected a putative QTL for internode length on LG9 at 0.0 cM with a LOD score of 2.9. Marker 600900 was linked to this locus and the phenotypic variance explained was 14%. Two other putative QTLs for internode length were located on LG13 and LG14 and were in association with AFLP markers P17b240ACG+CAA2 and P19b102AGC+CTA1, respectively. These three loci together explained about 38% of the phenotypic variation of *internode length* observed in this population.

***Carbon isotope discrimination*** (CID): A single QTL for CID was identified on LG7 at 93.9 cM with the LOD sore of 4.7. It was found to be associated strongly with marker P19b494AGC+CTA4 ($K^* = 17.4$ at $p = 0.0001$). This genomic region explained 22.1% of phenotypic variance observed for CID in the $F_2$ progeny of the wide cross (DipC x VSSP11).

***100-seed weight*** (g): Interval mapping also revealed a putative QTL for *100-seed weight* on LG7 at 87.6 cM, having a LOD score of 3.3 (significant threshold value = 3.6; by permutation, 10,000 replications). The marker P19b494AGC+CTA4 was the nearest to this QTL ($K^* = 15.7$; $p = 0.001$) and the interval mapping suggested that this explained 16.9% of total phenotypic variance. This marker was also found to be associated with the QTL of CID

## 6.3 Discussion

The majority of biological traits are genetically complex. Mapping quantitative trait loci (QTL) is a powerful means for estimating many parameters of the genetic architecture for a trait and potentially identifying the genes responsible for the expressed phenotypic variation. QTL mapping is a key tool for studying the genetic architecture of complex traits in plants, facilitating estimation of the minimum number of genome regions that affect a trait, the distribution of gene effects, and the relative importance of additive and non-additive gene action (Laurie *et al.*, 2004; Mackay, 2004).

The populations used for the map construction of both 'narrow' and 'wide' cross were segregating for a number of agronomic traits, which either showed a continuous or discrete distribution.

### 6.3.1 QTL detected

In order to obtain more reliable QTLs for the agronomically important traits a single QTL-model analysis of Interval Mapping was used to try to identify QTL for all the trait data that followed a normal distribution. The Kruskal-Wallis test of Marker-QTL associations was implemented for non-normally distributed traits at a significance threshold $p<0.05$ (Van Ooijen and Maliepaard, 2001).

### 6.3.2 LOD significance threshold for QTL detection

It was possible in the current study to carry out a permutation test to determine the empirical significance thresholds (Churchill and Doerge, 1994). MapQTL6 offers this

test for interval mapping to determine the significance threshold of the LOD score. The significance threshold in this study varied from 2.4 to 3.6 for different trait in different mapping populations. They were determined based on the actual data rather than an assumed normally distributed data, being freedom from constraints on probability distribution and size of samples (Senko and Kuznetsova, 2006). Calculation of this threshold was dependent on the genome-wide and population type under study. LOD scores greater than the genome-wide significance threshold value was considered as 'significant QTL' for traits studied and those above LOD = 2, but below the calculated thresholds as 'putative QTL'.

### 6.3.3 QTLs detected in the 'narrow' cross map

A stable QTL was detected for *internode length* in the $F_3$ population of the narrow cross for both glasshouse and field trials dataset. The phenotypic variation explained in association with the marker bgPabg-596988 was >40%. The significant association of this trait with both *seed no./plant* and *seed weight/plant* could be of importance, as selection for *internode length* should help to improve seed yield in bambara groundnut by selection for this early trait. *Internode length* is considered as one of the most variable traits between different landraces and could be important in selection of genotypes in a breeding program (Siise and Massawe, 2012). *Internode length* also had a strong negative correlation with *growth habit* in both environments (r = -0.8) and this association might be used to structure the plant architecture according to the target environment. Bunch types appear easier to manage in mixed cropping systems in low input subsistence farming. While for controlled crossing spreading types could be easier to deal with because the flowers of bambara groundnut are very small and are more compact in the bunch type, while genotypes with longer *internode lengths*

offer easier access to flower buds for emasculation and pollination. Furthermore, QTL detected for *internode length* has indicated a positive additive effect for the alleles contributed from the parents to increase the internode length (Table 6-2 and Table 6-3). The finding of residual internode variation in a domesticated x domesticated cross is interesting, as it is one of the major domestication traits detected (Basu *et al.*, 2007c) and suggests that variation for this trait has not been completely lost.

Both position and the magnitude of the QTLs for internode length on the map were stable across FutureCrop glasshouse and Indonesian field for the $F_3$ population derived from 'narrow' cross. This QTL could be a good candidate for MAS in the yield enhancement programs to help to suit different morphotypes to different environmental conditions. It could also be used to develop material for better management of this crop in the field.

Interval mapping analysis of the $F_3$ generation in the FutureCrop glasshouse for the 'narrow' map recorded the highest QTL LOD peak (9.7) for *peduncle length* on LG4. A significant QTL for *double seeded pods/plant* was also found in this genomic region and its association with *peduncle length* was identified previously through their significant trait correlation (r= +0.53 at p= 0.000). Higher additive effects compared to the dominance effects associated with the QTL of these two traits were detected and found to be in the same direction of effect, which is consistent with their associations (Appendix 27). Marker bgPt-423527 was the nearest marker to this locus and only 0.5 cM from the centre of the main QTL for *internode length*. The level of pods produced under the soil surface could be predicted through measurement of *peduncle length* in the population grown in FutureCrop glasshouse. Those lines having longer peduncles located the pods under the soil surface. Burying pods under the soil is considered to be an advantage of groundnut compared to other legumes,

making harvesting time less critical, as the pods can be left in the soil for considerable time without serious loses (Whitmore, 2000). This trait is also associated in this cross with *biomass dry weight* and *seed weight/plant*, through an increase in pods containing double seeds for those lines having a longer peduncle. All of these trait relationships were previously suggested through Pearson's correlation coefficient analysis. Thus, this QTL could be one of the candidates for marker assisted selection for the yield in bambara groundnut.

An emphasis on *biomass* and distribution from the source into different sinks is a key focus of breeding programmes. QTL analysis can assist in understanding the genetic basis for this and how the environment may effect this accumulation and distribution of carbon. The marker bgPabg-596774 on LG1 is a target marker for a number of QTL, being a major QTL for *biomass dry weight*, *node no./plant*, *pod weight* and *seed weight/plant* and as a putative QTL for *pod no.* and *seed no./plant.* Their relationships have been confirmed through strong trait associations in FutureCrop glasshouse data for the narrow cross population. These associations are reflected in the direction of additive effects observed for the coincident QTLs of these traits. This situation of multiple traits affected by a single marker bgPabg-596774 might come from the linkage of multiple QTLs or it could be pleitropy; a single gene affecting several biochemical pathways leading to many different phenotypes at different levels of organization. This has been identified in a wide range of species (Prokop, 2004). Previous studies on soybean also identified more than one QTL for different traits mapped on the same loci (Wenxin *et al.*, 2008; Zhang *et al.*, 2004).

Microsatellite marker Bam2coL63 is found to be a candidate marker for the QTL of *leaflet length* analysed in both $F_2$ and $F_3$ populations in the controlled environment glasshouses for the 'narrow' cross population. This marker could potentially be used

to predict *biomass dry weight* as *terminal leaf length* and *biomass dry weight* are associated traits (r= +0.273; p=0.019 and +0.441; p= 0.000, respectively).

In total 18 significant QTL and 7 putative QTL were detected using interval mapping in the 'narrow' cross map, in addition to loci for *days to emergence, growth habit* and *eye pattern around hilum* detected with nonparametric mapping. Most of the QTLs detected were clustered on linkage groups 1, 4 and 12. A major QTL for *internode length* was detected on LG4 of the $F_3$ generation grown in glasshouse and in the field. The highest LOD value of 9.7 for *peduncle length* was located close to the *internode length* locus. A QTL at marker locus bgPabg-596774 was also detected for *seed yield* and some other yield-related traits located on LG1.

## 6.3.4 QTL analysis in the 'wide' cross map

 QTL detection was based on LOD thresholds estimated by permutation tests (10000 permutations, $P = 0.05$), to account for the effects of known major QTL of multiple tests through their association with genetic markers for the entire experiment (Doerge and Churchill, 1996). The thresholds adopted here ranged from LOD = 2.9-3.6 for different traits studied in this cross. Two major QTLs were detected for the traits of *stem no./plant* and *CID* on LG13 and LG7 (LOD= 4.9 and 4.7), respectively. However the QTLs detected for *leaf area*, *specific leaf area* and *100-seed weight* had LOD scores $\geq$ 3, they were considered as putative because they were detected with LOD score lower than significant threshold value of permutation tests. Two and three QTLs for the traits *stem no./plant* and *internode length* were located on different linkage groups, respectively. Only two traits (*days to emergence* and *days to flowering*) were not normally distributed and their association with markers on different linkage groups was determined based on Kruskal-Wallis analysis with a

significance level of 0.05- 0.01. Insufficient parental data on *days to emergence* and *days to flowering* hindered the identification of discrete classes of parental values to analyse the segregation of $F_2$ data for these two traits (Basu, 2005). The AFLP marker P19b494AGC+CTA4 located at 93.9 cM was detected as the reference marker for the QTL effect on *days to emergence, CID and 100-seed weight*. The significant association between these three traits was detected previously by Basu (2005) using Pearson's correlation coefficient analysis.

Due to the difficulty in characterizing phenotype of traits based on single plants the wide cross linkage groups probably harboured fewer QTLs for agronomic traits. The lack of replication is considered one of the major disadvantages of using $F_2$ population in QTL analysis, because the phenotypic evaluation on single plants is usually not considered reliable for some quantitative traits (Rakshit *et al.*, 2012; Semagn *et al.*, 2010). It has been suggested that studies conducted in a single environment are likely to underestimate the number of QTLs that can influence a trait (Paterson *et al.*, 1991). For such environment-specific QTL, one would only be able to know that the QTL acts at locations where the environmental conditions are the same (Bolek *et al.*, 2005). Moreover, the limited number of common markers between 'wide' and 'narrow' maps for both SSR and DArT markers has also limited the attempt to combine linkage groups of these two maps, which has affected the proper identification of links between the QTLs detected on both maps. The only link detected between the QTLs of the combined maps of both crosses was for *days to emergence* on the linkage groups 13 and 14 of 'narrow' and 'wide' cross maps, respectively. The QTL for this trait in both linkage groups was in close association with the common DArT markers 601384 and 601748 and these two linkage groups were joined together in the combined map of group 6 (Figure 6-3).
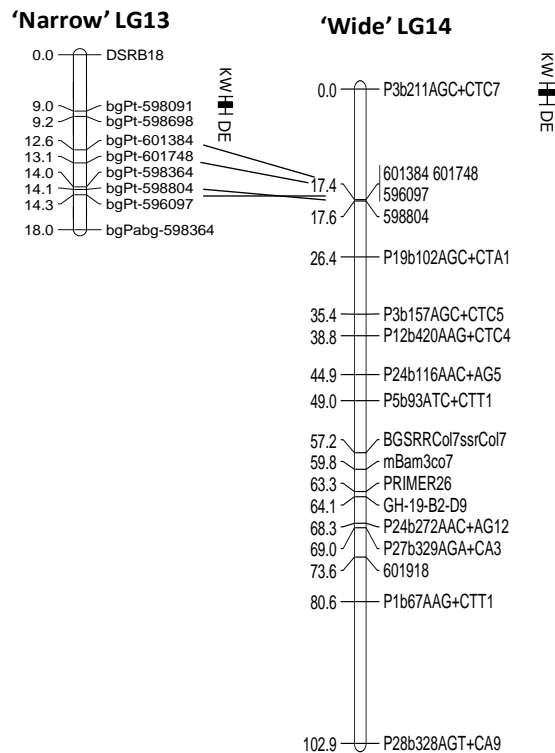
**Figure 6-3: Related linkage groups identified through comparison of markers in the populations derived from both wide and narrow crosses in bambara groundnut.**

One hundred and seven markers were not mapped to any linkage groups in the 'wide' map. The largest number of unmapped markers were AFLP markers (64 out of 141 markers) followed by DArTs (30 out of 106), although the highest proportion of unmapped markers (54%) was for SSR markers from the total number of 24 markers involved. It is expected that increasing the number of markers and the addition of new marker types to this population such as conserved ortholog set markers from other related legumes (soybean and *Medicago*) could help to identify and locate other QTLs, by increasing the coverage.

212

## 6.3.5 Summary

The results of QTLs analysis are presented and several genomic regions containing QTLs were found to explain variation for agronomic traits. The aim of this work has been to provide an understanding of biological mechanisms controlling these traits. The 'wide' map constructed between a wild and cultivated bambara groundnut accession has been used to determine the basis for domestication traits (Basu *et al.*, 2007c; Basu, 2005). While the 'narrow' map constructed in a domesticated (DipC) x domesticated (Tiga necaru) cross was expected to segregate for agronomic traits, rather than domestication or major morphological traits.

Microsatellite markers involved in constructing the linkage map were derived from both genomic and transcriptomic regions. No marker-trait association was identified for transcritome SSR markers in the current study for both 'narrow' and 'wide' map. The possible reason is that both maps contain a limited number of transcriptome SSRs (2 and 3) for 'narrow' and 'wide' linkage map, respectively, and they might be associated with other traits not investigated here.

However, while the populations under study here were not big enough to construct a high resolution map for QTL study (Collard *et al.*, 2005), the loci in the framework maps of 'narrow' and 'wide' maps were within 10 cM (3.04 cM and 5.91 cM, respectively) generally recommended for QTL analysis. The QTL identified could provide a resource for identifying the regions of bambara genome which contain genes for agronomically important traits and developing molecular markers for MAS in the breeding program of this crop. It has been declared that uniformly distributed loci every 10–20 cM within the entire genome could give a significant increase in the relative effectiveness of MAS and QTL identification (Stuber *et al.*, 1999).

# Chapter 7.  GENERAL DISCUSSION

## 7.1 Introduction

Investigating the genetic architecture of crop plants has paid significant dividends in the case of many crops and especially cereals, such as wheat and rice, to meet the continually growing needs of Humankind in terms of increased yield and improved quality (Cholin, 2009). Plant breeding efforts are yet to make such an impact in the case of legumes as here yields have not kept pace with those of cereals. Despite the reported drought tolerance in bambara groundnut coupled with reasonable resistance to diseases, pests and adaption to poor soils, crop production is at relatively low levels, possibly due to low and unpredictable yields (Mayes *et al.*, 2009). Bambara groundnut is cultivated throughout tropical regions of Africa and considered as an important food security crop in sub-Saharan Africa. The farmers there still rely on local landraces which can be inherently low yielding due to poor physical and genetic quality of seeds and poor crop management. Development of high yielding and adapted genotypes of bambara groundnut is an important strategic approach to try to increase food security through use of bambara groundnut.

Most of the agronomically important traits in bambara groundnut are quantitatively inherited (Olukolu *et al.*, 2012). Due to the genotype-by-environment interaction in these traits most yield components have low heritable variation. Hence, phenotypic selection based on conventional breeding techniques alone will have some limitations in breeding programs for yield improvement in this crop. Employing new biotechnological tools such as DNA markers for mapping and identifying genes for desirable traits could solve some aspects of this problem. Construction of molecular

linkage maps is now a routine way to trace valuable alleles in a segregating population.

In this study we aimed to generate a framework map of an $F_3$ segregating population derived from the intra-subspecific cross (DipC x Tiga necaru) and to improve the partial genetic map developed previously (Basu, 2005) from an inter-subspecific cross between a domesticated landrace and a wild ancestor. A number of microsatellites (derived from different microsatellite-enriched libraries) and Diversity Arrays Technology (DArT) markers were used to carry out map construction and allow a QTL analysis of agronomically important traits. An attempt to combine the 'wide' cross and the 'narrow' cross was made to identify common markers to QTL effects for bambara groundnut and to improve our understanding of domestication. This could help breeders to effectively pyramid genes for agronomically important traits into single cultivars in a much shorter time than would be possible by conventional breeding.

## 7.2 Molecular markers development, characterization and validating the $F_3$ population under study

An overview of molecular marker (SSR and DArT) characterization and the polymorphic markers used in both 'narrow' and 'wide' cross populations under study is described in Figure 7-1.
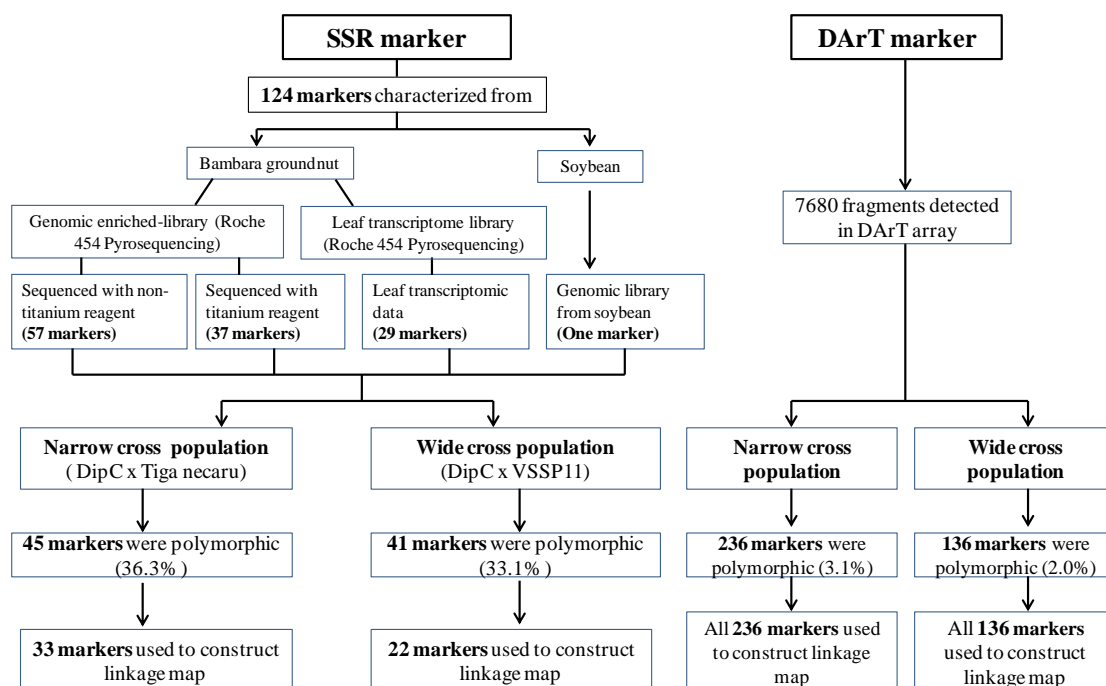
**Figure 7-1: Diagram of SSR and DArT markers, their characterization and the number of polymorphic markers used in both 'narrow' and 'wide' cross populations of this study.**

## 7.2.1 SSR markers

Microsatellite markers can be found in both coding and non-coding regions in all prokaryotic and eukaryotic genomes (Kalia *et al.*, 2011; Trivedi, 2004). The markers used in this study were derived from different microsatellite-enriched libraries of both regions. Out of a total 124 SSRs screened 94 markers were developed from a genomic-enriched repeat library sequenced with non-titanium and titanium reagents (Roche 454 Pyrosequencing), representing the first two sets. The other set of 29 SSR markers were developed from a bambara groundnut leaf transcriptome library (Roche 454 Pyrosequencing). In addition, one SSR marker from soybean was also used in this investigation. Transferring SSR markers between the legume crops has become more common recently through linkage mapping and assessing genetic diversity (Chandra, 2011). A high rate of amplification of SSR markers from other legume species (azuki

bean, cowpea, and mungbean) in bambara groundnut was also reported by Somta *et al.* (2011b). Wang *et al.* (2004a) suggested that transferring SSRs across species and genera could be an efficient approach to develop DNA markers, especially for minor crops.

Commercial application and quick detection of SSR allele using agarose or simple polyacrylamide gel systems can be achieved by any laboratory with minimum infrastructure, meaning that this technique has been widely used in studies of genetic diversity (Senior *et al.*, 1998; Stewart *et al.*, 2011). However, such systems cannot resolve small difference in allele size (Jones *et al.*, 1997).

Agarose gel (2.5%) was used in our investigation to detect the parental alleles in the populations under study for 4 primer pairs (D.35497, Primer10, Primer82 and D.48339) which had 34bp, 31bp, 28bp and 36bpbp allelic differences, respectively. A 3% agarose gel has been used by a number of researchers to screen for polymorphism of SSR markers with more than 20bp differences in allelic size (Ashkani *et al.*, 2012; Beyene *et al.*, 2005; Legesse *et al.*, 2007; Wietholter *et al.*, 2008).

Allelic size differences for other primers in both 'narrow' and 'wide' cross populations which could not be clearly scored with an agarose gel were determined using the CEQ$^{TM}$ 8000 (Genetic Analysis System, Beckman Coulter, USA). The product sizes were scored manually from the electrophoretograms, taking into account the entire microsatellite profile as the automated calling system is unable to cope simply with changes in relative peak heights between alleles and the shift in overall microsatellite 'shape' is more informative. To reduce the costs and facilitate screening of large numbers of potential microsatellites with the Beckman CEQ$^{TM}$ 8000, a three primer system was used (Schuelke, 2000). Further reduction in the cost of screening

the microsatellite markers was achieved by multiplexing PCR products which showed

a wide product size differences between microsatellites in their individual pre-screens

for size and polymorphism. Up to three SSRs were pooled. All the primers were

labeled only with fluorescent blue dye D4. This was to avoid the problem of colour

bleed through (e.g. as reported by Molosiwa, 2012) and the comparatively week D2

dye (Black).

**Polymorphism of SSRs**

Out of a total of 124 SSR markers, 45 were detected as polymorphic (36.3%) for the

parental alleles in the $F_3$ population derived from the cross between DipC and Tiga

necaru, while 41 markers were polymorphic (33.1%) in the $F_2$ population from the

cross between DipC and VSSP11. Low levels of genetic polymorphism in the specific

gene pool of bambara groundnut had been reported previously (Basu *et al.*, 2007a).

However, as all 124 markers had been prescreened against 24 genotypes believed to

reflect existing variation within bambara groundnut (Stadler, 2009), there is a

presumption that they are functional microsatellites. In an assessment of genetic

diversity in bambara groundnut by Somta *et al.* (2011b) 35 primers out of 188

amplified SSR markers (19%) were polymorphic in eight accessions collected from

the major growing regions of the world. Among these polymorphic markers, four

came from those previously developed in bambara groundnut (10 SSR in total) by

Basu *et al.* (2007a). The others were from related species: azuki bean, cowpea and

mungbean. Variable levels of polymorphism in SSR markers were detected in

different legume seed crops. Kajonphol *et al.* (2012) identified 16.1% polymorphic

SSR markers when screened in the parental genotypes of wild and cultivated

mungbean in an $F_2$ population using 945 markers. In a QTL study screening for

*Fusarium* wilt resistance gene in chickpea, Qadir *et al.* (2007) identified a high polymorphism rate (70.3%) for SSR markers between the parents of the mapping population (BG 256 and WR 315). The highest polymorphism rate (81.3%) was detected in pigeon pea by screening 40 genotypes (representing different *Cajanus* species including eight wild accessions) with 16 SSRs (Saxena *et al.*, 2010).

Individuals genotypes from the DipC landrace (*V. subterranea* subterranea) were involved in both 'wide' and 'narrow' crosses, being the female parent crossed with landraces of VSSP11 (*V. subterranea* spontenea) and Tiga necaru (*V. subterranea* subterranea), respectively. Higher polymorphism was expected in the inter-subspecific cross (DipC x VSSP11) than in the intra-subspecific cross (DipC x Tiga necaru) but a contrasting result was initially obtained with more polymorphic markers for narrow cross. Overall, similar levels of SSR polymorphism was observed in both crosses despite one segregating for domestication traits and the other for agronomic traits and diversity analysis (Stadler, 2009; Molosiwa, 2012) suggested that the DipC parent used in the 'wide' cross is actually closely related to the wild relative used in the same cross. Stadler (2009) also revealed considerable genetic distance between the DipC and Tiga necaru landraces when analyzing the genetic relationships of 87 bambara groundnut genotypes using 296 polymorphic DArT markers.

However, there was a small sample size of the common SSR markers (26 polymorphic SSRs in common from 45 and 41 polymorphic SSRs in the 'narrow' and 'wide' crosses, respectively) (Table 7-1). The authenticity of the DipC female parental genotype of both crosses was confirmed through amplification of the same allelic size with 17 common markers for both populations. Different allele sizes for the DipC parent in both populations with other common markers (9) is likely to come from the fact that different individuals of the DipC landrace were used as the maternal

parent. Massawe *et al.* (2005) reported genotype variations within bambara groundnut

landraces and DipC appears to be a particularly variable landrace (Mayes et al, 2009).

**Table 7-1: Common microsatellite markers and the alleles of the DipC parent used in both 'wide' and 'narrow' crosses**

| Primers | DipC parent used in narrow cross | DipC parent used in wide cross | allele | Source of primer |
|---|---|---|---|---|
| PRIMER10 | 260 | 268 | different | |
| PRIMER15 | 238 | 240 | different | |
| PRIMER19 | 273 | 270 | different | |
| PRIMER26 | 183 | 183 | match | |
| PRIMER32 | 247 | 220 | different | |
| PRIMER38 | 194 | 194 | match | |
| PRIMER48 | 238 | 244 | different | |
| PRIMER66 | 225 | 219 | different | |
| PRIMER98 | 264 | 274 | different | Genomic DNA library of bambara groudnut |
| GH-19-B2-D9 | 236 | 236 | match | |
| Bam2coL80 | 220 | 220 | match | |
| Bam2coL33 | 253 | 239 | different | |
| mBam3co7 | 267 | 267 | match | |
| BN145 | 150 | 143 | different | |
| PRIMER65 | 172 | 172 | match | |
| PRIMER16 | 189 | 189 | match | |
| PRIMER85 | 248 | 248 | match | |
| AG81 | 202 | 202 | match | Genomic DNA library of Soybean |
| D.42026 | 238 | 238 | match | |
| D.8148 | 244 | 244 | match | |
| D.8999 | 203 | 203 | match | |
| D.37053 | 181 | 181 | match | Leaf transcriptome library of bambara groundnut |
| D.12522 | 328 | 328 | match | |
| D.24269 | 246 | 246 | match | |
| D.51646 | 185 | 185 | match | |
| D.2094 | 224 | 224 | match | |

## 7.2.2 Validation of the F$_3$ populations under study

The segregating F$_3$ population was tested for residual heterozygosity before

construction of the genetic map. All 73 individual lines were screened with 33

polymorphic SSR markers. Residual heterozygosity of the population was found to be

24.9% which matches the predicted Mendelian ratios expected from the population

history. In addition, none of the lines was identified as being out-crossed. The tests

made confirmed that the population under study is an $F_3$ which can be used confidently in construction of a genetic map.

### 7.2.3 DArT assay

DArT as a high-throughput and hybridized-based microarray platform was developed to overcome some of the limitations of other molecular marker technologies, such as RFLP, AFLP and SSR (Akbari *et al.*, 2006). Since no DNA sequence information is required in developing DArT this technique allows for the typing of tens of thousands of loci in parallel with reduced cost per data point after the initial platform development (Kilian *et al.*, 2005; Wenzl *et al.*, 2004). It has proved to be the most cost-and time-efficient approach for under-utilized crops such as bambara groundnut (Mayes *et al.*, 2009; Olukolu *et al.*, 2012; Stadler, 2009). The procedure of generating DArT markers, screening for polymorphisms and genotyping was conducted by Diversity Arrays Pty. Ltd., Yarralumla, Australia. Due to the high locus specificity of these markers, they can be easily arranged into genetic linkage maps (Akbari *et al.*, 2006; Wenzl *et al.*, 2006). In the current study a total of 236 polymorphic markers (3.1%) were detected for the individuals of the $F_3$ population derived from the cross between DipC and Tiga necaru. They were integrated with polymorphic SSR markers to construct a genetic linkage map of bambara groundnut in this population. DArT was also used to generate additional markers for the existing initial map derived from the inter-subspecific cross in bambara groundnut with approximately 2% of DArT array elements revealing polymorphism. Higher levels of polymorphism at 5.5% were detected by Stadler (2009) in the construction of initial DArT marker discovery array from 38 landrace individuals using the restriction endonucleases combination *Pst*I/*Alu*I. However, this would be expected to reflect the overall levels of

polymorphism within the species, rather than in specific controlled crosses. In other crops such as barley, 2.9 to 10.4% polymorphism levels were found in discovery arrays prepared from nine 768-clone *Pst*I libraries from two genetically distant cultivars (Clipper and Sahara) (Wenzl *et al.*, 2004).

## 7.3 Rapid verification of presumed hybrids

Although a convenient hybridization system has been developed to obtain a large number of hybrid seeds in bambara groundnut (Suwanprasert *et al.*, 2006), the cleistogamous nature of *V. subterranea* flower causes a high rate of self-pollination in this crop in nature (Uguru and Agwatu, 2006). Due to the maternal control of seed coat colour, distinguishing between genuine $F_1$ hybrid seed and self-pollinated seed is not possible in the $F_1$ and the segregation of seed colour can only be observed in the $F_3$ seeds produced (Basu *et al*, 2007c). To shortcut this process PCR-based molecular marker fingerprinting systems have been developed (e.g. Sundaram *et al.* 2008). For further acceleration of this process we have investigated the possibility of extracting DNA directly from the seed endosperm instead of growing the presumed hybrids for DNA extraction, aiming to rapidly verify the $F_1$ seeds in a way that does not reduce their viability. Genomic DNA was extracted successfully using the GenElute Plant Genomic DNA kit (Sigma Aldrich) giving the greatest DNA yields with 30 minutes incubation. The viability of the seeds was checked by a germination test and statistical analysis revealed that germination frequency of the seeds was not affected by this treatment. A similar method has been used by Kamiya and Kiguchi (2003) to extract DNA from soybean seeds. This simple and rapid method could be a useful tool even in the next filial generations to perform marker-assisted selection, even before sowing of the seeds.

**7.4 Normality tests and association between the traits**

The distribution of trait data was tested using Anderson Darling Normality test for the $F_3$ populations derived from 'narrow' cross in both Future Crop glasshouse and the field trial, and for the $F_2$ population from the same cross evaluated in the TCRU. Normal distribution of traits was confirmed in most of the trait values over different environments for both $F_2$ and $F_3$ generations (*flower no./plant*, *plant height*, *petiole length*, *terminal leaflet length* and *width*, *leaf area, plant spread*, *stem no./plant*, *node no./plant*, *internode length*, *pod no./plant, double seeded pods/plant, peduncle length*, *pod weight/plant*, *pod length*, *pod width*, *double seeded pods length* and *width*, *seed length, biomass dry weight, shelling%* and *100-seed weight*), which indicates their control is more likely to be under multiple gene effects. Although some of these traits did not exhibit the same distribution pattern over all trials for both generations, Box-Cox transformation pulled many of them back to a continuous and normal distribution. Our results are in agreement with what was reported by other researchers on the effect of multiple additive genes for a number of these agronomic traits in bambara groundnut (Jonah *et al.*, 2012; Karikari, 2000).

The other traits of *days to emergence, days to flowering*, *growth habit* and *eye pattern around hilum* were characterized as discrete traits even after transformation. The trait distribution of both *growth habit* and *eye pattern around hilum* was found to follow a segregation pattern consistent with Mendelian inheritance. Incomplete and complete dominance were hypothesized to control both traits, respectively. The monogenic inheritance of *eye pattern around hilum* is in agreement with previous work by Basu (2005) in the segregating $F_2$ population derived from the 'wide' cross which has DipC as the maternal parent of the cross.

**7.5 Phenotypic correlation**

The low heritability of most quantitative traits of economic interest complicates the selection process, while understanding of their relationship with those of high heritability could aid indirect selection (Shimelis and Shiringani, 2010).

The characters studied here are most likely to have complex inter-relationships. In order to determine the phenotypic relationships that exist among traits, Pearson's correlation coefficient analysis was calculated for the population derived from intra-subspecific cross between DipC and Tiga necaru. Pearson's correlation coefficient describes the direction and degree to which one variable is linearly related to another (Bolboaca and Jantschi, 2006). This test was performed for the $F_3$ population evaluated in the FutureCrop glasshouse and the Indonesian field trial and also for the $F_2$ population derived from the same cross in the TCRU.

Most of the vegetative growth traits and yield component traits were in a strong association with yield in bambara groundnut for the different trials. These results are in accordance with earlier published association between these traits (Makanda *et al.*, 2009; Ntundu *et al.*, 2006; Ouedraogo *et al.*, 2008).

The negative association of *days to flowering* with the other traits, especially *seed yield* may lead to the hypothesis that the genotypes with early flowering may have a longer reproductive phase to deliver the stored and newly captured carbon into the seeds, leading to higher yields. Increasing the *leaf number*, *leaf area*, *internode length* and the optimal canopy spread should contribute positively to agronomical performance and increase the accumulation of biomass that goes into the seed.

A strong positive correlation was observed between yield and other traits studied here. *Pod no./plant* is a component of *seed yield* in bambara groundnut and considered as the most important yield component traits (Adu-Dapaah and Sangwan, 2004; Chijioke

*et al.*, 2010; Makanda *et al.*, 2009; Ofori, 1996). This has been confirmed in this study by the detection of a highly significant positive correlation (r=> 0.8 at p= 000) between both traits across different environments. It can been concluded that yield in bambara groundnut could be effectively improved through selection of these yield contributing traits. Given that bambara groundnut has a photoperiod requirement in some landraces for pod set/filling, The apparently self-evident correlation may not hold in non 12-hour photoperiod environments (Harris and Azam-Ali, 1993; Linnemann and Craufurd, 1994).

The association results between the traits studied was in agreement to the previously reported in the present of a strong correlation between the yield and other related traits in bambara groundnut. However these results needs further investigation to be validated, they may support the hypothesis that using early or simple traits associated with the later or more difficult will serve the assessment of the traits in selection of desirable genotypes for bambara groundnut.

## 7.6 Construction of genetic linkage map

The genetic studies reported earlier by Basu (2005) were based on an inter-subspecific cross. The genetic analysis of this cross has identified a number of genes important for the domestication process in Bambara groundnut (Basu *et al.*, 2007c). The second genetic map based on an intra-subspecific cross exploits variation within the domesticated landraces gene pool, using an $F_3$ population derived from the cross between two domesticated landraces (DipC and Tiga necaru). This cross was expected to show variation for traits of breeding interest (agronomic traits rather than domestication traits). An overview of both map construction and the attempt to combine the maps is described in Figure 7-2.
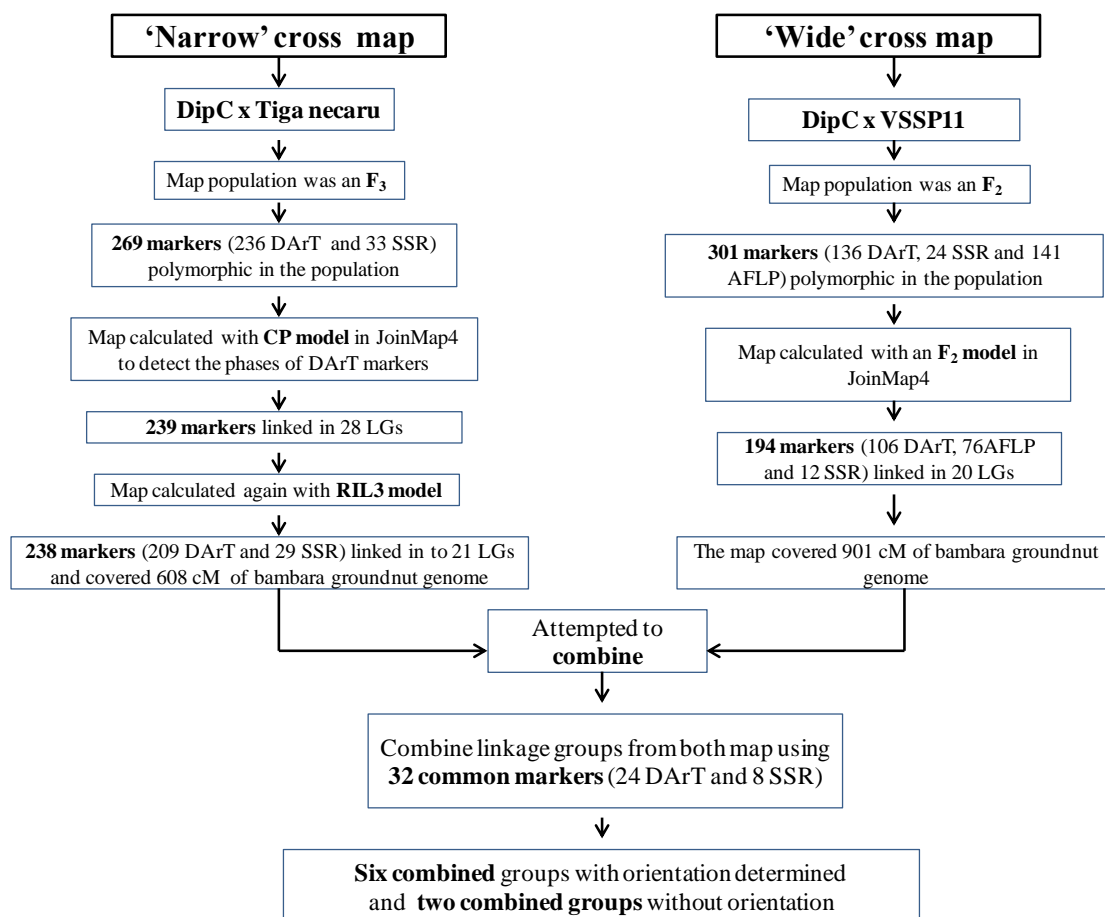
**Figure 7-2: A flow diagram of the process of construction of the genetic linkage maps in the 'narrow' and 'wide' cross population and their combination.**

## 7.6.1 Population size to construct the maps

A total number of 73 $F_3$ progeny were used to construct the 'narrow' cross genetic map. They were derived from a controlled cross between two single genotype accessions of domesticated background. Although, a similar population size of 75 recombinant inbred lines was used to construct a linkage map in common bean (Freyre *et al.*, 1998), larger populations are required to determine the marker order for high-resolution mapping (Mohan *et al.*, 1997). The size of population is considered the limiting factor for segregating populations and is determined by the number of seed from a single $F_1$ cross. As landraces are known to be composed of numerous

inbred lines in bambara groundnut (Massawe *et al.*, 2005), it is not possible to use multiple genotype accessions from a landrace to generate multiple $F_1$ seed whose self-pollinated $F_2$ can be combined into a single large $F_2$ population. In an $F_2$ population replications over time or space also cannot be carried out as each single plant represents a different genotype (Semagn *et al.*, 2010), while in the $F_3$ population single plant assessment can be replaced by the assessment of $F_3$ families.

The population of 98 $F_2$ progeny maintained as dried leaf in silica was re-extracted to be used in the improvement of the previous partial genetic map for an inter-subspecific cross between a domesticated landrace and a wild ancestor (Basu *et al.*, 2007c).

## 7.6.2 Marker distortion

With the development of molecular linkage maps, numerous examples of segregation distortion have been reported on many plant species (He *et al.*, 2001). In the current study the segregation patterns of the markers and detection of any distortion was tested by JoinMap4 through performance of a Chi-square test ($p<0.05$). Thirty two percent of the markers (SSR and DArT) were found to be distorted in the narrow cross, based on a RIL3 model. While in the wide cross population 27% of all three marker types (SSR, DArT and AFLP) were distorted, based on an $F_2$ model of segregation. The results of linkage analysis for both maps revealed that markers with distorted segregation were distributed throughout the genome. Our markers for both maps showed lower distortion than the 41% distortion reported by Basu (2005) in the linkage map constructed in bambara groundnut using 115 AFLPs and one SSR marker. However, lower distortion (19.93%) of SSR markers was reported in 106

RILs (F$_9$) derived from the cross of two soybean (*Glycine max*) genotypes (BD2 and BX10) (Liang *et al.*, 2010).

Segregation distortion is influenced by many factors, some of these are: mapping population, genetic transmission, gametic and zygotic selection, non-homologous recombination, gene transfer, transposable elements and environmental agents (Knox and Ellis, 2002; Yamagishi *et al.*, 2010). Wu *et al.* (2001) demonstrated that 21.7% segregation distortion was caused by a genetic transmission effect using an RIL population from a cross between two cultivated soybean Kefeng1 and Nannong1138-2. It was also reported that segregation distortion could be partially caused by gametophytic and sterility factors (Zhang *et al.*, 2006). In constructing a genetic map in barley a total of 22 loci among 65 polymorphic SSR loci (33.8%) were distorted in 260 F$_2$ individuals derived from a male sterile line and an elite cultivar. Distorted loci were found to be skewed toward the male sterile parent (Liu *et al.*, 2011). These results suggested that the phenomenon of segregation distortion occurs commonly in hybrid populations (Konishi *et al.*, 1990). In progeny derived from inter- and intra-specific crosses the distortion is thought to be caused by competition among gametes for preferential fertilization or from abortion of the gamete or zygote (Faris *et al.*, 1998).

### 7.6.3 Linkage map and marker distribution

Map construction was conducted with JoinMap4. Out of 269 markers involved in construction the 'narrow' map 238 marker (29 SSRs and 209 converted DArTs) were assigned in to 21 linkage groups which cover 608.6 cM of the bambara groundnut genome. Map construction was undertaken by running the population under the RIL3

model after the initial analysis of the population as a Cross Pollinator (CP) to determine the phases of DArT markers, by comparison with SSRs.

In the 'wide' cross map three marker types of SSR, DArT and AFLP were involved with a total number of 301 markers. One hundred and ninety four markers were assigned to 20 linkage groups spanning a total of 901 cM.

Some of the linkage groups in both maps had only a few markers and incomplete coverage of the genome led to both maps having more than the expected 11 linkage groups corresponding to chromosomes ($2n=2x=22$) for the bambara genome. Adding more markers to these two maps should join these small groups together, although undetected genetic reasons for fragmented groups may also exist.

The map of 'narrow' cross covered approximately 54% of bambara groundnut genome, however more comprehensive coverage was expected based on the marker-marker linkage (89%). Clustering of the developed markers to particular regions of the genome or inhibition of recombination by parental dissimilarity could be one reason of the lower than expected map length. The linkage map of the $F_2$ population derived from the 'wide' cross had greater genome coverage of 79.6% compared to 'narrow' map, however it had lower marker-marker linkage of 67% than the 'narrow' map. The combination of three marker types could be one of the reasons for greater coverage of the bambara genome by the 'wide' map. Adawy *et al.* (2005) reported that different marker systems differ in detecting polymorphism and their genome coverage.

Unusual clustering of DArT markers was observed in both 'wide' and 'narrow' maps, which seems to confirm earlier observations of two major clusters of accessions, based on of DArT markers including DipC in the smaller group containing the

unusual repeat cluster (Stadler, 2009). However, the source of this clustering is not quite clear an introgression from an alien species might be the reason or it could represent a repetitive sequence which has proliferated locally within the chromosome from a specific progenitor sequence within specific lineages. In the genomes of chick pea, barley and wheat clustering of DArT markers in a particular genome region has also been reported (Akbari *et al.*, 2006; Semagn *et al.*, 2006c; Thudi *et al.*, 2011).

We have developed the first linkage maps based on a combination of SSR and DArT markers in bambara groundnut. This map will be useful for comparative genomic analysis between the mapping populations in this crop and also between bambara groundnut and other related legume crops. This map along with the improved wide cross map using a combination of SSR, DArT and AFLP markers, represents an important step toward genetic analysis of interesting agronomic traits in bambara groundnut. These markers developed could be used in marker-assisted selection for complex agronomic traits and phylogenetic analysis in bambara groundnut.

### 7.6.4 Combined maps

Both 'narrow' and 'wide' maps were initially combined using the JoinMap4 software through 32 common markers (SSRs and DArTs). Thirty common markers were combined in 6 linkage groups with at least 2 common markers each. However, two other groups were linked with only one SSR marker each on the same chromosome, so resolving their relative orientation was not possible. The combined map will assist breeders to localize QTLs to genomic regions for marker-assisted selection. Addition of more common markers will assist in map integration. It will also make comparative mapping with other related legumes easier in the future (Millan *et al.*, 2010), which

ultimately will give valuable information about QTL regions in related crops which may also be important for bambara groundnut.


## 7.7 QTL analysis

The genetic architecture of complex traits in plants can be investigated by QTL mapping which can identify the genomic regions responsible, elucidate the effect of the gene(s) and (eventually) the structure of gene itself (Laurie *et al.*, 2004; Mackay, 2004). QTL analysis was carried out in this study to identify regions of bambara groundnut genome that are responsible of genetic mechanisms for the agronomic traits.

The first step in QTL mapping is to have a linkage map with good coverage of markers. Interval mapping (IM) was used to detect loci accounting for traits data variation in the populations derived from the intra-subspecific and inter-subspecific crosses of bambara groundnut. The IM model is considered more powerful than the analysis of variance at individual marker loci being more flexible for missing genotype data and is more feasible when the markers are widely spaced (Broman, 2001). LOD scores are used to measure the strength of evidence for the presence of a QTL effecting a trait. Where the LOD score is higher than the permutation test threshold, a QTL for that trait can be declared. Generally, the most likely position of the QTL is at the maximum LOD, with a 1 LOD drop in probability on either side of the maximal point defining the confidence interval. The detection of QTL for different generations in different environments for both 'narrow' and 'wide' cross populations are summarised in Figure 7-3.
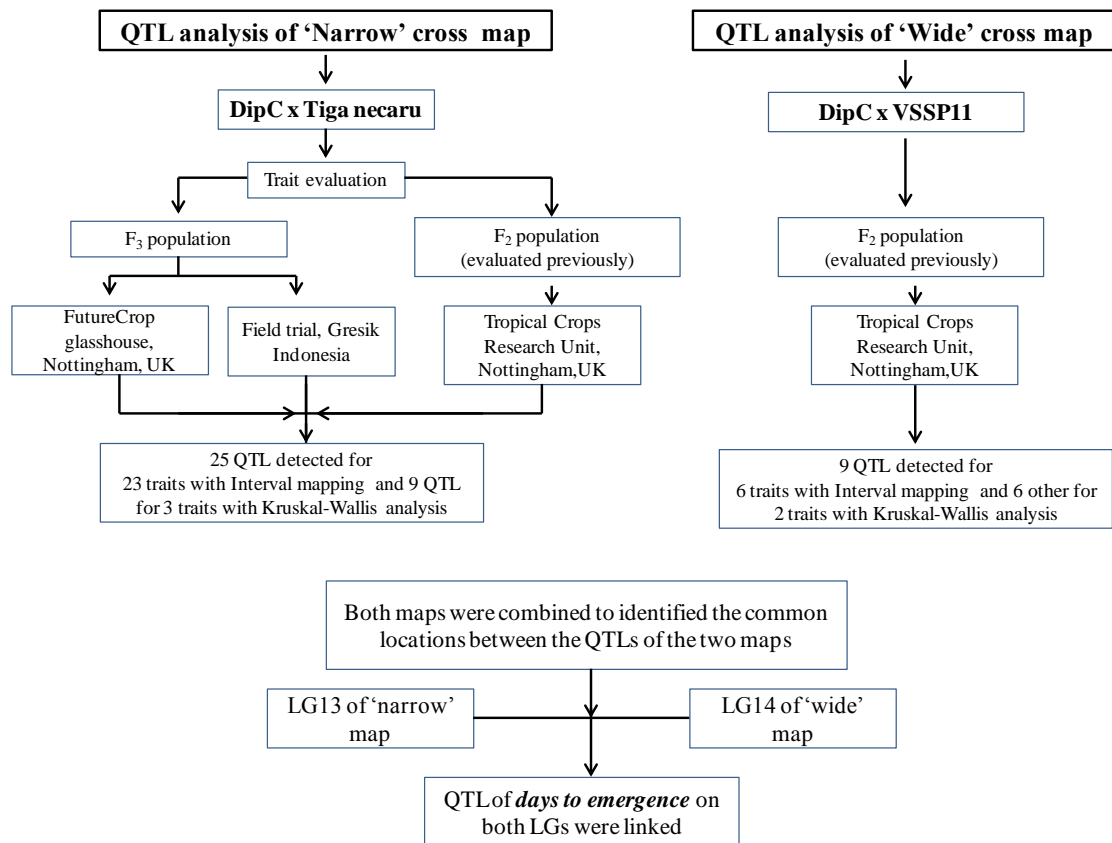
**Figure 7-3: Diagram of the QTL detection scheme for traits in the 'narrow' and 'wide' cross populations, evaluated in different generations and for different environments.**

### 7.7.1 QTL analysis of 'narrow' cross map

We have found several genomic regions detecting QTLs for agronomic traits in both FutureCrop glasshouse and the Indonesian field. These results were also supported by data collected from the same intra-subspecific cross for the $F_2$ generation grown in TCRU.

Eighteen significant QTL and 7 putative QTL were detected for 23 agronomic traits using an interval mapping in addition to nine others detected for *days to emergence*, *growth habit* and *eye pattern around helium* using non-parametric mapping for the $F_3$ population evaluated in FutureCrop glasshouse. Individual QTLs detected with interval mapping explained 13.5 to 50.3% of the total phenotypic variation. Two

QTLs were detected for each of the traits *node no./stem* and *pod length of double seeded*, explaining 39% and 36.2% of phenotypic variance observed for both traits in the $F_3$ progeny evaluated in FutureCrop glasshouse. A smaller number of QTL for these traits were detected and recorded in the Indonesian field trial and the TCRU for the $F_3$ and $F_2$ generations, respectively. The possible reason is that the typical tropical growth condition of this crop in the FutureCrop glasshouse allowed the genes responsible to these traits to be expressed properly more than the field condition, as most of the agronomic traits are quantitative with multi-gene response having minor phenotypic effect and they are sensitive to the environmental changes (Chenu *et al.*, 2011; Choudhary *et al.*, 2008). Furthermore, the technical measurement issue has caused the number of traits involved in QTL analysis to be less in both field and TCRU trials compared to the FutureCrop glasshouse. The QTL for *internode length* and *biomass dry weight* were considered stable across the FutureCrop glasshouse and the Indonesian field trial for the $F_3$ population derived from the cross of DipC x Tiga necaru. These QTLs could be a good candidates for MAS in a yield enhancement program under different environment conditions. Association of the trait values with the same marker in the FutureCrop glasshouse and the field was also recorded for the traits of *days to emergence* and *growth habit* detected by Kruskal-Wallis analysis (due to the non-normal distribution of the trait). Other QTLs for *plant spread*, *leaflet length* and *pod no./plant* were detected on the same linkage groups for the $F_3$ population evaluated in the FutureCrop glasshouse and TCRU.

Potentially pleiotropic effects of single loci on multiple phenotypic traits have been identified in current study; the marker bgPabg-596774 on LG1 was associated with six QTL related to yield. Three other QTL for the pod and seed dimensions were also positioned on LG12 and associated with marker bgPt-598767. The causal relationship

between these traits was confirmed through their strong association in the data analysis of the FutureCrop glasshouse population, as described in Chapter 4, and the allelic effects of coincident QTLs in the same direction as traits correlated supported the hypothesis that traits variation attributed to this coincident QTLs (Appendix 28). Pleiotropy effects have been identified in a wide range of species, including legumes (Bobby *et al.*, 2012; Irzykowska and Wolko, 2004; Prokop, 2004; Wenxin *et al.*, 2008; Zhang *et al.*, 2004).

The QTLs detected for both *peduncle length* and *internode length* were located close to each other within 1cM distance on the same linkage group (LG4). These two traits gave the highest LOD scores in the analysis, with values of 9.7, and 7.9, respectively. Most of the QTLs detected here could be considered valuable and of potential use in MAS of this crop, aiming to develop the yield and the performance of bambara groundnut.

The QTL of traits *Internode length*, *growth habit, plant spread* and *peduncle length* located on LG4 were within the combined confidence interval, indicating that they could be affected by the same gene (s). Using conserved synteny of the marker sequences at this location with the other related legumes might accelerate comparisons of the gene order and identifying candidate gene (s) responsible to these traits. In Cowpea, homologous genes were identified for 85 and 80% of the SNPs markers when soybean and *M. truncatula* genome sequences were compared (Lucas *et al.*, 2011; Muchero *et al.*, 2009).

Both dominance and additive effects were evaluated for the QTL detected with the interval mapping. Most of QTLs detected for the $F_3$ population evaluated in the FutureCrop glasshouse were more likely to be under the effect of additive effect with

the presence of a few dominance effects such as *flower no./plant, plant spread, internode length, seed no./plant* and *biomass dry weight*, while the QTL of *leaf area* appeared to have more dominance effect. Additive effect was also predominant for QTLs of the field trial except for the traits of *node no./plant* and *biomass dry weigh*t showing dominant gene action for their QTLs. While the $F_2$ population evaluated for the same 'narrow' cross the effect of dominance effects for the QTL of different traits seem to be more than those detected for the segregating $F_3$ population (Appendix 27, Appendix 28 and Appendix 29). QTLs with large additive effects are valuable for breeders as they might result in larger difference in the traits of interest (Bradshaw *et al.*, 2008). The coincidence of seed yield QTL with that of yield component such as (biomass dry weight and flower no./plant ) with high additive effect offers a means for selecting for seed yield by efficient selection for its component.

## 7.7.2 QTL analysis in the 'wide' cross map

The eight traits evaluated previously in the $F_2$ population derived from the wide cross (DipC x VSSP11) (Basu, 2005) were included in the QTL analysis of the updated 'wide' map. QTLs detected for *stem no./plant* and *CID* were significant, a putative QTL was detected for other traits of *leaf area*, *specific leaf area* and *100-seed weight*. Two and three QTLs for the traits *stem no./plant* and *internode length* were located on different linkage groups, and the loci for each traits together explained about 34% and 38% of the phenotypic variation, respectively. *Days to emergence* and *days to flowering* were non-normally distributed and they were found in significant association with AFLPs and SSRs marker loci on different linkage groups (7, 10, 14 and 15). Insufficient parental data on *days to emergence* and *days to flowering* and the possibility of transgressive segregation for these traits might hindered the

identification of discrete classes of parental value to analyse the segregation patterns in the $F_2$ for these two traits. The three traits of *days to emergence, CID and 100-seed weight* as found in significant association with each other, QTLs being located at the same position of marker P19b494AGC+CTA4 on LG7.

The QTL identified in any one mapping experiment are likely to describe only a small part of the global genetic architecture of the trait within the species as a whole. For a better understanding of the genetic architecture of natural variation for the trait, the use of multiple populations connected through common markers could help to understand the species level variation for the trait of interest (Symonds *et al.*, 2005).

In the combined linkage groups of 'wide' and 'narrow' maps a common location was determined between the QTLs for *days to emergence*. This trait in both maps was in close association with the common DArT markers 601384 and 601748. The limited QTL linkage between the two maps is likely to be a function of a limited number of integrated linkage groups and also due to some limitation of phenotypic assessment in an $F_2$ population based on single plant assessment which led to fewer QTLs for agronomic traits (Paterson *et al.*, 1991; Rakshit *et al.*, 2012; Semagn *et al.*, 2010). Increasing the number of common marker from other related legumes (soybean and *Medicago*) might help to link more QTLs through combining more linkage groups from both maps or additional marker systems in bambara groundnut (potentially DArT Seq scored in both populations) could significantly improve integration of the maps.

In the QTL analysis with the interval mapping those detected for *leaf area, Carbon isotope discrimination* and *100-seed weight* had dominance effect, while the others either had only additive effect or both additive and dominance effects (Appendix 30).

### 7.7.3 An overview of QTL analysis

To determine the basis for domestication traits and the recruitment of the wild ancestor into a domesticated 'landrace', the 'wide' map was constructed between a wild and cultivated accession (Basu *et al.*, 2007c; Basu, 2005). The subsequent development of domesticated landraces and trait variation within domesticated material was investigated through the 'narrow' map constructed here between domesticated landraces. The latter cross was developed to provide a better understanding of the genetic control of biological mechanisms controlling agronomic traits. The results of the QTL analysis have been presented and several genomic regions containing QTLs for agronomic traits identified, with a number of these stable across different environments. The QTLs detected here are specific to the populations' analyses for the given study; it is quite possible that future studied will discover additional QTL regions contributing to agronomic traits phenotypes because of genetic variance across and within landraces. Although the number of markers integrated to construct the genetic linkage maps, the loci in both framework maps were generally within recommended distance for QTL analysis (<10 cM) (Doerge, 2002). The QTL identified could provide a resource for identifying the regions of the bambara genome which contains genes for agronomically important traits and to develop molecular markers for MAS in the breeding of bambara groundnut.

**7.8 Summary of progress achieved in the study**

This project has investigated various aspects of developing molecular markers, phenotypic evaluation of the populations, construction of genetic maps and QTL analysis. The results of these investigations can be summarised into a number of main points.

1- Development of polymorphic SSR markers for both 'wide' and 'narrow' cross populations, generated from a genomic-enriched repeat library sequenced with non-titanium and titanium reagents (Roche 454 Pyrosequencing), a leaf transcriptome library (Roche 454 Pyrosequencing) and a single cross-species SSR from the soybean genome (Chapter 3).

2- Developing polymorphic DArT marker data for the $F_3$ population, derived from 'narrow' cross of domesticated bambara groundnut (Chapter 3).

3- Confirmation of the residual heterozygosity present in the 'narrow' cross. A set of 33 polymorphic SSR was used and the individuals of $F_3$ population were found to match with the predicted Mendelian ratios expected from the population history (Chapter 3).

4- Rapid extraction of genomic DNA from the seed endosperm without affecting seed viability, using the GenElute Plant Genomic DNA kit (Sigma Aldrich) (Chapter 3).

5- Associations between important agronomic traits in the segregating $F_2$ and $F_3$ populations derived from the 'narrow' cross evaluated in controlled environment glasshouse and Indonesian field experiments (Chapter 4).

6- Construction of a genetic linkage map in bambara groundnut using a total of 238 SSR and DArT polymorphic markers in the $F_3$ population (Chapter 5).

7- Improving the wide cross map using a combination of SSR, DArT and AFLP markers in the $F_2$ population (Chapter 5).

8- Attempting to combine both 'narrow' and 'wide' maps using JoinMap4 with 30 common markers (SSRs and DArTs) (Chapter 5).

9- QTL mapping of polygenic agronomic traits, identify 18 QTLs for the traits evaluated in the population derived from the 'narrow' cross and 6 QTLs for traits of 'wide' cross population evaluated previously (Chapter 6).

10- Co-location of a QTL for *days to emergence* in both 'wide' and 'narrow' (Chapter 6).

**7.9 Future work**

Based on the results obtained from the present study, the following future lines of work can be proposed briefly.

- Already validated candidate marker (bgPabg-596774) could be used for MAS. The QTL region could be fine mapped for use in map-based cloning or subjected to a functional genomics approach. Development of the population into full RILs (it is now at $F_6$) or the development of Near Isogenic Lines (NILs) could allow the evaluation of the effects of this QTL alone and also the development of heterozygous plants for the QTL region, allowing a large-scale fine mapping programme.

- Further validation of prominent candidate markers/QTLs for various traits is required in different populations, across locations and/seasons, before using them in the breeding program of bamabara groundnut.

- Saturate linkage maps with more markers so that the extensive phenotypic data, including drought and heat tolerance and photoperiodic control of pod filling, could be efficiently used for further QTL detection. DArT Seq data was generated on the narrow cross, but time did not allow a full analysis of this for genetic mapping.

- Increasing the number of common markers using COS primers (screened previously with the pooled DNA from bambara groundnut (Basu, 2005) along with other COS primers designed from related legume crops, *Medicago truncatula* Gaertn and soybean (Glycine max) http://www.phytozome.net/soybean (Kaur *et al.*, 2011; Varshney *et al.*, 2009b), to screen the polymorphism in the populations under study. This might help to link more QTLs through combining more linkage groups from both 'wide' and 'narrow' maps and would also allow alignment of

plant genomes between of bambara groundnut and other legume species to predict the gene content and gene order in bambara groundnut through comparative genetic maps.

- Further development of selected lines in future generation to identify superior RILs for various traits which can be used in the future crop improvement program of bambara groundnut before release of new cultivars.

- More sequencing data, especially transcriptomic sequence, can be generated for this crop at relatively low cost using Next Generation Sequencing technology. The sequences developed could be used directly to develop more markers (SSR) markers or to develop synteny of bambara groundnut to close relative legume sequences to analyse gene expression and potentially allow tracking of candidate gene from other relatives to bambara groundnut.

- Developing this segregating population through repeat crossing and development of more lines. This would allow greater QTL analysis power and allow the detection of further QTLs for MAS. It would also allow better localization of effects (due to increased meiotic crossing over). Single seed descent could be used to develop fixed lines for further analysis.

# References

Abate T., Alene A.D., Bergvinson D., Shiferaw B., Silim S., Orr A., Asfaw S. (2012) Tropical Grain Legumes in Africa and South Asia: Knowledge and Opportunities. PO Box 39063, Nairobi, Kenya: International Crops Research Institute for the Semi-Arid Tropics. 112 pp. ISBN: 978-92-9066-544-1. Order Code: BOE 056.

Abberton M.T., Michaelson-Yeates T.P.T., Bowen C., Marshall A.H., Prewer W., Carlile E. (2003) Bulked segregant AFLP analysis to identify markers for the introduction of the rhizomatous habit from Trifolium ambiguum into T. repens (white clover). Euphytica 134:217-222.

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., Kerlavage A.R., Mccombie W.R., Venter J.C. (1991) Complementary-DNA Sequencing - Expressed Sequence Tags and Human Genome Project. Science 252:1651-1656.

Adawy S.S., Hussein E.H.A., Ismail S.E.M.E., El-Itriby H.A. (2005) Genomic diversity in date palm (Phoenix dactylifera L.) as revealed bynAFLP's in comparison to RAPD's and ISSR's. Arab J. of Biotech. 8:99-114.

Adu-Dapaah H.K., Sangwan R.S. (2004) Improving bambara groundnut productivity using gamma irradiation and *in vitro* techniques. African Journal of Biotechnology 3:260-265.

Ajmone Marsan P., Castiglioni P., Ferrarini M., van Wiik R., Motto M., V. K.N. (1999) Polymorphism, distribution, and segregation of EcoRI and PstI based AFLP markers in a molecular linkage map in maize Agro Business Park 90, P.O. Box 216, 6700 AE Wageningen, The Netherlands.

Akbari M., Wenzl P., Caig V., Carling J., Xia L., Yang S., Uszynski G., Mohler V., Lehmensiek A., Kuchel H., Hayden M.J., Howes N., Sharp P., Vaughan P., Rathmell B., Huttner E., Kilian A. (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409-20. DOI: 10.1007/s00122-006-0365-4.

Akkaya M.S., Bhagwat A.A., Cregan P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-9.

Akpalu M.M. (2010) Growth, Yield and Nutritional Quality of Five Bambara Groundnut (*Vigna subterranea* (L) Verdc.) Landraces to Different Plant Population Densities, School of Postgraduate Studies, Kwame Nkrumah University of Science and Technology.

Aldrian E., Djamil Y.S. (2008) Spatio-temporal climatic change of rainfall in east Java Indonesia. International Journal of Climatology 28:435-448. DOI: Doi 10.1002/Joc.1543.

Alexander H.J., Wittenberg V.D.L., T.,, Cayla C., Kilian A., Visser R.G., Schouten H.J. (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Molecular Genetics and Genomics journal 274:30-39. DOI: 10.1007/s00438-005-1145-6.

Amadou H.I., Bebeli P.J., Kaltsikes P.J. (2001) Genetic diversity in Bambara groundnut (*Vigna subterranea* L.) germplasm revealed by RAPD markers. Genome 44:995-9.

Amarteifio J.O., Moholo D. (1998) The chemical composition of four legumes consumed in Botswana. Journal of Food Composition and Analysis 11:329-332. DOI: 10.1006/jfca.1998.0595.

Angeloni F., Wagemaker N., Vergeer P., Ouborg J. (2012) Genomic toolboxes for conservation biologists. Evolutionary Applications 5:130-143. DOI: DOI 10.1111/j.1752-4571.2011.00217.x.

Anonymous. (1947) Jugo beans. In: Review of Crop Experiments; Bechuanaland Protectorate, Department of Agriculture Research, Botswana pp. 127-128.

Aremu M.O., Olaofe O., Akintayo E.T. (2006) Chemical Composition and Physicochemical Characteristics of Two Varieties of Bambara Groundnut (*Vigna subterrenea*) Journal of Applied Sciences 6:1900-1903.

Arends D., Prins P., Jansen R.C., Broman K.W. (2010) R/qtl: high-throughput multiple QTL mapping. Bioinformatics 26:2990-2992. DOI: DOI 10.1093/bioinformatics/btq565.

Arumuganathan K., Earle E. (1991) Nuclear DNA Content of Some Important Plant Species. Plant Molecular Biology Reporter 9:208-218.

Ashkani S., Rafii M.Y., Rusli I., Sariah M., Abdullah S.N.A., Rahim H.A., Latif M.A. (2012) SSRs for Marker-Assisted Selection for Blast Resistance in Rice (*Oryza sativa* L.). Plant Molecular Biology Reporter 30:79-86.

Ashri A. (1994) Genetic resources of sesame: present and future perspectives, in: R. K. Arora and K. W. Riley (Eds.), Sesame Biodiversity in Asia, IPGRI, New Delhi, India.

Atiku A.A. (2000) Bambara groundnut processing and storage practices in North Eastern Nigeria. Postgraduate seminar paper, Department of Agricultural Engineering, University of Maiduguri, Maiduguri, Nigeria.

Avise J.C. (1993) Molecular Markers, Natural History and Evolution. 1st Edn. Chapman and Hall, New York.

Azam-Ali S.N., Sesay A., Karikari S.K., Massawe F.J., Aguilar-Manjarrez J., Bannayan M., Hampson K.J. (2001) Assessing the potential of an underutilized crop: A case study using Bambara groundnut. Experimental Agriculture 37:433-472.

Bailey D.W. (1981) Recombinant inbreds trains and bilineal congenic strains, in: H. J. Fosterg, *et al.* (Eds.), The Mouse in Biomedical Research, Academic Press, New York. pp. 223-239.

Bakker P.I.W., Burtt N.P., Graham R.R., Guiducci C., Yelensky R., Drake J.A., Bersaglieri T., Penney K.L., Butler J., Young S., Onofrio R.C., Lyon H.N., O Stram D., Haiman C.A., Freedman M.L., Zhu X., Cooper R., Groop L., Kolonel L.N., Henderson B.E., Daly M.J., Hirschhorn J.N., Altshuler D. (2006) Transferability of tag SNPs in genetic association studies in multiple populations. Nature Genetics 38:1298-1303.

Bamlink annual report. (2008) Molecular, Environmental and Nutritional Evaluation of Bambara Groundnut (*Vigna subterranea* L.Verdc.) for Food Production in Semi-Arid Africa and India, EU INCO-DEV: SIXTH FRAMEWORK PROGRAMME.

Barkley N.A., Dean R.E., Pittman R.N., Wang M.L., Holbrook C.C., Pederson G.A. (2007) Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing. Genet Res 89:93-106. DOI: S0016672307008695 [pii] 10.1017/S0016672307008695.

Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823-37. DOI: S0092-8674(07)00600-9 [pii] 10.1016/j.cell.2007.05.009.

Basten C., Weir B., Zeng Z.B. (2001) QTL Cartographer. Department of Statistics, North Carolina State University, Raleigh, NC.

Basu S., Roberts J.A., Azam-Ali S.N., Mayes S. (2007a) Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.) - an underutilized African legume crop species. Molecular Ecology Notes 7:1326-1328.

Basu S., Robert J.A., Azam-Ali S.N., Mayes S. (2007b) Bambara groundnut, in: K. C. (Ed.), Genomic mapping and molecular breeding in plants: Pulses,sugar and tuber crops, Springer, New York, pp 159-173.

Basu S., Mayes S., Davey M., Roberts J.A., Azam-Ali S.N., Mithen R., Pasquet R.S. (2007c) Inheritance of 'domestication' traits in bambara groundnut (*Vigna subterranea* (L.) Verdc.). Euphytica 157:59-68.

Basu S.M. (2005) Genetic mapping and trait analysis in bambara groundnut *Vigna subterranea* (L.) Verdc, Biosciences School, The University of Nottingham, Nottingham. pp. 257.

Beckmann J.S., Soller M. (1990) Toward a Unified Approach to Genetic Mapping of Eukaryotes Based on Sequence Tagged Microsatellite Sites. Bio-Technology (New York) 8:930-932.

Begemann F. (1986) Ecogeographic differentiation of bambara groundnut (*Vigna subterranea* ) in the collection of the International Institute of Tropical Agriculture (IITA). Germany, Giessen.

Behn A., Hartl L., Schweizer G., Baumer M. (2005) Molecular mapping of QTLs for non-parasitic leaf spot resistance and comparison of half-sib DH populations

in spring barley. Euphytica 141:291-299. DOI: DOI 10.1007/s10681-005-7697-z.

Benedict C.O., Michael I.U. (2011) Interrelationships among Pod and Seed Yield Traits in Bambara Groundnut (*Vigna subterranea* L. Verdc.) in the Derived Savanna Agro-Ecology of South-Eastern Nigeria under Two Planting Dates. International Journal of plant breeding 5:106-111.

Berchie J.N., Sarkodie-Addo J., Adu-Dapaah H., Agyemang A., Addy S., Asare E., Donkor J. (2010) Yield Evaluation of Three Early Maturing Bambara Groundnut (*Vigna subterranea* L. Verdc.) Landraces at the CSIR-Crops Research Institute, Fumesua-Kumasi, Ghana Journal of Agronomy 9:175-179.

Berchie J.N., Opoku M., Adu-Dapaah H., Agyemang A., Sarkodie-Addo J., Asare E., Addo J., Akuffo H. (2012) Evaluation of five bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces to heat and drought stress at Tono-Navrongo, Upper East Region of Ghana. African Journal of Agricultural Research 7:250-256. DOI: 10.5897/AJAR11.817.

Berkman P.J., Lai K., Lorenc M.T., Edwards D. (2012) Next-Generation Sequencing Applications for Wheat Crop Improvement. American Journal of Botany 99:365-371.

Beyene Y., Botha A.-M., Myburg A.A. (2005) A comparative study of molecular and morphological methods of describing genetic relationships in traditional Ethiopian highland maize. African Journal of Biotechnology 4:586-595.

Bhag M. (2007) Neglected and underutilized crop genetic resources for sustainable agriculture. Indian Journal of Plant Genetic Resources 20:1-14.

Bharucha Z., Pretty J. (2010) The roles and values of wild foods in agricultural systems. Philosophical Transactions of the Royal Society B-Biological Sciences 365:2913-2926. DOI: DOI 10.1098/rstb.2010.0123.

Bobby R., Bazzelle R., Clark W., Kantartzi S., Meksem K., Akond M., Kassen M. (2012) Genetic Analysis of Yield Components in the PI 438489B by 'Hamilton' Recombinant Inbred Line (RIL) Population of Soybean [*Glycine max* (L.) Merr.]. Journal of Agricultural Science 4.

Bolboaca S.D., Jantschi L. (2006) Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationship of biologic active compounds. Leornado Journal of Sciences 9:179-200.

Bolek Y., El-Zik K.M., Pepper A.E., Bell A.A., Magill C.W., Thaxton P.M., Reddy O.U.K. (2005) Mapping of verticillium wilt resistance genes in cotton. plant Science 168:1581-1590.

Bonetta L. (2006) Genome sequencing in the fast lane. Nature Methods 3:141-147.

Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C., Taberlet P. (2004) How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13:3261-3273.

Botstein D., White R.L., Skolnick M., Davis R.W. (1980) Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. American Journal of Human Genetics 32:314-331.

Bouck A., Vision T. (2007) The molecular ecologist's guide to expressed sequence tags. Molecular Ecology 16:907-924. DOI: DOI 10.1111/j.1365-294X.2006.03195.x.

Bowen S., Wheals A.E. (2006) Ser/Thr-rich domains are associated with genetic variation and morphogenesis in Saccharomyces cerevisiae. Yeast 23:633-40. DOI: 10.1002/yea.1381.

Box G.E.P., Cox D.R. (1964) An Analysis of Transformations. Journal of the Royal Statistical Society:211-243.

Bradshaw J.E., Hackett C.A., Pande B., Waugh R., Bryan G.J. (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theor Appl Genet 116:193-211. DOI: 10.1007/s00122-007-0659-1.

Brink M., Belay G. (2006) Plant Resources of Tropical Africa 1. Cereals and Pulses. PROTA Foundation, Wageningen, Netherlands.

Broman K.W. (2001) Review of statistical methods for QTL mapping in experimental crosses. Lab Animal 30:44-52.

Brookfield H. (2002) Agrodiversity and Agrobiodiversity, in: H. Brookfield, *et al.* (Eds.), *Cultivating Biodiversity*, ITDG Publishing, London.

Brough S.H., Azam-Ali S.N. (1992) The effect of soil moisture on the proximate composition of bambara groundnut (*Vigna subterranea* (L.) Verdc.). Journal of the Science of Food and Agriculture 60:197-203.

Brough S.H., Azam-Ali S.N., Taylor A.J. (1993) The potentials of bambara groundnut (*Vigna subterranea n*) in vegetable milk production and basic protein functionality systems. Food Chemistry 47:277-283.

Caporale L.H. (2003) Natural selection and the emergence of a mutation phenotype: An update of the evolutionary synthesis considering mechanisms that affect genome variation, Annual Review of Microbiology. Volume 57, Annual Reviews, 4139 El Camino Way, P. O. Box 10139, Palo Alto, CA, 94303-0139, USA. pp. 467-485.

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., Waugh R. (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847-54.

Cattan-Toupance I., Michalakis Y., Neema C. (1998) Genetic structure of wild bean populations in their South-Andean centre of origin. Theoretical and applied genetics 96:844-851.

Chakraborty S., Zeng Z.-B. (2011) QTL Mapping for Days to Flowering under Drought Condition in Rice (*Oryza sativa* L.) Genome. Not Bot Hort Agrobot Cluj 39:58-63.

Chambers G.K., MacAvoy E.S. (2000) Microsatellites: Consensus and controversy. Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology 126B:455-476.

Chandra A. (2011) Use of EST database markers from M. truncatula in the transferability to other forage legumes. Journal of Environmental Biology 32:347-354.

Chao S., Zhang W., Akhunov E., Sherman J., Ma Y., Luo M.-C., Dubcovsky J. (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Molecular Breeding 23:23-33.

Chauhan A.K., Varma A. (2009) A Textbook of Molecular Biotechnology, I.K. International Publishing House Pvt. Ltd., New Delhi, India, 1336 p.

Chenu K., Cooper M., Hammer G.L., Mathews K.L., Dreccer M.F., Chapman S.C. (2011) Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. Journal of Experimental Botany 62:1743-1755.

Chijioke O.B., Ifeanyi U.M., Blessing A.C. (2010) Comparative study on growth and development of some accessions of local Germplasm of Bambara groundnut (*Vigna subterrenea* L. Verdc.) of Nigeria in two cropping seasons Journal of Crop Science and Biotechnology 13:21-27.

Chittaranjan K. (2007) Genome Mapping and Molecular Breeding in Plants. Springer-Verlag, Berlin Heidelberg, New York.

Choi H.K., Kim D.J., Zhu H., Mun J.H., Back J.M., Roe B., Ellis M., Young M.D., Doyle J., Kiss G., Cook D.R. (2004a) Conserved gene order between crop and model legume species. Proc. Natl. Acad. Sci. USA, 101, str. 15289-15294

Choi H.K., Mun J.H., Kim D.J., Zhu H.Y., Baek J.M., Mudge J., Roe B., Ellis N., Doyle J., Kiss G.B., Young N.D., Cook D.R. (2004b) Estimating genome conservation between crop and model legume species. Proceedings of the National Academy of Sciences of the United States of America 101:15289-15294. DOI: DOI 10.1073/pnas/0402251101.

Cholin S.S. (2009) Construction of Genetic Linkagemap and QTL Analysis for Foliar Disease Resistance, Nutritional Quality and Productivity Traits in Groundnut (*Arachis hypogaea* L) Department of Genetics and plant breeding, University of Agricultural Sciences, Dharwad, Karnataka State, India pp. 187.

Choudhary K., Choudhary O.P., Shekhawat N.S. (2008) Marker Assisted Selection: A Novel Approach for Crop Improvement. American-Eurasian Journal of Agronomy 1:26-30.

Chowdhury D.M.S., Rathjen J.M., Tate M.E., McDonald G. (2004) Genetics of colour traits in common vetch (Vicia sativa L.). Euphytica 136:249-255.

Churchill G.A., Doerge R.W. (1994) Empirical Threshold Values for Quantitative Trait Mapping. Genetics 138:963-971.

Collard B.C., Mackill D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557-72. DOI: 8731828765572240 [pii] 10.1098/rstb.2007.2170.

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

Collinson S.T., Clawson E.J., AzamAli S.N., Black C.R. (1997) Effects of soil moisture deficits on the water relations of bambara groundnut (*Vigna subterranea* L Verdc.). Journal of Experimental Botany 48:877-884.

Condit R., Hubbell S.P. (1991) Abundance and DNA Sequence of Two-Base Repeat Regions in Tropical Tree Genomes. Genome 34:66-71.

Conforti P. (2011) Looking Ahead in World Food and Agriculture: Perspectives to 2050, Office of Knowledge Exchange, Research and Extension, FAO, Viale delle Terme di Caracalla, 00153 Rome, Food and Agriculture Organization of the United Nations, Italy. pp. 560.

Cooper M., Podlich D.W., Smith O.S. (2005) Gene-to-phenotype models and complex trait genetics. Australian Journal of Agricultural Research 56:895-918.

Cordeiro G.M., Casu R., McIntyre C.L., Manners J.M., Henry R.J. (2001) Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum. Plant Science (Shannon) 160:1115-1123.

Cornelissen R., Deswarte J.C., Azam-Ali S.N. (2002) A preliminary strategy for model development in bambara groundnut, in: A. Sesay (Ed.), Workshop on Increasing the productivity of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) for sustainable food production in semi-arid Africa, University of Swaziland, Kwaluseni, Swaziland. 28-30 August 2001. pp. 167-176.

Cotton R.G.H. (1997) Slowly but surely towards better scanning for mutations (vol 13, pg 43, 1997). Trends in Genetics 13:208-208.

Cuc L.M., Mace E.S., Crouch J.H., Quang V.D., Long T.D., Varshney R.K. (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (Arachis hypogaea). BMC Plant Biology 8:Article No.: 55.

Dakin E.E., Avise J.C. (2004) Microsatellite null alleles in parentage analysis. Heredity (Edinb) 93:504-9. DOI: 10.1038/sj.hdy.68005456800545 [pii].

Dakora F.D. (1998) Nodule function in symbiotic bambara groundnut (*Vigna subterranea* L.) and Kersting's bean (Macrotyloma geocarpum L.) is tolerant of nitrate in the root medium. Annals of Botany 82:687-690.

Damayanti F., Lawn R.J., Bielig L.M. (2010) Expression of qualitative and quantitative traits in hybrids between domesticated and wild accessions of the

tropical tuberous legume *Vigna vexillata* (L.) A. Rich. Crop & Pasture Science 61:798-811. DOI: Doi 10.1071/Cp10084.

Dani M.A., Heinneman M.B., Dani S.U. (2008) Brazilian Nelore cattle: a melting pot unfolded by molecular genetics. Genet Mol Res 7:1127-37. DOI: gmr499 [pii].

Dansi A., Vodouhe R., Azokpota P., Yedomonhan H., Assogba P., Adjatin A., Loko Y.L., Dossou-Aminon I., Akpagana K. (2012) Diversity of the neglected and underutilized crop species of importance in Benin. ScientificWorldJournal 2012:932947. DOI: 10.1100/2012/932947.

Darvasi A., Soller M. (1995) Advanced Intercross Lines, an Experimental Population for Fine Genetic-Mapping. Genetics 141:1199-1207.

Davies P.D.O. (2003) The world-wide increase in tuberculosis: How demographic changes, HIV infection and increasing numbers in poverty are increasing tuberculosis. Annals of Medicine 35:235-243.

Dellaporta S.L., Wood J., B. H.J. (1983) A plant DNA minipreparation version II. Plant Molecular Biology Reporter 1:19-21.

Devaux P., Pickering R. (2005) Haploids in the improvement of Poaceae, in: C. E. Palmer, *et al.* (Eds.), Biotechnology in Agriculture and Forestry, Haploids in Crop Improvement II, Springer-Verlag, Berlin Heidelberg Publ.

DeWoody J., Nason J.D., Hipkins V.D. (2006) Mitigating scoring errors in microsatellite data from wild populations. Molecular Ecology Notes 6:951-957.

Dimakatso R.M. (2006) Evaluation of bambara groundnut (*Vigna subterranea* ) for yield stability and yield related characteristics. Agricultural Research Council - Grain Crop Institute (ARC-GCI).

Directorate Plant Production. (2009) Production guideline for Bambara groundnuts. www.nda.agric.za/publications. (DPP) South Africa.

Directorate plant production. (2011) Production guideline for Bambara groundnuts, Directorate Agricultural Information Services, Department of Agriculture, Forestry and Fisheries Private Bag X144,, Pretoria, 0001 South Africa, from www.daff.gov.za.

Dirlewanger E., Cosson P., Tavaud M., Aranzana J., Poizat C., Zanetto A., Arús P., Laigret F. (2002) Development of microsatellite markers in peach [ *Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry ( *Prunus avium* L.). Theor Appl Genet. 105:127-138.

Doerge R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43-52. DOI: 10.1038/nrg703 nrg703 [pii].

Doerge R.W., Churchill G.A. (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285-294.

Doku E.V. (1968) Flowering, pollination and pod formation in Bambara Groundnut (*Voandzeia subterranea*) in Chana. Experimental Agriculture 4:41-48.

Doku E.V., Karikari S.K. (1970) Flowering and pod production of bambara groundnut. Ghana Journal of Agricultural Science:17-26.

Doku E.V., Karikari S.K. (1971) The role of ants in pollination and pod production of bambara groundnut. Economic Botany 25:357-362.

Doyle J.J., Luckow M.A. (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol 131:900-10. DOI: 10.1104/pp.102.018150.

Dutta S., Kumawat G., Singh B.P., Gupta D.K., Singh S., Dogra V., Gaikwad K., Sharma T.R., Raje R.S., Bandhopadhya T.K., Datta S., Singh M.N., Bashasab F., Kulwal P., Wanjari K.B., R K.V., Cook D.R., Singh N.K. (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]. BMC Plant Biol 11:17. DOI: 1471-2229-11-17 [pii] 10.1186/1471-2229-11-17.

Easterling W.E., Aggarwal P.K., Batima P., Brander K.M., Erda L., Howden S.M., Kirilenko A., Morton J., Soussana J.F., Schmidhuber J., Tubiello F.N. (2007) Food, fibre and forest products. Contribution of Working Group II to the Fourth Assessment Report of IPCC, in: M. L. Parry, *et al.* (Eds.), Climate change 2007: Impacts, adaptation and vulnerability, Cambridge University Press., Cambridge, UK, . pp. 273-313.

Edwards K.J., Barker J.H.A., Daly A., Jones C., Karp A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758-60.

El-Assal S.E.D., Alonso-Blanco C., Peeters A.J.M., Raz V., Koornneef M. (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. Nature Genetics 29:435-440.

Ellah M.M., Singh A. (2008) Bambara Groundnut (*Vigna subterranea* L. Verdc.) Yield as Influenced by Phosphorus and Cultivars in the Semi-Arid Savanna of Nigeria. Journal of Plant Sciences 3:176-181.

Ellegren H. (2004) Microsatellites: Simple sequences with complex evolution. Nature Reviews Genetics 5:435-445.

Eltayeb A.R.S.M., Ali A.O., Abou-Arab A.A., Abu-Salem F.M. (2011) Chemical composition and functional properties of flour and protein isolate extracted from Bambara groundnut (*Vigna subterranea n*). African Journal of Food Science 5:82-90.

Erickson D.L., Fenster C.B., Stenoien H.K., Price D. (2004) Quantitative trait locus analyses and the study of evolutionary process. Mol Ecol 13:2505-22. DOI: 10.1111/j.1365-294X.2004.02254.x MEC2254 [pii].

Erskine W. (2003) Prospects of genetic improvement of grain legumes for drought resistance, in: N. P. Saxena (Ed.), Management of agricultural drought: Agronomic and genetic options, Science Publishers, Inc., Enfield. pp. 81-88.

Eujayl I., Sorrells M.E., Baum M., Wolters P., Powell W. (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399-407. DOI: 10.1007/s001220100738.

Eujayl I., Sledge M.K., Wang L., May G.D., Chekhovskiy K., Zwonitzer J.C., Mian M.A. (2004) Medicago truncatula EST-SSRs reveal cross-species genetic markers for Medicago spp. Theor Appl Genet 108:414-22. DOI: 10.1007/s00122-003-1450-6.

Eyzaguirre P.B., Padulosi S., Hodgkin T. (1999) IPGRI's strategy for neglected and underutilized species and the human dimension of agrobiodiversity, in: S. Padulosi (Ed.), Priority-setting for underutilized and neglected plant species of the Mediterranean region, ICARDA, Aleppo, Syria pp. 1-20.

Fairbanks D.J., Andersen W.R. (1995) Molecular Analysis of Genetic Diversity: Advantages and Limitations, in: B. A. Roundy, *et al.* (Eds.), Proceedings: wildland shrub and arid land restoration symposium; 1993 October 19-21; Las Vegas, NV. Gen. Tech. Rep. INT-GTR-315. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 384 p.

FAO. (2011a) The State of Food Insecurity in the World 2011: How does international price volatility affect domestic economies and food security?, Research and Extension, FAO, Viale delle Terme di Caracalla, 00153 Rome, Italy., Rome.

FAO. (2011b) Biodicersity for a world without hunger, Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization, Rome, Italy.

Faris J.D., Laddomada B., Gill B.S. (1998) Molecular mapping of segregation distortion loci in Aegilops tauschii. Genetics 149:319-327.

Farooq S., Azam F. (2002) Molecular Markers in Plant Breeding-I: Concepts and Characterization. Pakistan Journal of Biological Sciences 5:1135-1140.

Faville M.J., Vecchies A.C., Schreiber M., Drayton M.C., Hughes L.J., Jones E.S., Guthridge K.M., Smith K.F., Sawbridge T., Spangenberg G.C., Bryan G.T., Forster J.W. (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (Lolium perenne L.). Theor Appl Genet 110:12-32. DOI: 10.1007/s00122-004-1785-7.

Ferreira A., da Silva M.F., Silva L.D.C.E., Cruz C.D. (2006) Estimating the effects of population size and type on the accuracy of genetic maps. Genetics and Molecular Biology 29:187-192.

Firouzian A. (2003) Heritability and genetic advance of grain yield and its related traits in spring wheat. Pak. J. Biological Sciences 6:2020-2032.

Foley J.A., Ramankutty N., Brauman K.A., Cassidy E.S., Gerber J.S., Johnston M., Mueller N.D., O'Connell C., Ray D.K., West P.C., Balzer C., Bennett E.M., Carpenter S.R., Hill J., Monfreda C., Polasky S., Rockstrom J., Sheehan J., Siebert S., Tilman D., Zaks D.P. (2011) Solutions for a cultivated planet. Nature 478:337-42. DOI: nature10452 [pii] 10.1038/nature10452.

Foolad M.R., Jones R.A., Rodriguez R.L. (1993) RAPD markers for constructing intraspecific tomato geneticmaps. Plant Cell Reports 12:293-297.

Forster B.P., Heberle-Bors E., Kasha K.J., Touraev A. (2007) The resurgence of haploids in higher plants. Trends in Plant Science 12:368-375. DOI: DOI 10.1016/j.tplants.2007.06.007.

Francia E., Tacconi G., Crosatti C., Barabaschi D., Bulgarelli D., Dall'Aglio E., Vale G. (2005) Marker assisted selection in crop plants. Plant Cell Tissue and Organ Culture 82:317-342. DOI: DOI 10.1007/s11240-005-2387-z.

Frary A., Nesbitt T.C., Grandillo S., Knaap E., Cong B., Liu J., Meller J., Elber R., Alpert K.B., Tanksley S.D. (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85-8. DOI: 8651 [pii].

Freyre R., Skroch P.W., Geffroy V., Adam-Blondon A.F., Shirmohamadali A., Johnson W.C., Llaca V., Nodari R.O., Pereira P.A., Tsai S.M., Tohme J., Dron M., Nienhuis J., Vallejos C.E., Gepts P. (1998) Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. Theoretical and applied genetics 97:847-856.

Gaitán-Solís E., Duque M.C., Edwards K.J., Tohme J. (2002) Microsatellite Repeats in Common Bean (Phaseolus vulgaris): Isolation, Characterization, and Cross-Species Amplification in Phaseolus ssp. Crop Science 42:2128-2136.

Garcia A.A.F., Benchimol L.L., Barbosa A.M.M., Geraldi I.O., Souza C.L., de Souza A.P. (2004) Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genetics and Molecular Biology 27:579-588.

Garg R., Patel R.K., Tyagi A.K., Jain M. (2011) De Novo Assembly of Chickpea Transcriptome Using Short Reads for Gene Discovery and Marker Identification. DNA Research 18:53-63. DOI: DOI 10.1093/dnares/dsq028.

Gepts P., Papa R. (2002) Evolution during Domestication:During the process of domestication, plants and animals have evolved under conscious and unconscious selection by humans. ENCYCLOPEDIA OF LIFE SCIENCES. Macmillan Publishers Ltd, Nature Publishing Group / www.els.net. DOI: 10.1038/npg.els.0003071.

Gholizadeh M., Mianji G.R. (2007) Use of microsatellite markers in poultry research. International Journal of Poultry Science 6:145-153.

Goli A.E. (1997) Bibliographical review of Bambara groundnut, IPGRI, Cotonou.

Gondo T., Sato S., Okumura K., Tabata S., Akashi R., Isobe S. (2007) Quantitative trait locus analysis of multiple agronomic traits in the model legume Lotus japonicus. Genome 50:627-637. DOI: Doi 10.1139/G07-040.

Gonzalez-Chavira M.A., Torres-Pacheo I., Villordo-Pineda A., G.R. G.-G. (2006) DNA markers, Advances in Agricultural and Food Biotechnology, 6: 99 - 134.

Grist S.A., Firgaira F.A., Morley A.A. (1993) Dinucleotide repeat microsatellite loci from *Elymus alaskanus* and length polymorpolymorphisms isolated by the polymerase chain reaction. Bio phism in several *Elymus* species (*Triticeae: Poaceae*). Genome 41:304-309.

Grodzicker T., Williams J., Sharp P., Sambrook J. (1975) Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harb Symp Quant Biol 39 Pt 1:439-46.

Gujaria N., Kumar A., Dauthal P., Dubey A., Hiremath P., Prakash A.B., Farmer A., Bhide M., Shah T., Gaur P.M., Upadhyaya H.D., Bhatia S., Cook D.R., May G.D., Varshney R.K. (2011) Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (Cicer arietinum L.). Theoretical and applied genetics 122:1577-1589. DOI: DOI 10.1007/s00122-011-1556-1.

Gupta P.K., Roy J.K., 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 (Bangalore) 80:524-535.

Gupta P.K., Varshney R.K., C. S.P., Ramesh B. (1999) Molecular markers and their applications in wheat breeding. Plant Breeding 118:369-390.

Gupta S.C., Saxena K.B., Sharma D. (1981) Inheritance of days to flower and of seed size in pigeonpea. In : Proc. International Workshop on Pigeonpea, ICRISAT, India:61-66.

Gur-Arie R., Cohen C.J., Eitan Y., Shelef L., Hallerman E.M., Kashi Y. (2000) Simple sequence repeats in Escherichia coli: Abundance, distribution, composition, and polymorphism. Genome Research 10:62-71.

Hackauf B., Wehling P. (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breeding 121:17-25.

Haley C.S., Anderson L. (1997) Linkage mapping of quantitative trait loci in plants and animals., in: P. H. Dear (Ed.), Genome mapping: a practical approach, Oxford University Press, New York.

Hammock E.A.D., Young L.J. (2005) Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308:1630-1634. DOI: DOI 10.1126/science.1111427.

Han Z., Wang C., Song X., Guo W., Gou J., Li C., Chen X., Zhang T. (2006) Characteristics, development and mapping of Gossypium hirsutum derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112:430-9. DOI: 10.1007/s00122-005-0142-9.

Hancock J.M., Simon M. (2005) Simple sequence repeats in proteins and their significance for network evolution. Gene 345:113-118. DOI: DOI 10.1016/j.gene.2004.11.023.

Hanelt P. (2001) Leguminosae subfamily (Fabaceae). In Mansfield's Encyclopaedia of Agricultural and Horticultural Crops, 2: 635 -957 (Eds. Hanelt P.), Springer, Berlin, Germany.

Harris D., Azam-Ali S.N. (1993) Implications of daylength sensitivity in bambara groundnut (*Vigna subterranea* ) for production in Botswana. Journal of Agricultural Science 120:75-78.

Hartl D., Jones E. (2001) Genetics: Analysis of Genes and Genomes, Jones and Bartlett Publishers, Sudbury, MA.

Hasan M.N., Islam A.K.M.A., Mian M.A.K., Hossain T. (2006) Inheritance of Yield Related Traits in French Bean (Phaseolus vulgaris L.). Bangladesh Journal of Life Sciences 18:125-130.

He G., Woullard F.E., Marong I., Guo B.Z. (2006) Transferability of Soybean SSR Markers in Peanut (Arachis hypogaea L.). Peanut Science 33:22-28.

He P., Li J.Z., Zheng X.W., Shen L.S., Lu C.F., Chen Y., Zhu L.H. (2001) Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. Crop Science 41:1240-1246.

Hearnden P.R., Eckermann P.J., McMichael G.L., Hayden M.J., Eglinton J.K., Chalmers K.J. (2007) A genetic map of 1,000 SSR and DArT markers in a wide barley cross. Theor Appl Genet 115:383-91. DOI: 10.1007/s00122-007-0572-7.

Helentjaris T., Slocum M., Wright S., Schaefer A., Nienhuis J. (1986) Construction of Genetic Linkage Maps in Maize Zea-Mays and Tomato Lycopersicon-Spp Using Restriction Fragment Length Polymorphisms. Theoretical and applied genetics 72:761-769.

Heller J., Begemann F., Mushonga J. (1997) Bambara groundnut. *Vigna subterranea* (L.) Verdc. *In* Promoting the conservation and use of underutilised and neglected crops. 9. Proceedings of the Workshop on Conservation and Improvement of Bambara Groundnut (*Vigna subterranea* (L.) Verdc.), 14-16 November 1995, Harare, Zimbabwe. International plant genetic resources institute (IPGRI), Rome, Italy.

Heller J., Begemann F., Mushonga J. ( 1995) Bambara groundnut. *Vigna subterranea* (L.) Verdc. Promoting the conservation and use of underutilised and neglected crops. 9. Proceedings of the Workshop on Conservation and Improvement of Bambara Groundnut (*Vigna subterranea* (L.) Verdc.), 14-16 November 1995, Harare, Zimbabwe. International plant genetic resources institute (IPGRI), Rome, Italy.

Hittalmani S., Shashidhar H.E., Bagali P.G., Huang N., Sidhu J.S., Singh V.P., Khush G.S. (2002) Molecular mapping of quantitative trait loci for plant growth,

yield and yield related traits across three diverse locations in a doubled haploid rice population. Euphytica 125:207-214.

Holmens C.J. (1998) Socio-economic background: The world food problem, in: W. C.C. and W. P.N. (Eds.), *Agriculture in the tropics*, Blackwell Science Ltd.

Howell J.A., Eshbaugh W.H., Guttman S., Rabakonandrianina E. (1994) Common names given to bambara groundnut (*Vigna subterranea* : Fabacea) in central madagasqar. Economic Botany 48:217-221.

Huang X.Q., Paulo M.J., Boer M., Effgen S., Keizer P., Koornneef M., van Eeuwijk F.A. (2011) Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. Proceedings of the National Academy of Sciences of the United States of America 108:4488-4493. DOI: DOI 10.1073/pnas.1100465108.

Hulbert S.H., Ilott T.W., Legg E.J., Lincoln S.E., Lander E.S., Michelmore R.W. (1988) Genetic-Analysis of the Fungus, Bremia-Lactucae, Using Restriction Fragment Length Polymorphisms. Genetics 120:947-958.

Huynh B.L. (2009) Genetic characterization and QTL mapping for grain fructan in wheat (*Triticum aestivum* L.). School of Agriculture, Food and Wine : Plant and Food Science, University of Adilade. pp. 124.

ICUC. (2006) Annual Report 2005-2006. Colombo, Sri Lanka: International Centre for Underutilised Crops. ISSN 1800-2315.

Ijarotimi O.S., Esho T.R. (2009) Comparison of nutritional composition and anti-nutrient status of fermented, germinated and roasted bambara groundnut seeds (*Vigna subterranea* ). British Food Journal 111:376-386. DOI: Doi 10.1108/000700910951515.

Ikram S., Chardon F. (2010) Plant Quantitative Traits. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

Inoue T., Orgel L.E. (1983) A Nonenzymatic Rna Polymerase Model. Science (Washington D C) 219:859-861.

International Atomic Energy Agency. (2004) Genetic improvement of under-utilized and neglected crops in low income food deficit countries through irradiation and related techniques : proceedings of a final Research Coordination Meeting organized by the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture and held in Pretoria, South Africa, 19-23 May 2003. International Atomic Energy Agency, Vienna.

Iorizzo M., Senalik D.A., Grzebelus D., Bowman M., Cavagnaro P.F., Matvienko M., Ashrafi H., Van Deynze A., Simon P.W. (2011) De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. BMC Genomics 12. DOI: Artn 389 Doi 10.1186/1471-2164-12-389.

IPGRI. (1998) IPGRI's strategy for Neglected and Under-utilized Species and the Human Dimension of Agrobiodiversity, in: P. B. Eyzaguirre, *et al.* (Eds.),

Priority setting for Under-utilized and Neglected Plant Species of the Mediterranean Region, IPGRI-Aleppo, Syria.

IPGRI. (2000) Bambara groundnut (*Vigna subterranea* ). Descriptors for Bambara groundnut (*Vigna subterranea* ), IPGRI. Source / contributor: BAMNET (The International Bambara Groundnut Network), IITA (International Institute of Tropical Agriculture), IPGRI (International Plant Genetic Resources Institute). ISBN: 92-9043-461-9.:48.

IPGRI. (2002) Neglected and Underutilized Plant Species: Strategic Action Plan of the International Plant Genetic Resources Institute., International Plant Genetic Resources Institute, Rome, Italy.

Iqbal A., Khalil I.A., Ateeq N., Khan M.S. (2006) Nutritional quality of important food legumes. Food Chemistry 97:331-335.

Irzykowska L., Wolko B. (2004) Interval mapping of QTLs controlling yield-related traits and seed protein content in Pisum sativum. Journal of Applied Genetics 45:297-306. DOI: 211 [pii].

Isemura T., Kaga A., Konishi S., Ando T., Tomooka N., Han O.K., Vaughan D.A. (2007) Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. Annals of Botany 100:1053-1071. DOI: Doi 10.1093/Aob/Mcm155.

Islam A.K.M.A., Newaz M.A. (2000) Gene Action and Seed Yield and Yield Related Characters in Dry Bean (Phaseolus vulgaris L). Bangladesh Journal of Plant Breeding and Genetics 13:31-34.

Jaccoud D., Peng K., Feinstein D., Kilian A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:E25.

Jaenicke H., Ganry J., Hoeschle-Zeledon I., Kahane R. (2009) Proceedings of the International Symposium on Underutilized Plants for Food, Nutrition, Income and Sustainable Development Acta Horticulture. ISHS, Leuven, Belgium, Arusha, Tanzania,.

Jansen R.C., Stam P. (1994) High-Resolution of Quantitative Traits into Multiple Loci Via Interval Mapping. Genetics 136:1447-1455.

Jarne P., Lagoda P.J.L. (1996) Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11:424-429.

Jenczewski E., Gherardi M., Bonnin I., Prosperi J.M., Olivieri I., Huguet T. (1997) Insight on segregation distortions in two intraspecific crosses between annual species of Medicago (Leguminosae). Theoretical and applied genetics 94:682-691. DOI: 10.1007/s001220050466.

Jing H.C., Bayon C., Kanyuka K., Berry S., Wenzl P., Huttner E., Kilian A., Hammond-Kosack K.E. (2009) DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum*

*monococcum*. BMC Genomics 10. DOI: Artn 458 Doi 10.1186/1471-2164-10-458.

Johnson D.S., Mortazavi A., Myers R.M., Wold B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497-502. DOI: 1141319 [pii] 10.1126/science.1141319.

Johnson P.C., Haydon D.T. (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. Genetics 175:827-42. DOI: genetics.106.064618 [pii] 10.1534/genetics.106.064618.

Jonah P.M., Aliyu B., Adeniji T.O., Bello D. (2012) Seasonal Variation and Pearson Correlation in Yield and Yield Components in Bambara Groundnut. World Journal of Agricultural Sciences 8:26-32.

Jones N., Ougham H., Thomas H. (1997) Markers and mapping:We are all geneticists now. New Phytologist 137:165-177.

Jorgensen S.T., Liu F., Ouedraogo M., Ntundu W.H., Sarrazin J., Christiansen J.L. (2010) Drought Responses of Two Bambara Groundnut (*Vigna subterranea* L. Verdc.) Landraces Collected from a Dry and a Humid Area of Africa. Journal of Agronomy and Crop Science 196:412-422. DOI: DOI 10.1111/j.1439-037X.2010.00435.x.

Jorgensen S.T., Aubanton M., Harmonic C., Dieryck C., Jacobsen S.P., Simonsen H.L., Ntundu W.H., Stadler F., Basu S., Christiansen J.L. (2009) Identification of photoperiod neutral lines of bambara groundnut (*Vigna subterranea* ) from Tanzania. IOP Conference Series: Earth and Environmental Science 6:3720-23.

Joshi S.P., Ranjekar P.K., Gupta V.S. (1999) Molecular markers in plant genome analysis. Current Science (Bangalore) 77:230-240.

Joshi S.P., Gupta V.S., Aggarwal R.K., Ranjekar P.K., Brar D.S. (2000) Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theoretical and applied genetics 100:1311-1320. DOI: DOI 10.1007/s001220051440.

Jun T.H., Michel A.P., Mian M.A.R. (2011) Development of soybean aphid genomic SSR markers using next generation sequencing. Genome 54:360-367. DOI: Doi 10.1139/G11-002.

Kajonphol T., Sangsiri C., Somta P., Toojinda T., Srinives P. (2012) SSR Map Construction and Quantitative Trait Loci (QTL) Identification of Major Agronomic Traits in Mungbean (*Vigna radiata* (L.) Wilczek). SABRAO Journal of Breeding and Genetics 44:71-86.

Kalia R.K., Rai M.K., Kalia S., Singh R., Dhawan A.K. (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309-334.

Kalo P., Seres A., Taylor S.A., Jakab J., Kevei Z., Kereszt A., Endre G., Ellis T.H.N., Kiss G.B. (2004) Comparative mapping between Medicago sativa and Pisum

sativum. Molecular Genetics and Genomics 272:235-246. DOI: DOI 10.1007/s00438-004-1055-z.

Kamiya M., Kiguchi T. (2003) Rapid DNA extraction method from soybean seeds. Breeding Science 53:277-279.

Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510.

Kao C.H., Zeng Z.B., Teasdale R.D. (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203-1216.

Karikari S.K. (2000) Variability Between Local and Exotic Bambara Groundnut Landraces in Botswana. African Crop Science Journal 8:145-152.

Karikari S.K., Chaba O., Molosiwa B. (1999) Effects of intercropping bambara groundnut on pearl millet, sorghum and maize in Botswana. African Crop Science Journal 7:143-152.

Karunaratne A.S., Azam-Ali S.N., Steduto P. (2011) Calibration and Validation of Fao-Aquacrop Model for Irrigated and Water Deficient Bambara Groundnut. Experimental Agriculture 47:509-527. DOI: Doi 10.1017/S0014479711000111.

Kaur S., Cogan N.O., Pembleton L.W., Shinozuka M., Savin K.W., Materne M., Forster J.W. (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics 12:265. DOI: 1471-2164-12-265 [pii] 10.1186/1471-2164-12-265.

Kearsey M., Pooni H. (1996) The genetical analysis of quantitative traits. Chapman & Hall, London.

Kearsey M.J. (1998) The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany 49:1619-1623.

Kebour D., Boutekrabt A., Mefti M. (2012) Using inter simple sequence repeat (ISSR) markers to study genetic polymorphism of pistachio (Pistacia vera L.) in Algeria. African Journal of Biotechnology 11:7354-7360.

Keim P., Diers B.W., Olson T.C., Shoemaker R.C. (1990) Rflp Mapping in Soybean - Association between Marker Loci and Variation in Quantitative Traits. Genetics 126:735-742.

Kelkar Y.D., Tyekucheva S., Chiaromonte F., Makova K.D. (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Research 18:30-38. DOI: Doi 10.1101/Gr.7113408.

Kennard W.C., Poetter K., Dijkhuizen A., Meglic V., Staub J.E., Havey M.J. (1994) Linkages among RFLP, RAPD, isozyme, disease-resistance, and morphological markers in narrow and wide crosses of cucumber. Theoretical and applied genetics 89:42-48.

Keurentjes J.J.B., Willems G., van Eeuwijk F., Nordborg M., Koornneef M. (2011) A comparison of population types used for QTL mapping in *Arabidopsis thaliana*. Plant Genetic Resources-Characterization and Utilization 9:185-188. DOI: Doi 10.1017/S1479262111000086.

Khattak G.S.S., Ashraf M., Haq M.A., Srinives P. (2002) Genetics architecture of seed yield and yield components in mungbean (*Vigna radiata* (L.) Wilczek). Tropical Agriculture 79:260-264.

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C.J., Ohtsuki K., Shishiki T., Otomo Y., Murakami K., Iida Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y., Kurosaki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M., Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Niikura J., Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J., Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Hayatsu N., Imotani K., Ishii Y., Itoh M., Kagawa I., Kondo S., Konno H., Miyazaki A., Osato N., Ota Y., Saito R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T., Yoshino M., Hayashizaki Y., Yasunishi A. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-9. DOI: 10.1126/science.1081288 301/5631/376 [pii].

Kilian A., Huttner E., Wenzl P., Jaccoud D., Carling J., Caig V., Evers M., Heller-Uszynska K., Uszynski G., Cayla C., Patarapuwadol S., Xia L., Yang S., Thomson B. (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement, in: R. Tuberosa, *et al.* (Eds.), Proceedings of the International Congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution", Avenue media, Bologna,, Bologna, Italy. pp. 443-461.

Kilian B., Graner A. (2012) NGS technologies for analyzing germplasm diversity in genebanks*. Briefings in Functional Genomics 11:38-50. DOI: Doi 10.1093/Bfgp/Elr046.

Knott S.A., Elsen J.M., Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theoretical and applied genetics 93:71-80.

Knox M.R., Ellis T.H.N. (2002) Excess heterozygosity conttributes to genetic map espansmn in pea recombinant lnbred populations. Genetics 162:861-873.

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. (2002) Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol 43:1096-105.

Konishi T., Abe K., Matsuura S., Yano Y. (1990) Distorted segregation of the esterase isozyme genotypes in barley (*Hordeum vulgare* L.). Japanese Journal of Genetics 65:411-6.

Koornneef M., Alonso-Blanco C., Vreugdenhil D. (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. Annual Review of Plant Biology 55:141-172.

Kouassi N.J., Zoro B.I.A. (2010) Effect of sowing density and seedbed typed on yield and yield components in bambara groundnut (*Vigna subterranea* ) in woodland savannas of Côte d'Ivore. Experimental Agriculture 46:99-110.

Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J., Marra M.A. (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639-45. DOI: gr.092759.109 [pii] 10.1101/gr.092759.109.

Kumar P., Gupta V.K., Misra A.K., Modi D.R., Pandey B.K. (2009) Potential of Molecular Markers in Plant Biotechnology. Plant Omics 2:141-162.

Kumar V., Sharma S., Kero S., Sharma S., Sharma A.K., Kumar M., Bhat V. (2008) Assessment of genetic diversity in common bean (Phaseolus vulgaris L.) germplasm using amplified fragment length polymorphism (AFLP). Scientia Horticulturae 116:138-143. DOI: DOI 10.1016/j.scienta.2007.12.001.

Lander E.S., Botstein D. (1989) Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. Genetics 121:185-199.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newberg L.A. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-81. DOI: 0888-7543(87)90010-3 [pii].

Langridge P., Lagudah E.S., Holton T.A., Appels R., Sharp P.J., Chalmers K.J. (2001) Trends in genetic and genome analyses in wheat: A review. Australian Journal of Agricultural Research 52:1043-1077.

Laurie C.C., Chasalow S.D., LeDeaux J.R., McCarroll R., Bush D., Hauge B., Lai C.Q., Clark D., Rocheford T.R., Dudley J.W. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141-2155. DOI: DOI 10.1534/genetics.104.029686.

Lecomte L., Duffe P., Buret M., Servin B., Hospital F., Causse M. (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theoretical and applied genetics 109:658-668. DOI: DOI 10.1007/s00122-004-1674-0.

Legesse B.W., Myburg A.A., Pixley K.V., Botha A.M. (2007) Genetic diversity of African maize inbred lines revealed by SSR markers. Hereditas 144:10-17. DOI: DOI 10.1111/j.2006.0018-0661.01921.x.

Lehmann E.L. (1975) Nonparametrics. McGraw-Hill, New York.

Levinson G., Gutman G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-21.

Liang Q.A., Cheng X.H., Mei M.T., Yan X.L., Liao H. (2010) QTL analysis of root traits as related to phosphorus efficiency in soybean. Annals of Botany 106:223-234. DOI: Doi 10.1093/Aob/Mcq097.

Liang X., Chen X., Hong Y., Liu H., Zhou G., Li S., Guo B. (2009) Utility of EST-derived SSR in cultivated peanut (Arachis hypogaea L.) and Arachis wild species. BMC Plant Biol 9:35. DOI: 1471-2229-9-35 [pii] 10.1186/1471-2229-9-35.

Lincoln S., Daly M., Lander E. (1993) Constructing genetic linkage maps with MAPMAKER/EXP. Version 3.0. Whitehead Institute for Biomedical Research Technical Report, 3rd Edn.

Linnemann A.R. (1992) Bambara groundnut (*Vigna subterranea* ) literature : a revised and updated bibliography. Wageningen Agricultural University, Wageningen, Netherlands.

Linnemann A.R., Azam-Ali S.N. (1993) Bambara groundnut (*Vigna subterranea* ), in: J. T. Williams (Ed.), Pulses and Vegetables, Chapman & Hall, London. pp. 278-282.

Linnemann A.R., Craufurd P.Q. (1994) Effects of temperature and photoperiod on phenological development in three genotypes of bambara groundnut (*Vigna subterranea* ). Annals of Botany (London) 74:675-681.

Linnemann A.R., Westphal E., Wessel M. (1995) Photoperiod regulation of development and growth in bambaragroundnut (*Vigna subterranea*). Field Crops Research 40:39-47.

Liu H.B. (1998) Statistical Genomics, Linkage, Mapping and QTL Analysis. CRC, Boca Raton, Florida, 611 pp.

Liu X., You J., Guo L., Liu X., He Y., Yuan J., Liu G., Feng Z. (2011) Genetic Analysis of Segregation Distortion of SSR Markers in F2 Population of Barley. Journal of Agricultural Science 3.

Lopes F.C., Gomes R.L.F., Filho F.R.F. (2003) Genetic control of cowpea seed sizes. Scientia Agricola 60:315-318.

Lu H., Romero-Severson J., Bernardo R. (2002) Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105:622-628. DOI: 10.1007/s00122-002-0970-9.

Lucas M.R., Diop N.N., Wanamaker S., Ehlers J.D., Roberts P.A., Close T.J. (2011) Cowpea-Soybean Synteny Clarified through an Improved Genetic Map. Plant Genome 4:218-225. DOI: DOI 10.3835/plantgenome2011.06.0019.

Luikart G., England P.R., Tallmon D., Jordan S., Taberlet P. (2003) The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4:981-994.

Mace E.S., Xia L., Jordan D.R., Halloran K., Parh D.K., Huttner E., Wenzl P., Kilian A. (2008) DArT markers: diversity analyses and mapping in Sorghum bicolor. BMC Genomics 9:26. DOI: 1471-2164-9-26 [pii] 10.1186/1471-2164-9-26.

Mackay T.F.C. (2001) The genetic architecture of quantitative traits. Annual Review of Genetics 35:303-339.

Mackay T.F.C. (2004) The genetic architecture of quantitative traits: lessons from Drosophila. Current Opinion in Genetics & Development 14:253-257.

Mackill D.J., Ni J. (2000) Molecular mapping and marker assisted selection for major-gene traits in rice, in: G. S. Khush, *et al.* (Eds.), Proc. Fourth Int. Rice Genetics Symp, International Rice Research Institute., Los Ban˜os, The Philippines. pp. 137-151.

Madabhushi R.S. (1998) Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. Electrophoresis 19:224-230.

Maesen L.J.G.V.d., Somaatmadja S. (1989) Plant resources of South-East Asia. No. 1. Pulses. Pudoc, Wageningen, p 75.

Magbagbeola J.A.O., Adetoso J.A., Owolabi O.A. (2010) Neglected and underutilized species (NUS): A panacea for community focused development to poverty alleviation/poverty reduction in Nigeria. Journal of Economics and International Finance 2:208-211.

Mak C., Yap T.C. (1980) Inheritance of Seed Protein-Content and Other Agronomic Characters in Long Bean (*Vigna-Sesquipedalis* Fruw). Theoretical and applied genetics 56:233-239.

Makanda I., Tongoona P., Madamba R., Icishahayo D., Derera J. (2009) Path coefficient analysis of bambara groundnut pod yield components at four planting dates. Research Journal of Agriculture and Biological Sciences 5:287-292.

Manly K.F., Cudmore R.H., Meer J.M. (2001) Map manager QTX, cross-platform software for genetic mapping. Mammalian Genome 12:930-932.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L., Jarvie T.P., Jirage K.B., Kim J.B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F., Rothberg J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-80. DOI: nature03959 [pii] 10.1038/nature03959.

Markert C.L., Moller F. (1959) Multiple Forms of Enzymes: Tissue, Ontogenetic, and Species Specific Patterns. Proc Natl Acad Sci U S A 45:753-63.

Massawe F.J., Azam-Ali S.N., Roberts J.A. (2003) The impact of temperature on leaf appearance in bambara groundnut landraces. Crop Science 43:1375-1379.

Massawe F.J., Dickinson M., Roberts J.A., Azam-Ali S.N. (2002) Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces revealed by AFLP markers. Genome 45:1175-1180. DOI: 10.1139/g02-093.

Massawe F.J., Schenkel W., M. B.S., Temba E.M. (2004) Artificial hybridisation in bambara groundnut (*Vigna subterranea* (L.) Verdc.). In: conference proceedings of the international bambara groundnut symposium, 5-12 September 2003,Gaborone, Botswana:193-209.

Massawe F.J., Mwale S.S., Azam-Ali S.N., Roberts J.A. (2005) Breeding in bambara groundnut (*Vigna subterranea* (L.) Verdc.): strategic considerations. African Journal of Biotechnology 4:463-471.

Matthes M.C., Daly A., Edwards K.J. (1998) Amplified fragment length polymorphism (AFLP) Chapman and Hall, Cambridge.

Mayes S., Karunaratne A., Massawe F., Azam-Ali S., Crout N. (2009) Modelling landrace structure in underutilised crops- bambara groundnut as an example.

Mayes S., Massawe F.J., Alderson P.G., Roberts J.A., Azam-Ali S.N., Hermann M. (2011) The potential for underutilized crops to improve security of food production. J Exp Bot 63:1075-9. DOI: err396 [pii] 10.1093/jxb/err396.

Melchinger A.E., Utz H.F., Schon C.C. (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Metzker M.L. (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31-46. DOI: nrg2626 [pii] 10.1038/nrg2626.

Michelmore R. (1995) Molecular Approaches to Manipulation of Disease Resistance Genes. Annual Review of Phytopathology 33:393-427. DOI: DOI 10.1146/annurev.phyto.33.1.393.

Milczak M. (1971) A study of intervarietal hybrids of common vetch (Vicia sativa L.). Hodowla Roslin, Aklimatyzacja I Nasiennictwo 15:101-112.

Millan T., Winter P., Jungling R., Gil J., Rubio J., Cho S., Cobos M.J., Iruela M., Rajesh P.N., Tekeoglu M., Kahl G., Muehlbauer F.J. (2010) A consensus genetic map of chickpea (Cicer arietinum L.) based on 10 mapping populations. Euphytica 175:175-189. DOI: DOI 10.1007/s10681-010-0157-4.

Misangu R.N., Azmio A., Reuben S.O.W.M., Kusolwa P.M., Mulungu L.S. (2007) Path Coefficient Analysis among Components of Yield in Bambara Groundnut (*Vigna subterranea* L. Verdc.) Landraces under Screen House Conditions. . Journal of Agronomy 6:317-323. DOI: 10.3923/ja.2007.317.323.

Mohan J.S., Suprasanna P. (2011) Induced mutations for enhancing nutrition and food production. Gene Conserve 40:201-215.

Mohan M., Nair S., Bhagwat A., Krishna T.G., Yano M., Bhatia C.R., Sasaki T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3:87-103.

Mohler V., Singrun C. (2004) General considerations: marker-assisted selection, in: H. Lorz and G. Wenzel (Eds.), In Biotechnology in agriculture and forestry, vol. 55: Molecular marker systems, Springer, Berlin, Germany. pp. 305-317.

Molosiwa O.O. (2012) Genetic Diversity and Population structure analysis of bambara groundnut [*Vigna subterranea* (L.) Verdc.] Landraces using Morpho-agronomic Characters and SSR Markers, School of Biosciences, Division of Plant and Crop Sciences, The University of Nottingham Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD UK. pp. 284.

Mondini L., Noorani A., Pagnotta M.A. (2009) Assessing plant genetic diversity by molecular tools. Diversity 1:19-35. DOI: 10.3390/d1010019.

Moore S.S., Sargeant L.L., King T.J., Mattick J.S., Georges M., Hetzel D.J.S. (1991) The Conservation of Dinucleotide Microsatellites among Mammalian Genomes Allows the Use of Heterologous Pcr Primer Pairs in Closely Related Species. Genomics 10:654-660.

Morozova O., Marra M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255-64. DOI: S0888-7543(08)00165-1 [pii] 10.1016/j.ygeno.2008.07.001.

Morozova O., Hirst M., Marra M.A. (2009) Applications of New Sequencing Technologies for Transcriptome Analysis. Annual Review of Genomics and Human Genetics 10:135-151. DOI: DOI 10.1146/annurev-genom-082908-145957.

Moschini G. (2008) Biotechnology and the development of food markets: retrospect and prospects. European Review of Agricultural Economics 35:331-355. DOI: Doi 10.1093/Erae/Jbn014.

Muchero W., Diop N.N., Bhat P.R., Fenton R.D., Wanamaker S., Pottorff M., Hearne S., Cisse N., Fatokun C., Ehlers J.D., Roberts P.A., Close T.J. (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159-64. DOI: 0905886106 [pii] 10.1073/pnas.0905886106.

Muños S., Ranc N., Botton E., Bérard A., Rolland S., Duffé P., Carretero Y., Le Paslier M.C., Delalande C., Bouzayen M., Brunel D., Causse M. (2011) Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. Plant Phsiology 156:2244-2254.

Mwale S.S., Azam-Ali S.N., Massawe F.J. (2007) Growth and development of bambara groundnut (*Vigna subterranea* ) in response to soil moisture - 1. Dry matter and yield. European Journal of Agronomy 26:345-353. DOI: DOI 10.1016/j.eja.2006.09.007.

Nagabhushana K., Mane S.P., Hittalmani S. (2006) Comparative studies on QTL mapping by simple interval mapping and composite interval mapping models for selected growth and yield traits in rice (*Oryza sativa* L.). Indian Journal of Crop Science 1:97-101.

National Academy of Sciences N. (1979) National Academy of Sciences. Tropical Legumes: Resources for the Future, Natl. Acad. Press, Washington DC.

Nawab N.N., Subhani G.M., Mahmood K., Shakil Q., Saeed A. ( 2008) Genetic variability, correlation and path analysis studies in garden pea (Pisum sativum L.). Journal of Agricultural Research 46:333-340.

Naylor R.L., Falcon W.P., Goodman R.M., Jahn M.M., Sengooba T., Tefera H., Nelson R.J. (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. Food Policy 29:15-44. DOI: DOI 10.1016/j.foodpol.2004.01.002.

Nelson J.C. (1997) QGENE: Software for marker-based genomic analysis and breeding. Molecular Breeding 3:239-245.

Ner-Gaon H., Leviatan N., Rubin E., Fluhr R. (2007) Comparative cross-species alternative splicing in plants. Plant Physiology 144:1632-1641. DOI: DOI 10.1104/pp.107.098640.

Nguyen T.T., Taylor J., Redden J., Ford R. (2004) Genetic diversity estimates in Cicer using AFLP analysis. Plant Breeding 123:173-179.

Ntundu W.H., Bach I.C., Christiansen J.L., Andersen S.B. (2004) Analysis of genetic diversity in bambara groundnut (*Vigna subterrnea* (L.) Verdc.) landraces using amplified fragment length polymorphism (AFLP) markers. African Journal of Biotechnology 3:220-225.

Ntundu W.H., Shillah S.A., Marandu W.Y.F., Christiansen J.L. (2006) Morphological diversity of bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces in Tanzania. Genetic Resources and Crop Evolution 53:367-378. DOI: DOI 10.1007/s10722-004-0580-2.

Ofori I. (1996) Correlation and path-coefficient analysis of components of seed yield in bambara groundnut (*Vigna subterranea* ). Euphytica 91:103-107.

Okpuzor J., Ogbunugafor H.A., Okafor U., Sofidiya M.O. (2010) Identification of protein types in Bambara nut seeds: Perspectives for dietary protein supply in developing countries. EXCLI Journal 9:17-28.

Olaleke A.M., Olorunfemi O., Emmanuel A. (2006) A comparative study on the chemical and amino acid composition of some Nigerian under-utilized legume flours. Pakistan Journal of Nutrition 5:34-38.

Olukolu B.A., Mayes S., Stadler F., Ng N.Q., Fawole I., Dominique D., Azam-Ali S.N., Abbott A.G., Kole C. (2012) Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. Genetic Resources and Crop Evolution 59:347-358.

Omo-Ikerodah E.E., Fawole I., Fatokun C.A. (2008) Genetic mapping of quantitative trait loci (QTLs) with effects on resistance to flower bud thrips (*Megalurothrips sjostedti*) identified in recombinant inbred lines of cowpea (*Vigna unguiculata* (L.) Walp). African Journal of Biotechnology 7:263-270.

Oniang'o R.K., Shiundu K., Maundu P., Johns T. (2006) Diversity, nutrition and food security: the case of African leafy vegetables. M. S. Swaminathan Research Foundation, Chennai, India, Chennai, India, April 18 – 19, 2005.

Onwubiko N.I.C., Odum O.B., Utazi C.O., Poly-Mbah P.C. (2011) Studies on the Adaptation of Bambara Groundnut [*Vigna subterranea* (L.) Verdc.] in Owerri Southeastern Nigeria. New York Science Journal 4:60-67. DOI: 10.3923/aj.2011.60.65.

Ouedraogo M., Ouedraogo J.T., J.B. T., Bilma D., Dabire C.B., Konate G. (2008) Characterization and evaluation of accessions of Bambara groundnut (*Vigna subterranea* (L.) Verdcourt) from Burkina Faso. Sciences & Nature 5:191-197.

Padulosi S., Hoeschle-Zeledon I. (2004) Underutilized plant species: what are they? . LEISA Magazine 20:5-6.

Padulosi S., Bergamini N., Lawrence T. (2011) On farm conservation of neglected and underutilized species: status, trends and novel approaches to cope with climate change: Proceedings of an International Conference, Frankfurt, 14-16 June, 2011. Bioversity International, Rome.

Padulosi S., Hodgkin T., Williams J.T., Haq N. (2002) Underutilized crops: Trends, challenges and opportunities in the 21st century, Managing plant genetic diversity. Edition 1, CABI Publishing. International Plant Genetic Resources Institute (IPGRI). pp. 323-338.

Panthee D.R., Pantalone V.R., Saxton A.M., West D.R., Sams C.E. (2007) Quantitative trait loci for agronomic traits in soybean. Plant Breeding 126:51-57.

Papa R., Gepts P. (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (Phaseolus vulgaris L.) from Mesoamerica. Theor Appl Genet 106:239-50. DOI: 10.1007/s00122-002-1085-z.

Paraguison R.C., Faylon M.P., Flores E.B., Cruz L.C. (2012) Improved RAPD-PCR for Discriminating Breeds of Water Buffalo. Biochem Genet 50:579-84. DOI: 10.1007/s10528-012-9502-8.

Paran I., Zamir D. (2003) Quantitative traits in plants: beyond the QTL. Trends in Genetics 19:303-306. DOI: Doi 10.1016/S0168-9525(03)00117-3.

Paran I., Goldman I., Tanksley S.D., Zamir D. (1995) Recombinant Inbred Lines for Genetic-Mapping in Tomato. Theoretical and applied genetics 90:542-548.

Pareek C.S., Smoczynski R., Tretyn A. (2011) Sequencing technologies and genome sequencing. Journal of Applied Genetics 52:413-435. DOI: DOI 10.1007/s13353-011-0057-x.

Park Y.-J., Lee J.K., Kim N.-S. (2009) Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crops. Molecules 14:4546-4569.

Pasquet R.S. (2003) Bambara groundnut and cowpea gene-pool organization and domestication: In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8- 12, August, 2003:263-272.

Pasquet R.S., Fotso M. (1997) The ORSTOM bambara groundnut collection, in: J. Heller, *et al.* (Eds.), *Proceedings of the workshop on conservation and improvement of bambara groundnut (Vigna subterranea* (L) Verdc.*)*, Harare, Zimbabwe. pp. 119-123.

Pasquet R.S., Schwedes S., Gepts P. (1999) Isozyme diversity in Bambara groundnut. Crop Science 39:1228-1236.

Paterson A.H. (1996) Making genetic maps. In: Paterson AH (ed.) Genome mapping in plants, San Diego, California: Academic Press, Austin, Texas, pp. 23-39.

Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E., Tanksley S.D. (1988) Resolution of Quantitative Traits into Mendelian Factors by Using a Complete Linkage Map of Restriction Fragment Length Polymorphisms. Nature 335:721-726.

Paterson A.H., Damon S., Hewitt J.D., Zamir D., Rabinowitch H.D., Lincoln S.E., Lander E.S., Tanksley S.D. (1991) Mendelian Factors Underlying Quantitative Traits in Tomato - Comparison across Species, Generations, and Environments. Genetics 127:181-197.

Peakall R., Gilmore S., Keys W., Morgante M., Rafalski A. (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Molecular Biology and Evolution 15:1275-1287.

Peng J.H., Lapitan N.L. (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics 5:80-96. DOI: 10.1007/s10142-004-0128-8.

Polhill R.M., Raven P.H., Stirton C.H. (1981) Evolution and Systematics of the Leguminosae, in: R. M. Polhill and P. H. Raven (Eds.), Advances in Legume Systematics Part 1, Royal Botanic Gardens, Kew, UK. pp. 1-26.

Pompanon F., Bonin A., Bellemain E., Taberlet P. (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847-59. DOI: nrg1707 [pii] 10.1038/nrg1707.

Pottorff M., Ehlers J.D., Fatokun C., Roberts P.A., Close T.J. (2012) Leaf morphology in Cowpea [*Vigna unguiculata* (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. BMC Genomics 13. DOI: Artn 234 Doi 10.1186/1471-2164-13-234.

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

Price A.H. (2006) Believe it or not, QTLs are accurate! Trends in Plant Science 11:213-216. DOI: DOI 10.1016/j.tplants.2006.03.006.

Prokop Z. (2004) Mapping polygenes - methods, constraints, and application in evolutionary ecology. Wiadomosci Ekologiczne 50:19-29.

Provan J., Russell J.R., Booth A., Powell W. (1999) Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. Mol Ecol 8:505-11.

Qadir S.A., Datta S., Singh N.P., Kumar S. (2007) Development of highly polymorphic SSR markers for chickpea (Cicer arietinum L.) and their use in parental polymorphism. Indian Journal of Genetics & Plant Breeding 67:329-333.

Qi X., Cui F., Yu L., Ding A., Li J., Chen G., Wang H. (2010) Molecular Tagging Wheat Powdery Mildew Resistance Gene Pm21 by EST-SSR and STS Markers. Molecular Plant Breeding 1.

Rafalski A. (2002) Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

Rakshit S., Rakshit A., Patil J.V. (2012) Multiparent intercross populations in analysis of quantitative traits. Journal of Genetics 91:111-117.

Rasheed S., Hanif M., Sadiq S., Abbas G., Asghar M.J., Haq M.A. (2008) Inheritance of seed yield and related traits in some lentil *(Lens culinaris* Medik) genotypes. Pakistan Journal of Agricultural Sciences 45:49-52.

Rauf S., Silva J.A.T., Khan A.A., Naveed A. (2010) Consequences of Plant Breeding on Genetic Diversity. International journal of Plant Breeding 4:1-21.

Reece K.S., Ribeiro W.L., Gaffney P.M., Carnegie R.B., Allen S.K. (2004) Microsatellite marker development and analysis in the eastern oyster (Crassostrea virginica): Confirmation of null alleles and non-Mendelian segregation ratios. Journal of Heredity 95:346-352. DOI: DOI 10.1093/jhered/esh058.

Ribaut J.-M., Betran J. (1999) Single large-scale marker-assisted selection (SLS-MAS). Molecular Breeding 5:531-541.

Ribaut J.M., Hoisington D.A. (1998) Marker-assisted selection: New tools and strategies. Trends in Plant Science 3:236-239.

Ribaut J.M., Jiang C., GonzalezdeLeon D., Edmeades G.O., Hoisington D.A. (1997) Identification of quantitative trait loci under drought conditions in tropical maize .2. Yield components and marker-assisted selection strategies. Theoretical and applied genetics 94:887-896. DOI: DOI 10.1007/s001220050492.

Richards R.I., Sutherland G.R. (1992) Heritable unstable DNA sequences. Nat Genet 1:7-9. DOI: 10.1038/ng0492-7.

Risch N. (1992) Genetic-Linkage - Interpreting Lod Scores. Science 255:803-804.

Roa A.C., Chavarriaga-Aguirre P., Duque M.C., Maya M.M., Bonierbale M.W., Iglesias C., Tohme J. (2000) Cross-species amplification of cassava (Manihot esculenta) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. Am J Bot 87:1647-55. DOI: 87/11/1647 [pii].

Roder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganal M.W. (1998) A microsatellite map of wheat. Genetics 149:2007-2023.

Rodriguez-Zas S.L., Southey B.R., Heyen D.W., Lewin H.A. (2002) Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. Journal of Dairy Science 85:3081-3091.

Rohman M.M., Iqbal Hussain A.S.M., Arifin M.D.S., Akhter Z., Hasanuzzaman M. (2003) Genetic Variability, Correlation and Path Analysis in Mungbean. Asian Journal of Plant Sciences 2:1209-1211. DOI: 10.3923/ajps.2003.1209.1211.

Roy S.K., Abdul Karim M.d., Aminul Islam A.K.M., Bari M.d.N., Khaleque Mian M.A., Hidaka T.H. (2006) Relationship between yield and its component characters of bush bean (Phaseolus vulgaris L.). South Pacific Studies 27:13-23(Published from Japan).

Rozen S., Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers, in: S. Krawetz and S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa. pp. 365-386.

Rungis D., Berube Y., Zhang J., Steven R., Ritland C.E., Ellis B.E., Douglas C., Bohlmann J., Ritland K. (2004) Robust simple sequence repeat markers for spruce (Picea spp.) from expressed sequence tags. Theoretical and applied genetics 109:1283-1294.

Saha M.C., Mian M.A.R., Eujayl I., Zwonitzer J.C., Wang L., May G.D. (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theoretical and applied genetics 109:783-791.

Sandal N., Krusell L., Radutoiu S., Olbryt M., Pedrosa A., Stracke S., Sato S., Kato T., Tabata S., Parniske M., Bachmair A., Ketelsen T., Stougaard J. (2002) A genetic linkage map of the model legume Lotus japonicus and strategies for fast mapping of new loci. Genetics 161:1673-1683.

Sanger F., Nicklen S., Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463-7.

Sansaloni C.P., Petroli C.D., Carling J., Hudson C.J., Steane D.A., Myburg A.A., Grattapaglia D., Vaillancourt R.E., Kilian A. (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. Plant Methods 6:16. DOI: 1746-4811-6-16 [pii] 10.1186/1746-4811-6-16.

Santra D.K., Tekeoglu M., Ratnaparkhe M., Kaiser W.J., Muehlbauer F.J. (2000) Identification and mapping of QTLs conferring resistance to ascochyta blight in chickpea. Crop Science 40:1606-1612.

Sato S., Isobe S., Tabata S. (2010) Structural analyses of the genomes in legumes. Current Opinion in Plant Biology 13:146-152.

Sax K. (1923) The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. Genetics 8:552-60.

Saxena R.K., Prathima C., Saxena K.B., Hoisington D.A., Singh N.K., Varshney R.K. (2010) Novel SSR Markers for Polymorphism Detection in Pigeonpea (Cajanus spp.). Plant Breeding 129:142-148.

Schloss J.A. (2008) How to get genomes at one ten-thousandth the cost. NATURE BIOTECHNOLOGY 26:1113-1115. DOI: Doi 10.1038/Nbt1008-1113.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233-4. DOI: 10.1038/72708.

Schuster-Gajzágó I. (2009) Nutritional Aspects of Legumes, in: G. Fuleky (Ed.), Cultivated Plants, Primarily as Food Sources, Volume 1, Eolss Publishers Company Limited, 2009, Hungary.

Schuster S.C. (2008) Next-generation sequencing transforms today's biology. Nature Methods 5:16-18.

Scott K.D., Eggler P., Seaton G., Rossetto M., Ablett E.M., Lee L.S., Henry R.J. (2000) Analysis of SSRs derived from grape ESTs. Theoretical and applied genetics 100:723-726.

Sellschop J.P.F. (1962) Cowpeas (*Vigna unguiculata* (L) Walp). Field Crop Abstracts 15:259-266.

Semagn K., Bjornstad A., Ndjiondjop M.N. (2006a) Principles, requirements and prospects of genetic mapping in plants. African Journal of Biotechnology 5:2569-2587.

Semagn K., Bjornstad A., Ndjiondjop M.N. (2006b) An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540-2568.

Semagn K., Bjornstad A., Xu Y. (2010) The genetic dissection of quantitative traits in crops. Electronic Journal of Biotechnology 13:Article No.: 14.

Semagn K., Skinnes H., Bjornstad A., Maroy A.G., Tarkegne Y. (2007) Quantitative trait loci controlling Fusarium head blight resistance and low deoxynivalenol content in hexaploid wheat population from 'Arina' and NK93604. Crop Science 47:294-303. DOI: DOI 10.2135/cropsci2006.02.0095.

Semagn K., Bjornstad A., Skinnes H., Maroy A.G., Tarkegne Y., William M. (2006c) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545-55. DOI: g06-002 [pii]10.1139/g06-002.

Senior M.L., Murphy J.P., Goodman M.M., Stuber C.W. (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Science 38:1088-1098.

Senko O.V., Kuznetsova A.V. (2006) The Optimal Valid Partitioning Procedures. Statistics on the Internet http://statjournals.net/, April, 2006.

Sesay A., Magagula C.N., Mansuetus A.B. (2008) Influence of sowing date and environmental factors on the development and yield of bambara groundnut (*Vigna subterranea* ) landraces in a sub-tropical region. Experimental Agriculture 44:167-183.

Sévérin B.B., Yao D. (2011) Variabilité morphologique et agronomique des variétés traditionnelles de voandzou [*Vigna subterranea* (L.) Verdc. (fabaceae)] de Cote d'Ivoire. Journal of Applied Biosciences 41:2820-2835.

Shafie M.S.B., Hasan S.M.Z., Zain A.M., Shah R.M. (2011) RAPD and ISSR markers for comparative analysis of genetic diversity in wormwood capillary (Artemisia capillaris) from Negeri Sembilan, Malaysia. Journal of Medicinal Plants Research 5:4426-4437.

Shimelis H., Shiringani R. (2010) Variance components and heritabilities of yield and agronomic traits among cowpea genotypes. Euphytica 176:383-389.

Shoemaker R.C., Grant D., Olson T., Warren W.C., Wing R., Yu Y., Kim H., Cregan P., Joseph B., Futrell-Griggs M., Nelson W., Davito J., Walker J., Wallis J., Kremitski C., Scheer D., Clifton S.W., Graves T., Nguyen H., Wu X., Luo M., Dvorak J., Nelson R., Cannon S., Tomkins J., Schmutz J., Stacey G., Jackson S. (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome 51:294-302. DOI: g08-010 [pii] 10.1139/G08-010.

Shultz J.L., Kazi S., Bashir R., Afzal J.A., Lightfoot D.A. (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081-90. DOI: 10.1007/s00122-007-0501-9.

Siise A., Massawe F.J. (2012) Microsatellites based marker molecular analysis of Ghanaian bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces alongside morphological characterization Genetic Resources and Crop Evolution Published online: 6 July 2012. DOI: 10.1007/s10722-012-9874-y.

Smailus D.E., Marziali A., Dextras P., Marra M.A., Holt R.A. (2005) Simple, robust methods for high-throughput nanoliter-scale DNA sequencing. Genome Research 15:1447-1450.

Smartt J. (1985) Evolution of Grain Legumes. III. Pulses in the Genus *Vigna*. Experimental Agriculture 21:87-100.

Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B., Hood L.E. (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674-9. DOI: 10.1038/321674a0.

Snedecor G.W., Cochran W.G. (1989) Statistical Methods, Eighth Edition. Iowa State University Press.

Solh M., Saxena M.C. (2011) Food security and climate change in dry areas: proceedings of an International Conference, 1-4 February 2010, Amman, Jordan. PO Box 5466, Aleppo, Syria: International Center for Agricultural Research in the Dry Areas (ICARDA). viii + 369 pp.

Somta P., Rungnoi O., Chankaew S., Srinives P. (2011a) Crossspecies amplification of microsatellite markers in Bambara groundnut (*Vigna subterranea* ) and their application in diversity study. Acta Horticulture: *In press.*

Somta P., Chankaew S., Rungnoi O., Srinives P. (2011b) Genetic diversity of the Bambara groundnut (*Vigna subterranea*  (L.) Verdc.) as assessed by SSR markers. Genome 54:898-910.

Somta P., Musch W., Kongsamai B., Chanprame S., Nakasathien S., Toojinda T., Sorajjapinun W., Seehalak W., Tragoonrung S., Srinives P. (2008) New microsatellite markers isolated from mungbean (*Vigna radiata* (L.) Wilczek). Molecular Ecology Resources 8:1155-1157. DOI: DOI 10.1111/j.1755-0998.2008.02219.x.

Song Q.J., Marek L.F., Shoemaker R.C., Lark K.G., Concibido V.C., Delannay X., Specht J.E., Cregan P.B. (2004) A new integrated genetic linkage map of the soybean. Theoretical and applied genetics 109:122-128. DOI: DOI 10.1007/s00122-004-1602-3.

Soranzo N., Provan J., Powell W. (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42:158-61.

Souci S.W., Fachmannn W., Kraut H. (2000) Food composition and nutrition tables. Wissenschaft verlags GmbH, Stuttgart.

Stadler F. (2009) Analysis of differential gene expression under water-deficit stress and genetic diversity in bambara groundnut [*Vigna subterranea*  (L.) Verdc.] using novel high-throughput technologies, Technischen Universität München. pp. 177.

Stam P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant Journal 3:739-744.

Stapley J., Reger J., Feulner P.G.D., Smadja C., Galindo J., Ekblom R., Bennison C., Ball A.D., Beckerman A.P., Slate J. (2010) Adaptation genomics: the next generation. Trends in Ecology & Evolution 25:705-712. DOI: DOI 10.1016/j.tree.2010.09.002.

Stephens J.M. (2003) Bambara groundnut: Voandzeia subterranea (L.) Thouars. University of Florida, IFAS Extension.http://edis.ifas.ufl.edu/pdffiles/MV/MV01400.pdf.

Stephens M.A. (1974) Edf Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association 69:730-737.

Stewart S., Wickramasinghe D., Dorrance A.E., Robertson A.E. (2011) Comparison of three microsatellite analysis methods for detecting genetic diversity in

Phytophthora sojae (Stramenopila: Oomycete). Biotechnology Letters 33:2217-2223. DOI: DOI 10.1007/s10529-011-0682-9.

Strickler S.R., Bombarely A., Mueller L.A. (2012) Designing a Transcriptome Next-Generation Sequencing Project for a Nonmodel Plant Species. American Journal of Botany 99:257-266.

Stuber C.W., Polacco M., Lynn M. (1999) Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Science 39:1571-1583.

Sturtevant A.H. (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. Journal of Experimental Zoology 14:43-59.

Sundaram R.M., Naveenkumar B., Biradar S.K., Balachandran S.M., Mishra B., IlyasAhmed M., Viraktamath B.C., Ramesha M.S., Sarma N.P. (2008) Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment. Euphytica 163:215-224. DOI: DOI 10.1007/s10681-007-9630-0.

Suwanprasert J., Toojinda T., Srinives P., Chanprame S. (2006) Hybridization technique for bambara groundnut. Breeding Science 56:125-129.

Swanevelder C.J. (1998) Bambara - food for Africa: (*Vigna subterranea* bambara groundnut). National Department of Agriculture, South Africa.

Symonds V.V., Godoy A.V., Alconada T., Botto J.F., Juenger T.E., Casal J.J., Lloyd A.M. (2005) Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. Genetics 169:1649-58. DOI: genetics.104.031948 [pii] 10.1534/genetics.104.031948.

Tanksley S.D. (1993) Mapping polygenes, Annual Review of Genetics, Annual Reviews Inc. {a}, P.O. Box 10139, 4139 El Camino Way, Palo Alto, California 94306, USA. pp. 205-233.

Tanyolac B., Ozatay S., Kahraman A., Muehlbauer F. (2010) Linkage mapping of lentil (Lens culinaris L.) genome using recombinant inbred lines revealed by AFLP, ISSR, RAPD and some morphologic markers. Journal of Agricultural Biotechnology and Sustainable Development 2:001-006.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17:6463-71.

Taylor M., Jaenicke H., Skelton P., Mathur P.N. (2009) Regional Consultation on Crops for The Future: Towards Food, Nutritional, Economic and Environmental Security in The Pacific, Secretariat of the Pacific Community. Asia-Pacific Association of Agricultural Research Institutions. Bioversity International. Crops for the Future. National Agricultural Research Institute, PNG. pp. 28.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441-52. DOI: 10.1101/gr.184001.

Thiel T., Michalek W., Varshney R.K., Graner A. (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and applied genetics 106:411-422.

Thies E. (2000) Promising and Underutilized Species, Crops and Breeds. Deutsche Gesellschaft Für Technische Zusammenarbeit (GTZ) GmbH, 23 pp. http://www.ilri.org/html/thies3.pdf.

Thoquet P., Gherardi M., Journet E.P., Kereszt A., Ane J.M., Prosperi J.M., Huguet T. (2002) The molecular genetic linkage map of the model legume Medicago truncatula: an essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol 2:1.

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S.t. (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286-9. DOI: 10.1038/90135 90135 [pii].

Thudi M., Bohra A., Nayak S.N., Varghese N., Shah T.M., Penmetsa R.V., Thirunavukkarasu N., Gudipati S., Gaur P.M., Kulwal P.L., Upadhyaya H.D., KaviKishor P.B., Winter P., Kahl G., Town C.D., Kilian A., Cook D.R., Varshney R.K. (2011) Novel SSR Markers from BAC-End Sequences, DArT Arrays and a Comprehensive Genetic Map with 1,291 Marker Loci for Chickpea (Cicer arietinum L.). PLoS One 6. DOI: ARTN e27275. DOI 10.1371/journal.pone.0027275.

Tinker N.A., Kilian A., Wight C.P., Heller-Uszynska K., Wenzl P., Rines H.W., Bjornstad A., Howarth C.J., Jannink J.L., Anderson J.M., Rossnagel B.G., Stuthman D.D., Sorrells M.E., Jackson E.W., Tuvesson S., Kolb F.L., Olsson O., Federizzi L.C., Carson M.L., Ohm H.W., Molnar S.J., Scoles G.J., Eckstein P.E., Bonman J.M., Ceplitis A., Langdon T. (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39. DOI: 1471-2164-10-39 [pii] 10.1186/1471-2164-10-39.

Toader M., Roman G.H.V., Ionescu A. (2011) Chemical composition and nutritional values of some alternative crops promoted in organic agriculture. UASVM Bucharest, Series A, Vol. LIV.

Toungos M.D., Sajo A.A., Gungula D.T. (2009) Recommended Fertilizer Levels on Bambara Groundnut (*Vigna subterranea* (L) Verdc.) in Yola Adamawa State, Nigeria. Agricultural Journal 4:14-21.

Trethowan R.M., Mujeeb-Kazi A. (2008) Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. Crop Science 48:1255-1265. DOI: DOI 10.2135/cropsci2007.08.0477.

Trivedi S. (2004) Microsatellites (SSRs): Puzzles within puzzle. Indian Journal of Biotechnology 3:331-347.

Turner N.C. (1979) Drought resistance and adaptation to water deficits in crops, in: H. Mussel and R. C. States (Eds.), Stress Physiology in Crop Plants, Wiley and Sons, New York. pp. 343-373.

Uguru M.I., Ezeh N.E. (1997) Growth, nodulation and yield of bambara groundnut (*Vigna subterranea* (L) Verdc.) on selected Nigerian soils. Journal of the Science of Food and Agriculture 73:377-382.

Uguru M.I., Agwatu U.K. (2006) Cytogenetic studies on bambara groundnut (*Vigna subterranea* (L.) Verdc.). Journal of genetic and breeding 60:125-132.

Utz H., Melchinger A. (1996) PLABQTL: A program for composite interval mapping of QTL. Journal of quantitative train loci 2:1-5.

Van Ooijen J.W. (2006) JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV: Wageningen, Netherlands.

Van Ooijen J.W. (2009) MapQTL 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands.

Van Ooijen J.W., Maliepaard C. (2001) MapQTL version 4.0: Software for the calculation of QTL positions on genetic maps. CPRO-DELO, Wageningen. The Netherlands.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538. DOI: DOI 10.1111/j.1471-8286.2004.00684.x.

Varshney R.K., Graner A., Sorrells M.E. (2005a) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48-55. DOI: S0167-7799(04)00322-1 [pii] 10.1016/j.tibtech.2004.11.005.

Varshney R.K., Nayak S.N., May G.D., Jackson S.A. (2009a) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27:522-30.

Varshney R.K., Close T.J., Singh N.K., Hoisington D.A., Cook D.R. (2009b) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202-10. DOI: S1369-5266(08)00202-1 [pii] 10.1016/j.pbi.2008.12.004.

Varshney R.K., Sigmund R., Borner A., Korzun V., Stein N., Sorrells M.E., Langridge P., Graner A. (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. plant Science 168:195-202. DOI: DOI 10.1016/j.plantsci.2004.08.001.

Vezzulli S., Troggio M., Coppola G., Jermakow A., Cartwright D., Zharkikh A., Stefanini M., Grando M.S., Viola R., Adam-Blondon A.-F., Thomas M., This P., Velasco R. (2008) A reference integrated map for cultivated grapevine (Vitis vinifera L.) from three crosses, based on 283 SSR and 501 SNP-based markers. Theoretical and applied genetics 117:499-511.

Vietmeyer D.N. (1979) Tropical legumes: Research for the future. National Accademic press, Washington.

Vinod K.K. (2009) Genetic mapping of quantitative genetic loci and marker assisted selection in plantation crops In: proceedings of the trainning programme on " In vitro techniques in plantation crops" Central Plantation Crops Research Institute, Kasaragod, India, pp 11-132.

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

Wanamaker S., Close T.J. (2011) HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu (verifi ed 27 Sept. 2011). University of California, Riverside, CA.

Wang B., Chee P.W. (2010) Application of advanced backcross quantitative trait locus (QTL) analysis in crop improvement. Journal of Plant Breeding and Crop Science 2:221-232.

Wang H., Walla J.A., Zhong S., Huang D., Dai W. (2012) Development and cross-species/genera transferability of microsatellite markers discovered using 454 genome sequencing in chokecherry (*Prunus virginiana* L.). Plant Cell Reports:DOI 10.1007/s00299-012-1315-z.

Wang J., McClean P.E., Lee R., Goos R.J., Helms T. (2008) Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. Theor Appl Genet 116:777-87. DOI: 10.1007/s00122-008-0710-x.

Wang J.L., Barkley N.A., Jenkins T.M. (2009) Microsatellite markers in plants and insects Part I: Applications of biotechnology. Genes, Genomes, and Genomics 3:54-67.

Wang M.L., Gillaspie A.G., Newman M.L., Dean R.E., Pittman R.N., Morris J.B., Pederson G.A. (2004a) Transfer of simple sequence repeat (SSR) markers across the legume family for germplasm characterization and evaluation. Plant Genetic Resources Characterization and Utilization 2:107-119.

Wang M.L., Huang L., Bongard-Pierce D.K., Belmonte S., Zachgo E.A., Morris J.W., Dolan M., Goodman H.M. (1997) Construction of an approximately 2 Mb contig in the region around 80 cM of *Arabidopsis thaliana* chromosome 2. Plant J 12:711-30.

Wang S., Basten C.J., Zeng Z.B. (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. ( http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wang X.W., Kaga A., Tomooka N., Vaughan D.A. (2004b) The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. Theoretical and applied genetics 109:352-360. DOI: DOI 10.1007/s00122-004-1634-8.

Weber D., Helentjaris T. (1989) Mapping Rflp Loci in Maize Using B-a Translocations. Genetics 121:583-590.

Wenxin L., Da-Hao Z., Kyujung V., Suk-Ha L. (2008) QTL Mapping for Major Agronomic Traits across Two Years in Soybean (*Glycine max* L. Merr.). crop Science biotechnology 11:171-176.

Wenzl P., Carling J., Kudrna D., Jaccoud D., Huttner E., Kleinhofs A., Kilian A. (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci U S A 101:9915-20. DOI: 10.1073/pnas.04010761010401076101 [pii].

Wenzl P., Li H., Carling J., Zhou M., Raman H., Paul E., Hearnden P., Maier C., Xia L., Caig V., Ovesna J., Cakir M., Poulsen D., Wang J., Raman R., Smith K.P., Muehlbauer G.J., Chalmers K.J., Kleinhofs A., Huttner E., Kilian A. (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:Article No.: 206.

White G., Powell W. (1997) Isolation and characterization of microsatellite loci in Swietenia humilis Meliaceae an endangered tropical hardwood species. Molecular Ecology 6:851-860.

Whitmore J.S. (2000) Drought Management on Farmland Kluwer Academic Publishers, Dordrecht, the Netherlands.

Wietholter P., Sereno M.J.C.D., Terra T.D., Silva S.D.D.E., Barbosa J.F. (2008) Genetic Variability En Corn Landraces from Southern Brazil. Maydica 53:151-159.

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. (1990) DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic Markers. Nucleic Acids Research 18:6531-6536.

Williams J.T., Brenner D. (1995) Grain amaranth (Amaranthus species), in: J. T. Williams (Ed.), Underutilised Crops: Cereals and Pseudocereals, Chapman & Hall, London, London. pp. 129-186.

Williams J.T., Haq N. (2002) Global research on underutilised crops: an assessment of current activities and proposal for an enhanced cooperation International Centre for Underutlised Crops.

Winter P., Kahl G. (1995) Molecular marker technologies for plant improvement. World Journal of Microbiology and Biotechnology 11:438-448.

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

Wu K.S., Tanksley S.D. (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225-35.

Wu X.L., He C.Y., Wang Y.J., Zhang Z.Y., Dongfang Y., Zhang J.S., Chen S.Y., Gai J.Y. (2001) [Construction and analysis of a genetic linkage map of soybean]. Yi Chuan Xue Bao 28:1051-61.

Wu Y.H., Bhat P.R., Close T.J., Lonardi S. (2008) Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. Plos Genetics 4. DOI: ARTN e1000212 DOI 10.1371/journal.pgen.1000212.

Xia L., Peng K., Yang S., Wenzl P., de Vicente M.C., Fregene M., Kilian A. (2005) DArT for high-throughput genotyping of Cassava (Manihot esculenta) and its wild relatives. Theoretical and applied genetics 110:1092-1098.

Xu P., Wu X.H., Wang B.G., Liu Y.H., Qin D.H., Ehlers J.D., Close T.J., Hu T.T., Lu Z.F., Li G.J. (2010) Development and polymorphism of *Vigna unguiculata* ssp unguiculata microsatellite markers used for phylogenetic analysis in asparagus bean (*Vigna unguiculata* ssp sesquipedialis (L.) Verdc.). Molecular Breeding 25:675-684. DOI: DOI 10.1007/s11032-009-9364-x.

Xu P., Wu X., Wang B., Liu Y., Ehlers J.D., Close T.J., Roberts P.A., Diop N.N., Qin D., Hu T., Lu Z., Li G. (2011) A SNP and SSR based genetic map of asparagus bean (*Vigna. unguiculata* ssp. sesquipedialis) and comparison with the broader species. PLoS One 6:e15952. DOI: 10.1371/journal.pone.0015952.

Xu Y.B., Crouch J.H. (2008) Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48:391-407. DOI: DOI 10.2135/cropsci2007.04.0191.

Xue D.W., Zhou M.X., Zhang X.Q., Chen S., Wei K., Zeng F.R., Mao Y., Wu F.B., Zhang G.P. (2010) Identification of QTLs for yield and yield components of barley under different growth conditions. Journal of Zhejiang University-Science B 11:169-176. DOI: DOI 10.1631/jzus.B0900332.

Yamagishi M., Takeuchi Y., Tanaka I., Kono I., Murai K., Yano M. (2010) Segregation distortion in F2 and doubled haploid populations of temperate japonica rice. Journal of Genetics 89:237-241.

Yang J., Hu C.C., Hu H., Yu R.D., Xia Z., Ye X.Z., Zhu J. (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24:721-723. DOI: DOI 10.1093/bioinformatics/btm494.

Yang S.Y., Pang W., Ash G., Harper J., Carling J., Wenzl P., Huttner E., Zong X.X., Kilian A. (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theoretical and applied genetics 113:585-595. DOI: DOI 10.1007/s00122-006-0317-z.

Yang S.Y., Saxena R.K., Kulwal P.L., Ash G.J., Dubey A., Harper J.D., Upadhyaya H.D., Gothalwal R., Kilian A., Varshney R.K. (2011) The first genetic map of pigeon pea based on diversity arrays technology (DArT) markers. J Genet 90:103-9.

Yeboah M.A., Chen X.H., Feng C.R., Liang G.H., Gu M.H. (2007) A genetic linkage map of cucumber (Cucumis sativus L) combining SRAP and ISSR markers. African Journal of Biotechnology 6:2784-2791.

Yi S.S., Jatoi S.A., Fujimura T., Yamanaka S., Watanabe J., Watanabe K.N. (2008) Potential loss of unique genetic diversity in tomato landraces by genetic colonization of modern cultivars at a non-center of origin. Plant Breeding 127:189-196. DOI: DOI 10.1111/j.1439-0523.2007.01446.x.

Yim Y.S., Davis G.L., Duru N.A., Musket T.A., Linton E.W., Messing J.W., McMullen M.D., Soderlund C.A., Polacco M.L., Gardiner J.M., Coe E.H. (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiology 130:1686-1696. DOI: Doi 10.1104/Pp.013474.

Young N.D. (1994) Constructing a plant genetic linkage map with DNA markers, p. 39-57, In: I. K.V. Ronald & L. Phillips (Eds.), DNA-based markers in plants. Kluwer, Dordrecht/ Boston/London.

Zabeau M., Vos P. (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Office, publication 0 534 858 A1, bulletin 93/13. Available from www.freepatentsonline.com.

Zalapa J.E., Cuevas H., Zhu H., Steffan S., Senalik D., Zeldin E., McCown B., Harbut R., Simon P. (2012) Using Next-Generation Sequencing Approaches to Isolate Simple Sequence Repeat (Ssr) Loci in the Plant Sciences. American Journal of Botany 99:193-208.

Zeng Z.B. (1993) Theoretical Basis for Separation of Multiple Linked Gene Effects in Mapping Quantitative Trait Loci. Proceedings of the National Academy of Sciences of the United States of America 90:10972-10976.

Zeng Z.B. (1994) Precision Mapping of Quantitative Trait Loci. Genetics 136:1457-1468.

Zhang F., Wan X.Q., Pan G.T. (2006) Genetic analysis of segregation distortion of molecular markers in maize F2 population. J Genet Genomics 32.

Zhang L.Q., Zhang L., Luo J.T., Chen W.J., Hao M., Liu B.L., Yan Z.H., Zhang B., Zhang H.G., Zheng Y.L., Liu D.C., Yen Y. (2011) Synthesizing double haploid hexaploid wheat populations based on a spontaneous alloploidization process. Journal of Genetics and Genomics 38:89-94. DOI: DOI 10.1016/j.jcg.2011.01.004.

Zhang W.K., Wang Y.J., Luo G.Z., Zhang J.S., He C.Y., Wu X.L., Gai J.Y., Chen S.Y. (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theoretical and applied genetics 108:1131-1139.

Zhao X., Kochert G. (1993) Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice (O*ryza sativa* L.). Plant Mol Biol 21:607-14.

Zietkiewicz E., Rafalski A., Labuda D. (1994) Genome Fingerprinting by Simple Sequence Repeat (Ssr)-Anchored Polymerase Chain-Reaction Amplification. Genomics 20:176-183.

# Appendices

**Appendix 1: First set of microsatellite makers developed from Roche 454 pyrosequencing technology used in bambara groundnut.**

| No. | Primer name | Sequence F | Length | Sequence R | Length |
|---|---|---|---|---|---|
| 1 | PRIMER1 | AACTTGCCATACGTGGAAGG | 20 | ACACGCTGCATAATTCACCA | 20 |
| 2 | PRIMER2 | CGTGGATACCCATACCGTCT | 20 | TAAGTCCATTTTGTCCGATTGA | 22 |
| 3 | PRIMER4 | CATTGTCTCTGCCACCATTTT | 21 | CAGACTGGGATTTGCATGTG | 20 |
| 4 | PRIMER5 | CTGCTGTGGTGAGCTTTTGT | 24 | CTCCTTGCAGCTAAGCGTCT | 21 |
| 5 | PRIMER6 | TACGGTCCTACACGGGAAAC | 20 | ACCTGTCCAGCCGCAATTA | 19 |
| 6 | PRIMER7 | GTAGGCCCAACACCACAGTT | 26 | GGAGGTTGATCGATGGAAAA | 20 |
| 7 | PRIMER8 | GGAAGAGTGCGTTTTGGTGT | 22 | CTGTGTGGACCCCAGAAAAT | 20 |
| 8 | PRIMER10 | TCAGTGCTTCAACCATCAGC | 24 | GACCAAACCATTGCCAAACT | 24 |
| 9 | PRIMER15 | AGGAGCAGAAGCTGAAGCAG | 21 | CCAATGCTTTTGAACCAACA | 21 |
| 10 | PRIMER16 | CCGGAACAGAAAACAACAAC | 21 | CGTCGATGACAAAGAGCTTG | 21 |
| 11 | PRIMER19 | AGGCAAAAACGTTTCAGTTC | 20 | TTCATGAAGGTTGAGTTTGTCA | 22 |
| 12 | G11-9-B2-D9 | ATCAAAATCAAGCAAATGAGA | 21 | ACCTTTTACGCTCATTTTAACCAG | 24 |
| 13 | PRIMER23 | CAGTAGCCATAATTTGCTATGAACA | 27 | CGAATCACCATTCAATACGC | 22 |
| 14 | PRIMER26 | CGCTCATTTTAACCAGACCTC | 24 | CAAACAAACCAACGGAATGA | 22 |
| 15 | PRIMER27 | ACACCGCCATCATGAGATTT | 26 | CATTTCAGGATTTGGGAGGA | 25 |
| 16 | PRIMER30 | AATGCAAGATTTTGGCTTGG | 24 | CCCACTCAAACCATACACCA | 23 |
| 17 | PRIMER31 | GCTAAGGTGGAGTGGTGGAA | 26 | CAATCATCTTTTGCGCTTCA | 25 |
| 18 | PRIMER32 | TTCACCTGAACCCCTTAACC | 21 | AGGCTTCACTCACGGGTATG | 20 |
| 19 | PRIMER37 | CCGATGGACGGGTAGATATG | 23 | GCAACCCTCTTTTTCTGCAC | 22 |
| 20 | PRIMER38 | TCACACTTGCAATGGTGCTT | 24 | TCGTTGTTTCTCTTTTCATTGC | 24 |
| 21 | PRIMER42 | TCGTACCGAATCACCATTCA | 23 | CAGTAGCCATAATCTGCTATGAACA | 27 |
| 22 | PRIMER43 | CTTGATGCTACCGAGAGAGAG | 27 | AGGCTCCAACAATGCGATAG | 25 |
| 23 | PRIMER44 | TGTGGGCGAAAATACACAAA | 27 | TCGTCGAATACCTGACTCATTG | 27 |
| 24 | PRIMER45 | CGTGGATACCCATACCGTCT | 22 | AAGTCCATTTTGTCCGATTGA | 23 |
| 25 | PRIMER47 | ACCCATTGCACGTCATAGAA | 21 | GGGTGAACTACACCACCTTCA | 21 |
| 26 | PRIMER48 | TACCTGCATTCGGGCACAGTT | 20 | TTCACTCTTTCTTGATCACATGC | 23 |
| 27 | PRIMER52 | TGCTTACCGCTTGAAGGACT | 20 | GCACCATCATCGACTTCCTT | 20 |
| 28 | PRIMER56(Z) | TGGCATGAGCATCATTCATT | 20 | CAAGGCTATGTCACGGGTTT | 20 |
| 29 | PRIMER63 | TTTAGGAATCAAGAAAGGCAAG | 22 | GCGAGGCAAAAGAAAATGAC | 20 |
| 30 | PRIMER65 | GGACGTGAATCGATGGAGAT | 20 | TCCTTCCCCCTTCTCTGATT | 20 |
| 31 | PRIMER66 | CGTTAGATCTGAGACGCCATT | 21 | CATCCATCACCTGTCACCAG | 20 |
| 32 | PRIMER72 | AACTTGCCATACGTGGAAGG | 20 | ACACGCTGCATAATTCACCA | 20 |
| 33 | PRIMER73 | CACCCTCCAAGACCATCTTC | 20 | CCGAATCCTGATTCTTCCAA | 20 |
| 34 | PRIMER74 | GAGCAGGTGGTGGTGGTT | 18 | GTGAGCTCGTCGTGTTTCTG | 20 |
| 35 | PRIMER81 | CCGGAACAGAAAACAACAAC | 20 | CGTCGATGACAAAGAGCTTG | 20 |
| 36 | PRIMER82 | AATCGTGGTCTCCCATTGAG | 20 | AGGAGCAGAAGCTGAAGCAG | 20 |
| 37 | PRIMER84(Z) | AGAGGAAACCACCATTGCAC | 20 | GCATCACACGAATTTCTCACA | 21 |
| 38 | PRIMER85 | TTTCCAGATTGGATCGTTGA | 20 | TGTCTTCACACCGGAATTTG | 20 |
| 39 | PRIMER88 | TGTGGTTGTGCTCCTTCTCA | 20 | GGGAAGAAGAGTGAAGTTGGAA | 22 |
| 40 | PRIMER91 | ATGGCTGTTAGTGGCGTGAG | 20 | AGACCTGGGAACTTCCACCT | 20 |
| 41 | PRIMER95 | AAGTCCATTTTGTCCGATTGA | 21 | CGTGGATACCCATACCGTCT | 20 |
| 42 | PRIMER96 | TCTTATGCCCTCACCCATTC | 20 | ATTCAGACCAGTGCCCAAAA | 20 |
| 43 | PRIMER98 | TTTTGTCACTGTTTGCCACAA | 21 | AGATTTATATCTGGATGAGAGAGAGAG | 27 |
| 44 | PRIMER103 | AAAATTCAAAGGCCTGGAAAAA | 21 | TTTTTGAGTTCTGCGAGCAA | 20 |
| 45 | PRIMER104 | ACGAGTGTGCGTGTGTGTTT | 20 | GGATCAGGCTCCACATAACC | 20 |
| 46 | PRIMER105 | CAGTGGCAATGAATATCTCAGG | 22 | TGGCTGTTCTCTTGTGAATCC | 21 |
| 47 | Bamcol17 | AACCTGAGAGAAGCGCGTAGAGAA | 24 | GGCTCCCTTCTAAGCAGCAGAACT | 24 |
| 48 | Bamcol24 | ATTGTGCTCAGCTAGGATGTCACG | 24 | GAACTTGGCGAGACTATGCCTCAT | 24 |
| 49 | Bam2col58 | GCGGATGGAAGTTTTGGAAGTGTA | 24 | TTGCTTCCTCCCCCAACATTACTA | 24 |
| 50 | Bam2col63 | AAAATCTCACTCGGATGGCATGTG | 24 | TGGAATCACCTGATAGTAGTGTATTGG | 27 |
| 51 | Bam2col80 | GAGTCCAATAACTGCTCCCGTTTG | 24 | ACGGCAAGCCCTAACTCTTCATTT | 24 |
| 52 | Bam2col13 | CCATCATGAGATTTGCCTTTCCTT | 24 | TAAGCATTTCAGGATTTGGGAGGA | 24 |
| 53 | Bam2col33 | ATGTTCCTTCGTCCTTTTCTCAGC | 24 | AAAACAATCTCTGCCCCAAAAAGA | 24 |
| 54 | mBam3co7 | GGGTTAGTGATAATAAATGGGTGTG | 25 | GTCATAGGAAAGGACCAGTTTCTC | 24 |
| 55 | mBam3co18 | TAGGTTATGAGGTAAGCATTTCAGG | 25 | TGGATATGTGTCTCTATGTTTCCAC | 25 |
| 56 | mBam3co33 | TGTGTCTGTTTGTGGGGATATGTA | 24 | TTATCCCGGTCCTAATTCATCTTA | 24 |
| 57 | mBam3co39 | CAGTAGCCATAATTTGCTATGAACA | 25 | CACATCAATCAAAAATCTCGGTAG | 24 |
| 58 | AG81* | ATTTTCCAACTCGAATTGACC | 21 | TCATCAATCTCGACAAAGAATG | 22 |

*primer from soybean

281

**Appendix 2: Second set of microsatellite makers developed from Roche 454 pyrosequencing technology used in bambara groundnut.**

| No. | Primer name | Sequence F | Length | Sequence R | Length |
|---|---|---|---|---|---|
| 1 | BN1 | CACTTCCGACTCGCTCTGTT | 20 | ggcagtctcagCTCTTGCTT | 20 |
| 2 | BN4 | AGCACCAACTACAGCAGCAA | 20 | GTAATCCCAGGAACTACCACCTC | 23 |
| 3 | BN6a | GAATGTTCCAGACACTACCCTGTT | 24 | GGTTACTCTGTTTGGCTTTGTTAG | 24 |
| 4 | BN6b | CACTACCCTGTTCTTCATCCGT | 22 | CATTGCACGTCATAGAATTTGG | 22 |
| 5 | BN10 | GTTACGATGACCAATATGCTTCC | 23 | CTTGGACTTCGCATAACTTCCTT | 23 |
| 6 | BN11 | aatacctaccgaccaacgacac | 22 | gtccgttcggctagttctca | 20 |
| 7 | BN19 | ATGTCGCACTCTCCTTCACC | 20 | CTTACGCGTGGACTAACCAA | 20 |
| 8 | BN21 | AGTTGCAATGCAGCAACAACAC | 22 | ACGCGTGGACTAACATGAGGA | 21 |
| 9 | BN24 | ATCCCTAGCCAAATGACCTACC | 22 | GCTTTAGGTGTGGCGGATT | 19 |
| 10 | BN28 | ACCTCCTTGGTAATACCTCTCTC | 23 | ggcagtctcagCTCTTGCTTAC | 22 |
| 11 | BN30 | CAAACCAACGGAATGAGTTATG | 22 | ggcagtctcagCTCTTGCTTAC | 22 |
| 12 | BN37 | GGAACAGAAACAGCAACAACC | 21 | acgcgtcgatgacaagaga | 19 |
| 13 | BN39 | CAATGTTTTCACTCAGCAGCA | 21 | tcgttggtaGGAACCGAAAC | 20 |
| 14 | BN41 | GGTTTGCAACCCTCGAATAA | 20 | acggtcgtcgactgacaagag | 21 |
| 15 | BN44 | CATTCACAGCAGCAACACTACTT | 23 | CGTACGTCTACGCGTCTTCTC | 21 |
| 16 | BN53 | CAGGCAGAGTATCAGCAGCA | 20 | ACTAACCCGACCTAGCCCATT | 21 |
| 17 | BN57 | GTTCCAGACACTACCCTGTTCTTC | 24 | ggcagtctcagCTCTTGCTTAC | 22 |
| 18 | BN63 | AGCGAGACGCAAGTTTCATTAC | 22 | ggcagtctcagCTCTTGCTTAC | 22 |
| 19 | BN64a | AGCGAGACGCAAGTTTCATTAC | 22 | GTGGACGTACGCGAGGAG | 18 |
| 20 | BN64b | AGCGAGACGCAAGTTTCATTAC | 22 | gtcgagtctcagCTCTTGTCTTAC | 24 |
| 21 | BN67a | AACAACCTTCACAACTACAACCAC | 24 | CCAGGCCCTGTTACATCAA | 19 |
| 22 | BN67b | CACACCCTTAATGATACAACCA | 22 | TGGACTAACAGATGGAAGTGAA | 22 |
| 23 | BN145 | GGCACTGGTAGCAACGAAA | 19 | CGTGGACGTAACAACACAACAC | 22 |
| 24 | BN151a | GGCACTGGTAGCAACGAAA | 19 | CGTCGGACGTAACAACAACA | 20 |
| 25 | BN151b | ATCGCTTCCACCGTAGAAAGTAG | 23 | TCGGACGTAACAACAACAACAC | 22 |
| 26 | BN164 | ACCGGGAATTCTGGCACT | 18 | cctaccctccttcggtgt | 19 |
| 27 | BN167 | CTGGTAGCAACGAAAAGGGTAA | 23 | TCTACACTCTCTCTTCTGTCGTCTTC | 26 |
| 28 | BN168a | ACATTCACAGCAGCAACACTACTT | 24 | CTCTACACACTTACCTTCCGGTTC | 24 |
| 29 | BN168b | CACTACTTTGCCACATTTTCCA | 23 | TCTACACACTTACCTTCCGGTTC | 23 |
| 30 | BN169 | ATCGCTTCCACCGTAGAAAGTAG | 23 | GTCTCTCGCTCTTCGTCTCGT | 21 |
| 31 | BN206 | GGTTTCCATTCATACAAGCCAT | 22 | GGCAGTCTCAGCTCTTGCTTAC | 22 |
| 32 | BN207 | GATTGGAGGGAATGAATGTGAT | 22 | GGACTACAGTGCTGGTTGCAT | 21 |
| 33 | BN209 | GTCATTGTGAGATCCTTTTCCA | 22 | GTACGCTCGTCTCGTCGTC | 20 |
| 34 | BN249 | CCTCACTTGAACGAAGCTAGGA | 22 | GAAACGAACGGACGAAGAAG | 20 |
| 35 | BN259 | CGATTGCACGTCATAGAATTTG | 22 | GTTCCAGACACTACCCTCGTTC | 22 |
| 36 | BN260 | AGGGAGCGTGGCAATGTAT | 19 | ACGCCTTCGTTCGTTTACC | 19 |
| 37 | BN275 | CTTGCATTTCTATGAGCAGTGG | 23 | CTATCCCCTCGTGTGCCTT | 19 |

**Appendix 3: Microsatellite makers developed from deep sequencing of bambara groundnut.**

| No. | Primer name | Sequence F | Length | Sequence R | Length |
|---|---|---|---|---|---|
| 1 | D.5953 | TCAATTCAAGACTTCCTGCTTTC | 23 | TAGGGTTTTTATTAGGCACTCTGC | 24 |
| 2 | D.15508 | AACTGTATGTAGCAGGGCTTCG | 22 | GGAGATAGAGAGCTACAGAGAATCG | 25 |
| 3 | D.25551 | GCACACAACATAACATAAGGGACAC | 25 | GTTGAAAACCGAACTCATCATTCTC | 25 |
| 4 | D.36186 | GGCGTTTCGATTTCTTCGACTAT | 23 | GAATATGACCTCAGAGCCATTCCT | 24 |
| 5 | D.42026 | TCGTGAAGAATGATTGTCTTTGAT | 24 | TCTATTCAGCTGTTGCTACTCTGG | 24 |
| 6 | D.8148 | CAAACAGTTGGTGTACTTGAATGG | 24 | CAATTAACAAGTTCCATGATGTGC | 24 |
| 7 | D.11587 | TTGAACATTGGAGTTTCTCTTGGT | 24 | AGAAAGGTCCAACACCTTCTCTTTC | 25 |
| 8 | D.8999 | GTGTTATCCGGTATTCGTCAAAAC | 24 | GTCCCTTCAACCTCTCATCTTCC | 23 |
| 9 | D.21950 | ACAGCAGAAACCCTAATTTGTGAG | 24 | AACCAAATTTCACTCTCCTTTTCAC | 25 |
| 10 | D.37053 | AAGCTTCTTAGCCCAGAACAGCTA | 24 | GTCATCCTATTTCTCCCACATCGT | 24 |
| 11 | D.12522 | AGAGGTTGAAGAGGCATAGAGTTG | 24 | ACCACTAACCCAATACAATAGATGG | 25 |
| 12 | D.14265 | GGCGAAGTGATTATTATATAAGATTAGG | 28 | ATCAGAATATCTTCGTTCGAAACTC | 25 |
| 13 | D.15619 | GTGTTCAGGCAAAATCCATGTG | 22 | ACAAAATTTCTTTGGTGGGTCTCT | 24 |
| 14 | D.16501 | TCATAGAGGAAGTTGAGGAACTGG | 24 | TCATCATCATTCATCCAATAACAG | 24 |
| 15 | D. 21310 | GTTTTGATTTCAAAGGGATTATCAC | 25 | ACAACTGAGATTGCAGAAACGC | 22 |
| 16 | D.24269 | AGGTTCATGATCGTAGATGTGGAT | 24 | ACGATATCATACTGACATGTTTCATAC | 27 |
| 17 | D.32937 | GGTGAAAATCAGCAGTACCGAAT | 23 | GTTTCGTCACCATTACCCTCTC | 22 |
| 18 | D.35497 | ACTTTTAGCTCTTGTCAGGAAACG | 24 | TCTTTCTACTTTTCTCTGGCTGGT | 24 |
| 19 | D.48339 | CTGACCCATACATATCCATCATCA | 24 | AGATAGTTCTTTCCACTGCCATGC | 24 |
| 20 | D.51646 | CGGCAATATAATTCAAAATCACTATG | 26 | GTGGAGTTTGAAGAAAACGGATCT | 24 |
| 21 | D.655 | CTTCAATCCCAAAGTCAATTCTG | 24 | CACATTTTCAACTCTGTTTCGTCA | 24 |
| 22 | D.1006 | CTCAGTTTCACCACTCTCCAAAAC | 24 | CATCAGCATTACCAAGAATGACC | 23 |
| 23 | D.7215 | ATCAAGGGTTTGAGGAGTTGAAAT | 24 | GGTGAAGTGAGTGTGGAAAAAGTT | 24 |
| 24 | D.8387 | CAATGTTGTTGTTGTTATGGTGGT | 24 | CTAGTCCATTGTTGTCCCAAAATC | 24 |
| 25 | D.11860 | GAAATCAACATCCAAACACATGAA | 24 | ACAAAATCACACACAAAACACTCG | 24 |
| 26 | D.125 | CCACAAGTACCTAACTCCCAGCTA | 24 | TCTGACTCTACCGTTTTAGGTTGC | 24 |
| 27 | D.1050 | ACAGATGCTTAGAAGATGGAGCTG | 24 | TTTTCAGGCTCTATTTTTGTCCTG | 24 |
| 28 | D.1591 | TTGCAGAAAAATTACAGAAACCAG | 24 | TTATTTACTCGTTTCCCCAATCAA | 24 |
| 29 | D.2094 | ACTCCTGGAAGACGAGTCTGTG | 22 | GTTCACACATGGTGCTAAATTGAA | 24 |

**Appendix 4: Normality plot for *leaf area* in the F$_3$ population of the field data.**



Field trial

Average: 672.182
StDev: 329.697
N: 70

Anderson-Darling Normality Test
A-Squared: 0.326
P-Value: 0.514
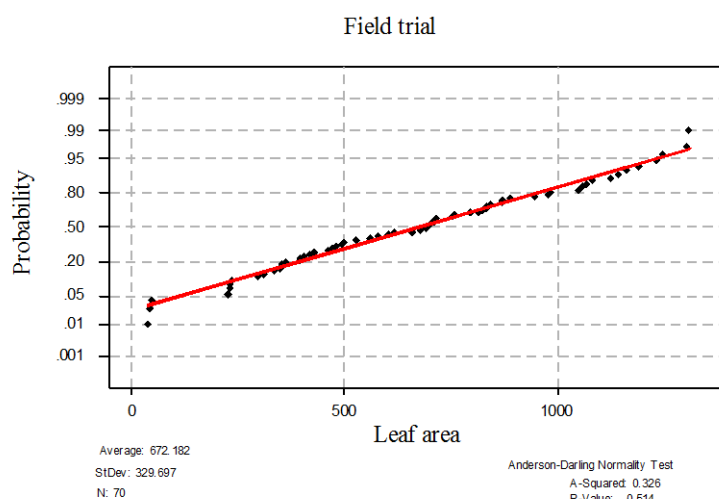
**Appendix 5: Histogram of normal distribution for *internode length* in the glasshouse.**
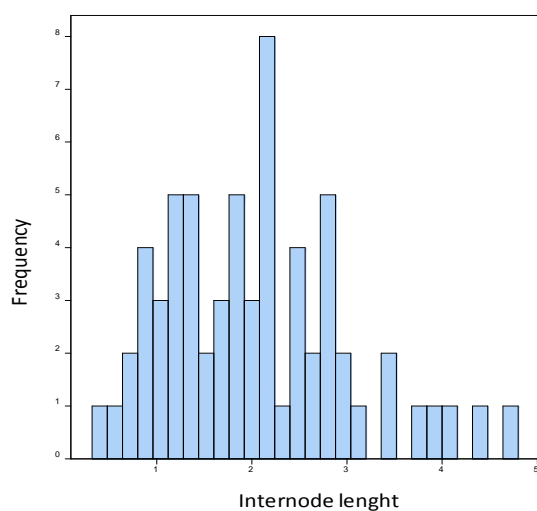


**Appendix 6: Histogram of normal distribution for *internode length* in the field.**
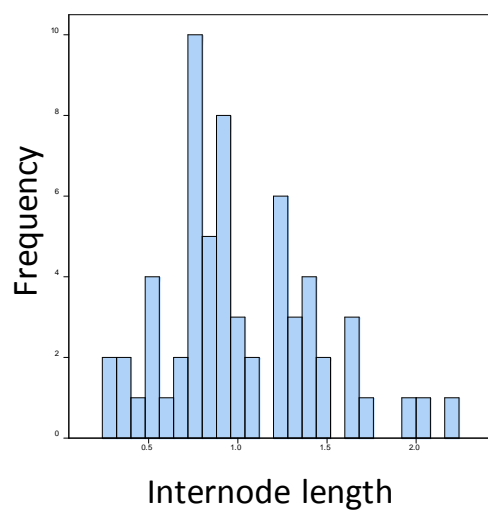


**Appendix 7: Histogram of non-normal distribution for *growth habit* in the glasshouse.**
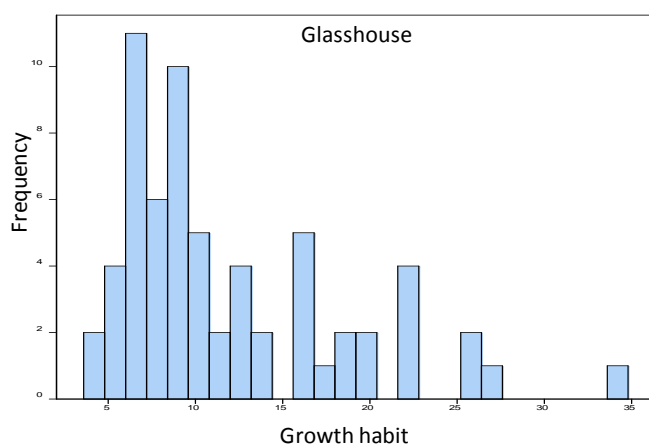
**Appendix 8: Histogram of non-normal distribution of *growth habit* in the Field.**
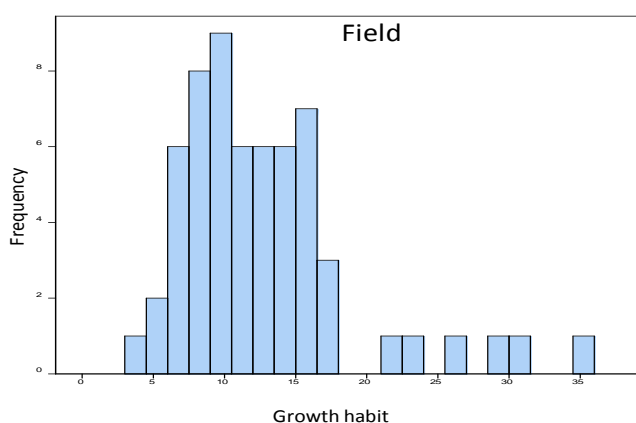


**Appendix 9: Normality plot for *pod no./plant* in the F$_3$ population of the glasshouse.**
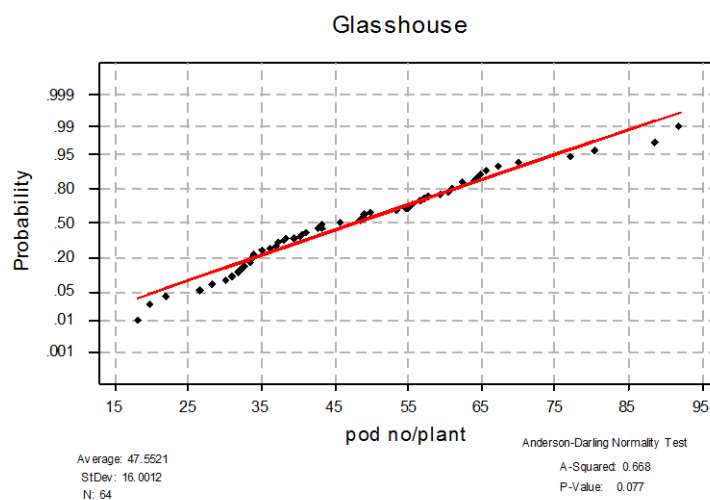


**Appendix 10: Normality plot for *pod no./plant* (transformed) in the F$_3$ population of the field.**
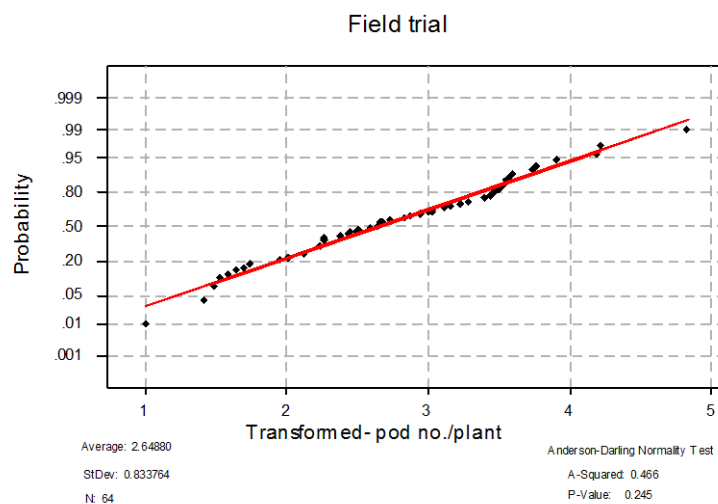
**Appendix 11: Normality plot for *double seeded pods/plant* (transformed) in the F$_3$ population of the glasshouse.**
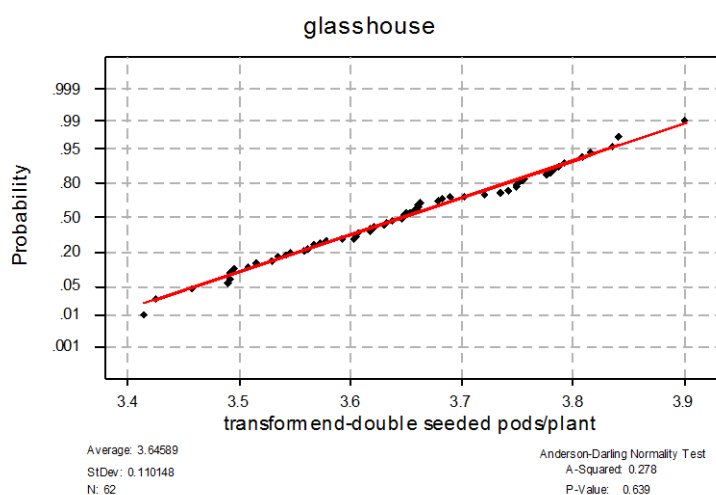


**Appendix 12: Normality plot for *seed weight* (transformed) in the F$_3$ population of the glasshouse.**
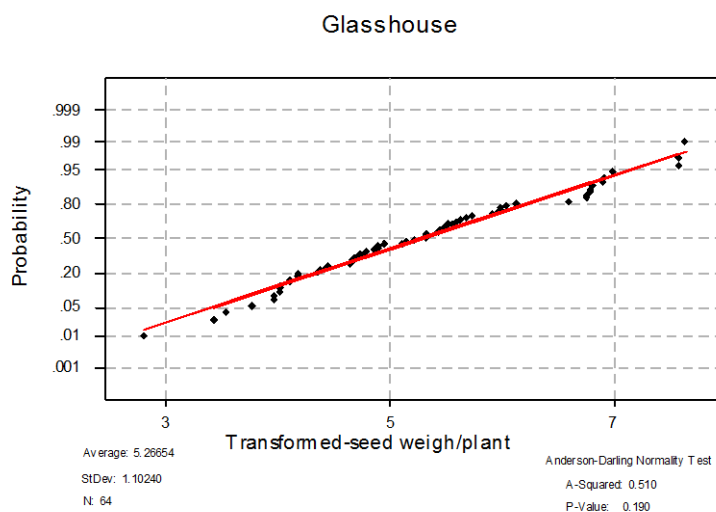
**Appendix 13: Histogram of normal distribution of the *length* and *width* of *terminal leaflet* for F$_2$ progenies data in TCRU.**



F$_2$ data in TCRU
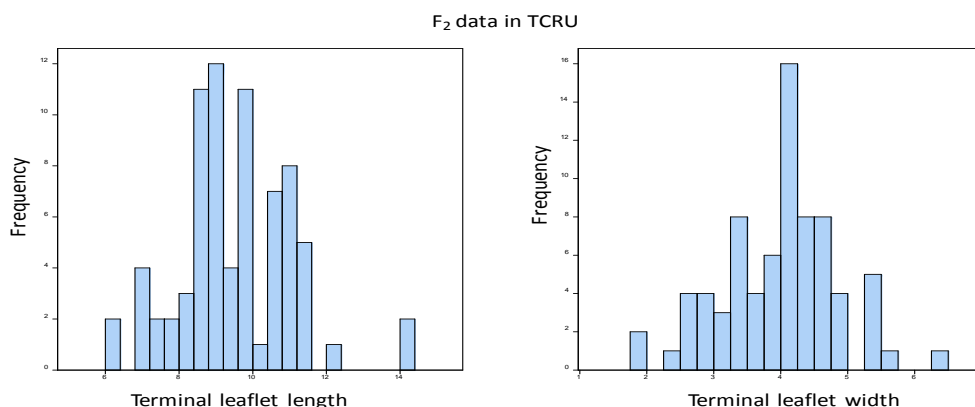
**Appendix 14: Normality plot for *seed no./plant* (transformed) in the F$_2$ population grown in TCRU.**



TCRU

Average: 10.1350
StDev: 3.58155
N: 73

Anderson-Darling Normality Test
A-Squared: 0.470
P-Value: 0.240
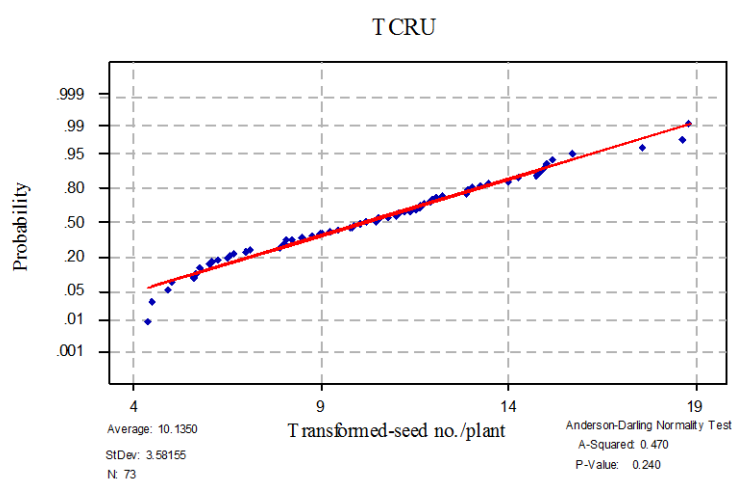
**Appendix 15: A fitted line regression plot of *flower no./plat* with both *plant spread* and *biomass dry weight* in F$_3$ population of glasshouse.**



Flowerno = 0.0467156 + 3.12495 plant spread
S = 28.3851    R-Sq = 33.2 %    R-Sq(adj) = 32.1 %

Flower no. = 43.3019 + 1.00936 biomass dry weight
S = 25.7613    R-Sq = 45.0 %    R-Sq(adj) = 44.1 %

287
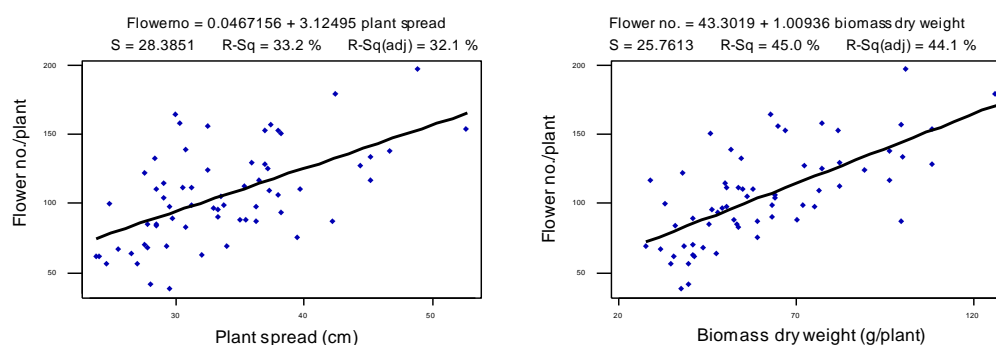
**Appendix 16: A fitted line regression plot of *plant spread* with both *growth habit* and *biomass dry weight* in F$_3$ population of glasshouse.**

plant spread = 40.8135 - 0.595520 growth habit

S = 5.05183      R-Sq = 37.8 %      R-Sq(adj) = 36.8 %

plant spread = 20.8257 + 0.209137 biomass dry weight

S = 4.21055      R-Sq = 56.8 %      R-Sq(adj) = 56.1 %
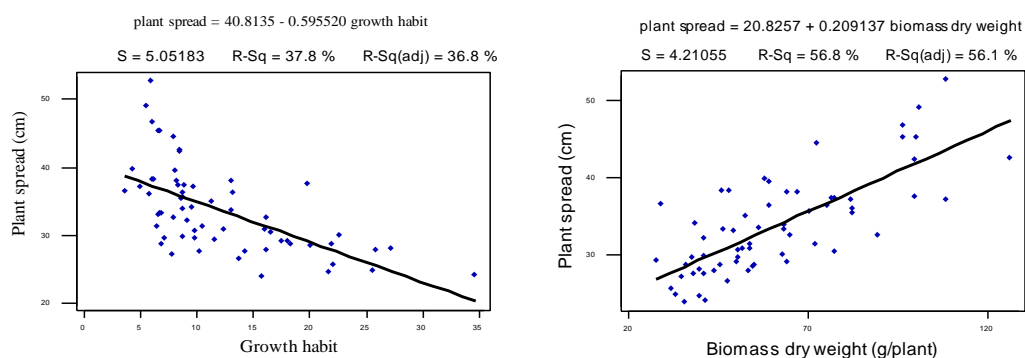
**Appendix 17: A fitted line regression plot of pod no./plant with both of node no./stem and seed weight in F$_3$ population of glasshouse.**

Podno./plant = 1.75814 + 5.28875 node no./stem

S = 11.1867      R-Sq = 51.9 %      R-Sq(adj) = 51.1 %

Podno./plant = 13.6861 + 1.17051 seed weight/plant

S = 7.64473      R-Sq = 77.5 %      R-Sq(adj) = 77.2 %
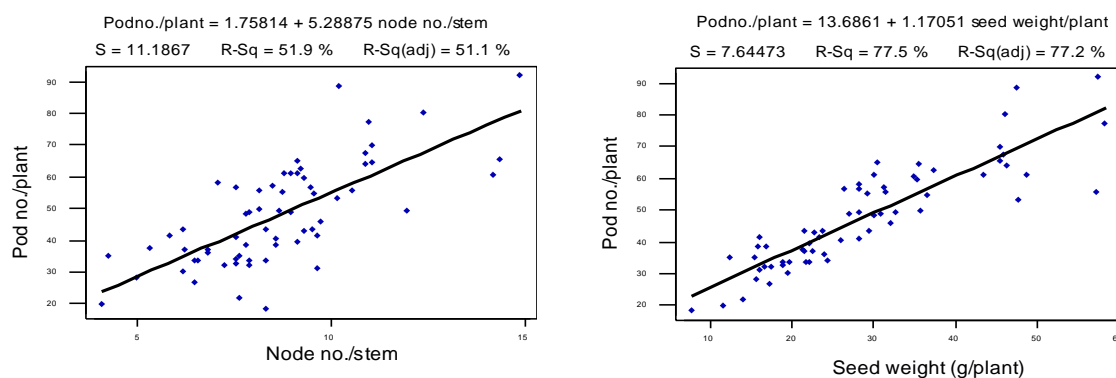
**Appendix 18: A fitted line regression plot for *plant height* with *biomass dry weight* in the F$_3$ population of the Field.**

Plant height = 8.55713 + 0.753757 biomass dry weight

S = 2.18371      R-Sq = 57.4 %      R-Sq(adj) = 56.7 %

288
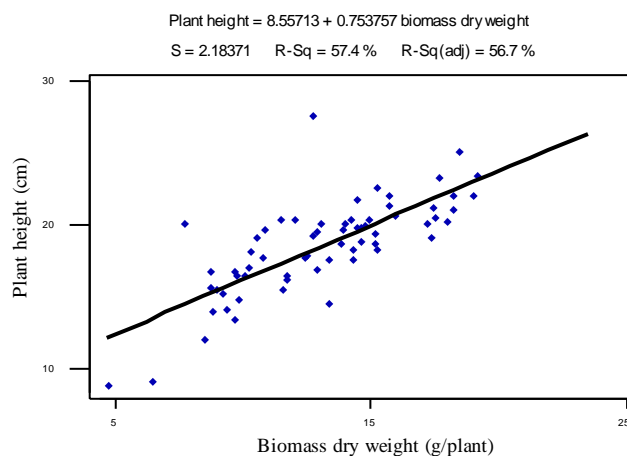
**Appendix 19: A fitted line regression plot for *internode length* with *growth habit* and *seed weight* in the F$_3$ population of the field.**



Internode length = 1.70899 - 0.0531969 growth habit

S = 0.257562     R-Sq = 63.5 %     R-Sq(adj) = 62.9 %

Internode length = 0.756156 + 0.0867348 seed weight

S = 0.379230     R-Sq = 15.4 %     R-Sq(adj) = 13.8 %
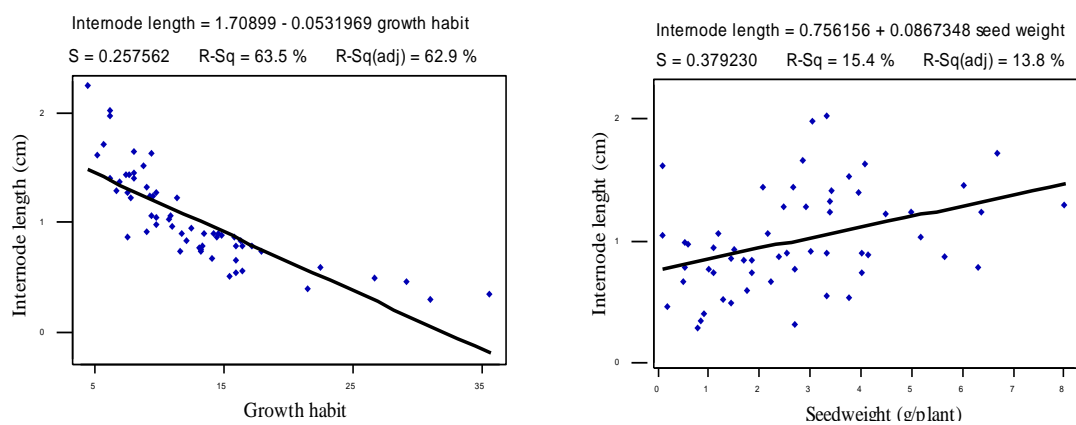
**Appendix 20: A fitted line regression plot for *seed weight* with *pod weight* in the F$_3$ population of the Field.**



Regression Plot

Seed weight = 0.0063412 + 0.650766 pod weight

S = 0.837295     R-Sq = 79.9 %     R-Sq(adj) = 79.6 %

289
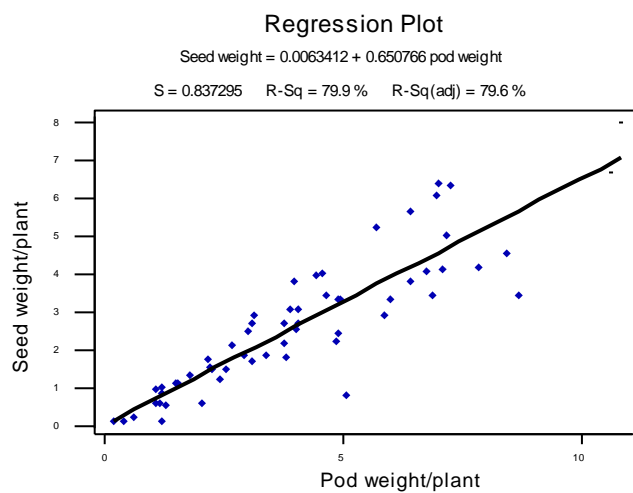
**Appendix 21: A fitted line regression plot for *plant spread* with *biomass dry weight* in the F$_2$ population of narrow cross.**

Biomass dry weight = -140.088 + 4.26199 plant spread

S = 39.8138    R-Sq = 65.1 %    R-Sq(adj) = 64.6 %
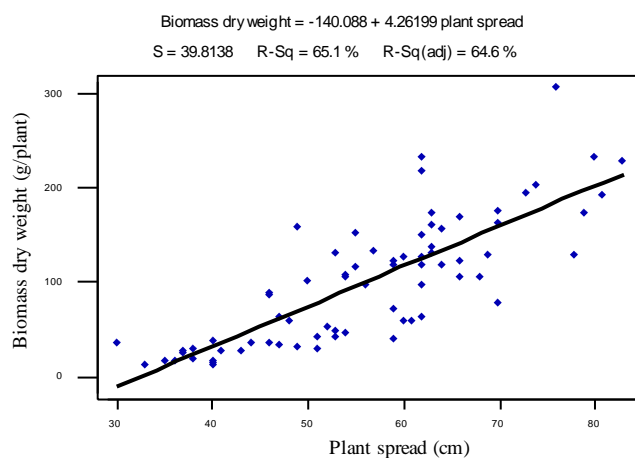


**Appendix 22: A fitted line regression plot for *growth habit* with *plant spread* in the F$_2$ population of narrow cross.**

Seed weight = -58.0990 + 1.74037 Plant spread

S = 17.3631    R-Sq = 62.1 %    R-Sq(adj) = 61.5 %



290
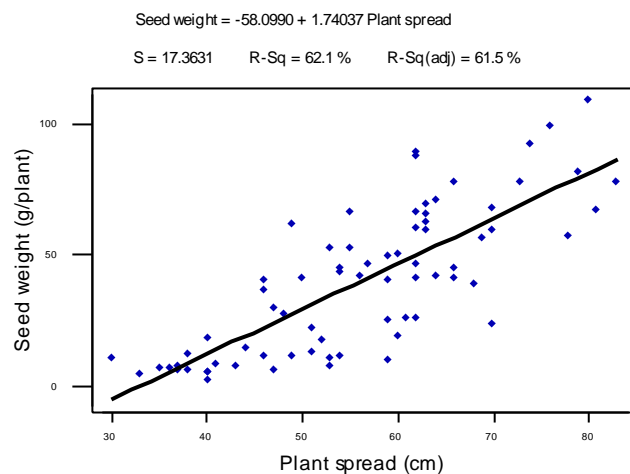
**Appendix 23: The locus genotype frequency for narrow cross map after phase detection.**

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| bgPabg-594562 | 25 | 0 | 0 | 42 | 0 | 6 | 0 | 1 | - | [a:h+b+c] |
| bgPabg-595666 | 25 | 0 | 0 | 42 | 0 | 6 | 0 | 1 | - | [a:h+b+c] |
| bgPt-594335 | 25 | 0 | 0 | 41 | 0 | 7 | 0 | 1 | - | [a:h+b+c] |
| bgPt-597197 | 22 | 0 | 0 | 37 | 0 | 14 | 0 | 1 | - | [a:h+b+c] |
| bgPt-602174 | 22 | 0 | 0 | 37 | 0 | 14 | 0 | 1 | - | [a:h+b+c] |
| bgPt-598456 | 24 | 0 | 0 | 41 | 0 | 8 | 0.01 | 1 | - | [a:h+b+c] |
| bgPt-598510 | 21 | 0 | 0 | 36 | 0 | 16 | 0.01 | 1 | - | [a:h+b+c] |
| bgPt-598594 | 21 | 0 | 0 | 36 | 0 | 16 | 0.01 | 1 | - | [a:h+b+c] |
| bgPt-600863 | 21 | 0 | 0 | 36 | 0 | 16 | 0.01 | 1 | - | [a:h+b+c] |
| bgPabg-595315 | 25 | 0 | 0 | 43 | 0 | 5 | 0.02 | 1 | - | [a:h+b+c] |
| bgPabg-597801 | 26 | 0 | 0 | 42 | 0 | 5 | 0.02 | 1 | - | [a:h+b+c] |
| bgPabg-594305 | 23 | 0 | 0 | 40 | 0 | 10 | 0.03 | 1 | - | [a:h+b+c] |
| bgPabg-596075 | 27 | 0 | 0 | 43 | 0 | 3 | 0.03 | 1 | - | [a:h+b+c] |
| bgPt-596460 | 23 | 0 | 0 | 40 | 0 | 10 | 0.03 | 1 | - | [a:h+b+c] |
| bgPt-601402 | 22 | 0 | 0 | 35 | 0 | 16 | 0.03 | 1 | - | [a:h+b+c] |
| bgPabg-423257 | 24 | 0 | 0 | 42 | 0 | 7 | 0.04 | 1 | - | [a:h+b+c] |
| bgPt-601656 | 0 | 0 | 21 | 0 | 33 | 19 | 0.04 | 1 | - | [a+h+d:b] |
| bgPabg-593892 | 25 | 0 | 0 | 44 | 0 | 4 | 0.05 | 1 | - | [a:h+b+c] |
| bgPabg-595273 | 25 | 0 | 0 | 44 | 0 | 4 | 0.05 | 1 | - | [a:h+b+c] |
| bgPt-594531 | 20 | 0 | 0 | 31 | 0 | 2 | 0.06 | 1 | - | [a:h+b+c] |
| bgPabg-597632 | 23 | 0 | 0 | 41 | 0 | 9 | 0.07 | 1 | - | [a:h+b+c] |
| bgPabg-601173 | 23 | 0 | 0 | 41 | 0 | 9 | 0.07 | 1 | - | [a:h+b+c] |
| bgPt-593893 | 23 | 0 | 0 | 41 | 0 | 9 | 0.07 | 1 | - | [a:h+b+c] |
| bgPt-598647 | 0 | 0 | 25 | 0 | 39 | 9 | 0.07 | 1 | - | [a+h+d:b] |
| bgPabg-422614 | 27 | 0 | 0 | 42 | 0 | 4 | 0.08 | 1 | - | [a:h+b+c] |
| DSRB18 | 28 | 17 | 27 | 0 | 0 | 1 | 0.09 | 2 | - | [a:h:b] |
| bgPt-598428 | 24 | 0 | 0 | 37 | 0 | 12 | 0.09 | 1 | - | [a:h+b+c] |
| bgPabg-422333 | 26 | 0 | 0 | 40 | 0 | 7 | 0.1 | 1 | - | [a:h+b+c] |
| bgPt-596097 | 22 | 0 | 0 | 40 | 0 | 11 | 0.11 | 1 | - | [a:h+b+c] |
| bgPt-596596 | 25 | 0 | 0 | 38 | 0 | 10 | 0.13 | 1 | - | [a:h+b+c] |
| bgPabg-594981 | 27 | 0 | 0 | 41 | 0 | 5 | 0.14 | 1 | - | [a:h+b+c] |
| bgPabg-596210 | 27 | 0 | 0 | 41 | 0 | 5 | 0.14 | 1 | - | [a:h+b+c] |
| bgPabg-596646 | 27 | 0 | 0 | 41 | 0 | 5 | 0.14 | 1 | - | [a:h+b+c] |
| bgPabg-597746 | 27 | 0 | 0 | 41 | 0 | 5 | 0.14 | 1 | - | [a:h+b+c] |
| bgPt-600898 | 0 | 0 | 20 | 0 | 37 | 16 | 0.14 | 1 | - | [a+h+d:b] |
| bgPabg-594814 | 28 | 0 | 0 | 42 | 0 | 3 | 0.19 | 1 | - | [a:h+b+c] |
| bgPabg-595707 | 28 | 0 | 0 | 42 | 0 | 3 | 0.19 | 1 | - | [a:h+b+c] |
| bgPt-594664 | 23 | 0 | 0 | 43 | 0 | 7 | 0.2 | 1 | - | [a:h+b+c] |
| bgPt-601022 | 23 | 0 | 0 | 34 | 0 | 16 | 0.2 | 1 | - | [a:h+b+c] |
| bgPt-601091 | 19 | 0 | 0 | 36 | 0 | 1 | 0.2 | 1 | - | [a:h+b+c] |
| bgPabg-595725 | 24 | 0 | 0 | 45 | 0 | 4 | 0.22 | 1 | - | [a:h+b+c] |
| bgPabg-597705 | 27 | 0 | 0 | 40 | 0 | 6 | 0.22 | 1 | - | [a:h+b+c] |
| bgPabg-597860 | 27 | 0 | 0 | 40 | 0 | 6 | 0.22 | 1 | - | [a:h+b+c] |
| bgPt-593939 | 27 | 0 | 0 | 40 | 0 | 6 | 0.22 | 1 | - | [a:h+b+c] |
| bgPabg-594999 | 0 | 0 | 24 | 0 | 45 | 4 | 0.22 | 1 | - | [a+h+d:b] |
| bgPabg-594261 | 0 | 0 | 24 | 0 | 35 | 14 | 0.25 | 1 | - | [a+h+d:b] |
| bgPt-596392 | 26 | 0 | 0 | 38 | 0 | 9 | 0.27 | 1 | - | [a:h+b+c] |
| bgPt-600935 | 26 | 0 | 0 | 38 | 0 | 9 | 0.27 | 1 | - | [a:h+b+c] |
| bgPabg-595685 | 0 | 0 | 22 | 0 | 42 | 9 | 0.27 | 1 | - | [a+h+d:b] |
| bgPabg-597086 | 28 | 0 | 0 | 41 | 0 | 4 | 0.28 | 1 | - | [a:h+b+c] |
| bgPt-598511 | 18 | 0 | 0 | 35 | 0 | 2 | 0.28 | 1 | - | [a:h+b+c] |
| bgPt-597279 | 0 | 0 | 19 | 0 | 37 | 17 | 0.3 | 1 | - | [a+h+d:b] |
| bgPabg-597623 | 20 | 0 | 0 | 39 | 0 | 14 | 0.33 | 1 | - | [a:h+b+c] |
| bgPt-598641 | 20 | 0 | 0 | 39 | 0 | 14 | 0.33 | 1 | - | [a:h+b+c] |
| bgPt-600823 | 0 | 0 | 20 | 0 | 39 | 14 | 0.33 | 1 | - | [a+h+d:b] |
| bgPabg-598428 | 22 | 0 | 0 | 43 | 0 | 8 | 0.37 | 1 | - | [a:h+b+c] |
| PRIMER10 | 25 | 20 | 28 | 0 | 0 | 0 | 0.39 | 2 | - | [a:h:b] |
| bgPabg-595682 | 0 | 0 | 24 | 0 | 47 | 2 | 0.41 | 1 | - | [a+h+d:b] |

Appendix 23 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| bgPt-601748 | 19 | 0 | 0 | 38 | 0 | 16 | 0.42 | 1 | - | [a:h+b+c] |
| bgPt-602039 | 0 | 0 | 19 | 0 | 38 | 16 | 0.42 | 1 | - | [a+h+d:b] |
| bgPt-596705 | 21 | 0 | 0 | 29 | 0 | 23 | 0.43 | 1 | - | [a:h+b+c] |
| bgPabg-594467 | 27 | 0 | 0 | 38 | 0 | 8 | 0.45 | 1 | - | [a:h+b+c] |
| bgPabg-596244 | 0 | 0 | 27 | 0 | 38 | 8 | 0.45 | 1 | - | [a+h+d:b] |
| bgPabg-597436 | 22 | 0 | 0 | 44 | 0 | 7 | 0.49 | 1 | - | [a:h+b+c] |
| bgPabg-597991 | 26 | 0 | 0 | 36 | 0 | 11 | 0.52 | 1 | - | [a:h+b+c] |
| bgPt-597945 | 0 | 0 | 24 | 0 | 33 | 16 | 0.52 | 1 | - | [a+h+d:b] |
| bgPt-598364 | 18 | 0 | 0 | 37 | 0 | 18 | 0.53 | 1 | - | [a:h+b+c] |
| bgPabg-595854 | 29 | 0 | 0 | 40 | 0 | 4 | 0.6 | 1 | - | [a:h+b+c] |
| bgPabg-596250 | 0 | 0 | 23 | 0 | 47 | 3 | 0.64 | 1 | - | [a+h+d:b] |
| bgPt-423805 | 0 | 0 | 17 | 0 | 36 | 20 | 0.67 | 1 | - | [a+h+d:b] |
| PRIMER66 | 24 | 20 | 29 | 0 | 0 | 0 | 0.68 | 2 | - | [a:h:b] |
| bgPt-598651 | 26 | 0 | 0 | 35 | 0 | 12 | 0.68 | 1 | - | [a:h+b+c] |
| bgPt-602268 | 26 | 0 | 0 | 35 | 0 | 12 | 0.68 | 1 | - | [a:h+b+c] |
| bgPt-597632 | 18 | 0 | 0 | 38 | 0 | 17 | 0.69 | 1 | - | [a:h+b+c] |
| bgPt-602071 | 22 | 0 | 0 | 29 | 0 | 22 | 0.69 | 1 | - | [a:h+b+c] |
| bgPabg-423248 | 0 | 0 | 24 | 0 | 32 | 17 | 0.69 | 1 | - | [a+h+d:b] |
| bgPt-596678 | 0 | 0 | 24 | 0 | 32 | 17 | 0.69 | 1 | - | [a+h+d:b] |
| bgPt-601384 | 20 | 0 | 0 | 42 | 0 | 11 | 0.73 | 1 | - | [a:h+b+c] |
| bgPabg-598184 | 21 | 0 | 0 | 44 | 0 | 8 | 0.75 | 1 | - | [a:h+b+c] |
| Bam2coL63 | 25 | 17 | 31 | 0 | 0 | 0 | 0.77 | 2 | - | [a:h:b] |
| PRIMER103 | 29 | 19 | 23 | 0 | 0 | 2 | 0.79 | 2 | - | [a:h:b] |
| bgPt-598122 | 17 | 0 | 0 | 37 | 0 | 19 | 0.83 | 1 | - | [a:h+b+c] |
| PRIMER88 | 28 | 21 | 24 | 0 | 0 | 0 | 0.84 | 2 | - | [a:h:b] |
| bgPt-597446 | 18 | 0 | 0 | 39 | 0 | 16 | 0.85 | 1 | - | [a:h+b+c] |
| bgPt-422660 | 0 | 0 | 24 | 0 | 31 | 18 | 0.88 | 1 | - | [a+h+d:b] |
| bgPabg-422621 | 20 | 0 | 0 | 43 | 0 | 10 | 0.89 | 1 | - | [a:h+b+c] |
| bgPabg-594142 | 21 | 0 | 0 | 45 | 0 | 7 | 0.91 | 1 | - | [a:h+b+c] |
| bgPt-423432 | 0 | 0 | 25 | 0 | 32 | 16 | 0.98 | 1 | - | [a+h+d:b] |
| bgPt-598804 | 17 | 0 | 0 | 38 | 0 | 18 | 1.02 | 1 | - | [a:h+b+c] |
| bgPabg-596040 | 30 | 0 | 0 | 39 | 0 | 4 | 1.05 | 1 | - | [a:h+b+c] |
| bgPt-595486 | 28 | 0 | 0 | 36 | 0 | 9 | 1.07 | 1 | - | [a:h+b+c] |
| bgPt-597731 | 20 | 0 | 0 | 44 | 0 | 9 | 1.07 | 1 | - | [a:h+b+c] |
| bgPabg-423332 | 0 | 0 | 20 | 0 | 44 | 9 | 1.07 | 1 | - | [a+h+d:b] |
| bgPabg-596546 | 0 | 0 | 28 | 0 | 36 | 9 | 1.07 | 1 | - | [a+h+d:b] |
| bgPabg-596618 | 21 | 0 | 0 | 46 | 0 | 6 | 1.08 | 1 | - | [a:h+b+c] |
| bgPabg-598364 | 26 | 0 | 0 | 33 | 0 | 14 | 1.09 | 1 | - | [a:h+b+c] |
| bgPt-598084 | 0 | 0 | 26 | 0 | 33 | 14 | 1.09 | 1 | - | [a+h+d:b] |
| bgPt-597858 | 29 | 0 | 0 | 37 | 0 | 7 | 1.17 | 1 | - | [a:h+b+c] |
| bgPabg-422590 | 0 | 0 | 29 | 0 | 37 | 7 | 1.17 | 1 | - | [a+h+d:b] |
| bgPt-422567 | 27 | 0 | 0 | 34 | 0 | 12 | 1.19 | 1 | - | [a:h+b+c] |
| GH-19-B2-D9 | 23 | 19 | 31 | 0 | 0 | 0 | 1.21 | 2 | - | [a:h:b] |
| bgPt-598767 | 25 | 0 | 0 | 31 | 0 | 17 | 1.22 | 1 | - | [a:h+b+c] |
| bgPabg-597446 | 19 | 0 | 0 | 43 | 0 | 11 | 1.24 | 1 | - | [a:h+b+c] |
| bgPabg-593983 | 0 | 0 | 30 | 0 | 38 | 5 | 1.27 | 1 | - | [a+h+d:b] |
| bgPabg-596774 | 0 | 0 | 30 | 0 | 38 | 5 | 1.27 | 1 | - | [a+h+d:b] |
| bgPabg-597624 | 0 | 0 | 30 | 0 | 38 | 5 | 1.27 | 1 | - | [a+h+d:b] |
| bgPabg-423067 | 0 | 0 | 28 | 0 | 35 | 10 | 1.3 | 1 | - | [a+h+d:b] |
| bgPabg-422473 | 0 | 0 | 29 | 0 | 36 | 8 | 1.4 | 1 | - | [a+h+d:b] |
| PRIMER48 | 32 | 15 | 26 | 0 | 0 | 0 | 1.43 | 2 | - | [a:h:b] |
| bgPt-596563 | 17 | 0 | 0 | 40 | 0 | 16 | 1.43 | 1 | - | [a:h+b+c] |
| bgPt-597606 | 16 | 0 | 0 | 38 | 0 | 19 | 1.43 | 1 | - | [a:h+b+c] |
| bgPt-601098 | 16 | 0 | 0 | 38 | 0 | 19 | 1.43 | 1 | - | [a:h+b+c] |
| bgPt-596916 | 19 | 0 | 0 | 44 | 0 | 10 | 1.45 | 1 | - | [a:h+b+c] |
| bgPt-598621 | 19 | 0 | 0 | 44 | 0 | 10 | 1.45 | 1 | - | [a:h+b+c] |
| bgPt-594453 | 20 | 0 | 0 | 46 | 0 | 7 | 1.46 | 1 | - | [a:h+b+c] |
| bgPt-596188 | 20 | 0 | 0 | 46 | 0 | 7 | 1.46 | 1 | - | [a:h+b+c] |
| bgPt-596726 | 20 | 0 | 0 | 46 | 0 | 7 | 1.46 | 1 | - | [a:h+b+c] |
| PRIMER15 | 28 | 22 | 23 | 0 | 0 | 0 | 1.48 | 2 | - | [a:h:b] |
| bgPabg-595822 | 30 | 0 | 0 | 37 | 0 | 6 | 1.51 | 1 | - | [a:h+b+c] |

Appendix 23 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| bgPabg-594335 | 31 | 0 | 0 | 38 | 0 | 4 | 1.62 | 1 | - | [a:h+b+c] |
| bgPabg-597858 | 31 | 0 | 0 | 38 | 0 | 4 | 1.62 | 1 | - | [a:h+b+c] |
| bgPabg-598611 | 20 | 0 | 0 | 47 | 0 | 6 | 1.67 | 1 | - | [a:h+b+c] |
| bgPt-595093 | 19 | 0 | 0 | 45 | 0 | 9 | 1.67 | 1 | - | [a:h+b+c] |
| bgPt-597667 | 19 | 0 | 0 | 45 | 0 | 9 | 1.67 | 1 | - | [a:h+b+c] |
| bgPt-598184 | 14 | 0 | 0 | 35 | 0 | 24 | 1.67 | 1 | - | [a:h+b+c] |
| bgPt-600818 | 19 | 0 | 0 | 45 | 0 | 9 | 1.67 | 1 | - | [a:h+b+c] |
| bgPabg-596988 | 21 | 0 | 0 | 49 | 0 | 3 | 1.68 | 1 | - | [a:h+b+c] |
| mBam3co7 | 22 | 21 | 30 | 0 | 0 | 0 | 1.72 | 2 | - | [a:h:b] |
| PRIMER32 | 32 | 14 | 27 | 0 | 0 | 0 | 1.78 | 2 | - | [a:h:b] |
| bgPt-595129 | 28 | 0 | 0 | 33 | 0 | 12 | 1.84 | 1 | - | [a:h+b+c] |
| BN6b | 22 | 19 | 32 | 0 | 0 | 0 | 1.87 | 2 | - | [a:h:b] |
| bgPabg-594355 | 20 | 0 | 0 | 48 | 0 | 5 | 1.9 | 1 | - | [a:h+b+c] |
| bgPt-600827 | 18 | 0 | 0 | 44 | 0 | 11 | 1.9 | 1 | - | [a:h+b+c] |
| bgPabg-596335 | 0 | 0 | 31 | 0 | 37 | 5 | 1.9 | 1 | - | [a+h+d:b] |
| bgPt-595387 | 0 | 0 | 26 | 0 | 30 | 17 | 1.9 | 1 | - | [a+h+d:b] |
| bgPt-597573 | 0 | 0 | 18 | 0 | 44 | 11 | 1.9 | 1 | - | [a+h+d:b] |
| PRIMER43 | 25 | 15 | 33 | 0 | 0 | 0 | 1.94 | 2 | - | [a:h:b] |
| bgPabg-423420 | 0 | 0 | 29 | 0 | 34 | 10 | 1.96 | 1 | - | [a+h+d:b] |
| bgPabg-422461 | 0 | 0 | 30 | 0 | 35 | 8 | 2.08 | 1 | - | [a+h+d:b] |
| bgPabg-423053 | 0 | 0 | 30 | 0 | 35 | 8 | 2.08 | 1 | - | [a+h+d:b] |
| bgPt-601853 | 0 | 0 | 17 | 0 | 43 | 13 | 2.15 | 1 | - | [a+h+d:b] |
| PRIMER26 | 24 | 15 | 33 | 0 | 0 | 1 | 2.17 | 2 | - | [a:h:b] |
| bgPt-598507 | 15 | 0 | 0 | 39 | 0 | 19 | 2.18 | 1 | - | [a:h+b+c] |
| bgPabg-423334 | 0 | 0 | 31 | 0 | 36 | 6 | 2.2 | 1 | - | [a+h+d:b] |
| bgPt-597491 | 26 | 0 | 0 | 29 | 0 | 18 | 2.24 | 1 | - | [a:h+b+c] |
| bgPabg-422379 | 0 | 0 | 29 | 0 | 33 | 11 | 2.28 | 1 | - | [a+h+d:b] |
| bgPabg-422586 | 0 | 0 | 29 | 0 | 33 | 11 | 2.28 | 1 | - | [a+h+d:b] |
| bgPabg-596678 | 0 | 0 | 29 | 0 | 33 | 11 | 2.28 | 1 | - | [a+h+d:b] |
| BamcoL17 | 31 | 21 | 21 | 0 | 0 | 0 | 2.38 | 2 | - | [a:h:b] |
| bgPt-596444 | 30 | 0 | 0 | 34 | 0 | 9 | 2.4 | 1 | - | [a:h+b+c] |
| bgPt-597962 | 17 | 0 | 0 | 44 | 0 | 12 | 2.41 | 1 | - | [a:h+b+c] |
| bgPt-598611 | 17 | 0 | 0 | 44 | 0 | 12 | 2.41 | 1 | - | [a:h+b+c] |
| PRIMER7 | 24 | 24 | 25 | 0 | 0 | 0 | 2.43 | 2 | - | [a:h:b] |
| bgPabg-593965 | 0 | 0 | 33 | 0 | 38 | 2 | 2.44 | 1 | - | [a+h+d:b] |
| BN259 | 23 | 16 | 34 | 0 | 0 | 0 | 2.58 | 2 | - | [a:h:b] |
| bgPabg-594537 | 29 | 0 | 0 | 32 | 0 | 12 | 2.62 | 1 | - | [a:h+b+c] |
| bgPabg-594877 | 19 | 0 | 0 | 49 | 0 | 5 | 2.65 | 1 | - | [a:h+b+c] |
| bgPt-594877 | 18 | 0 | 0 | 47 | 0 | 8 | 2.67 | 1 | - | [a:h+b+c] |
| PRIMER16 | 27 | 12 | 32 | 0 | 0 | 2 | 2.95 | 2 | - | [a:h:b] |
| PRIMER38 | 30 | 23 | 20 | 0 | 0 | 0 | 3.47 | 2 | - | [a:h:b] |
| PRIMER19 | 19 | 21 | 33 | 0 | 0 | 0 | 4.13 | 2 | - | [a:h:b] |
| PRIMER95 | 19 | 23 | 31 | 0 | 0 | 0 | 4.28 | 2 | - | [a:h:b] |
| mBam3co33 | 21 | 16 | 36 | 0 | 0 | 0 | 4.48 | 2 | - | [a:h:b] |
| bgPabg-423185 | 0 | 0 | 40 | 0 | 24 | 9 | 17.07 | 1 | ******* | [a+h+d:b] |
| bgPt-596933 | 0 | 0 | 48 | 0 | 10 | 15 | 50.69 | 1 | ******* | [a+h+d:b] |
| bgPabg-596933 | 0 | 0 | 54 | 0 | 12 | 7 | 55.31 | 1 | ******* | [a+h+d:b] |
| bgPt-423050 | 0 | 0 | 49 | 0 | 8 | 16 | 57.12 | 1 | ******* | [a+h+d:b] |
| bgPt-601086 | 0 | 0 | 44 | 0 | 4 | 25 | 60.09 | 1 | ******* | [a+h+d:b] |
| bgPabg-594508 | 0 | 0 | 58 | 0 | 12 | 3 | 61.44 | 1 | ******* | [a+h+d:b] |
| bgPabg-423807 | 0 | 0 | 54 | 0 | 9 | 10 | 62.49 | 1 | ******* | [a+h+d:b] |
| bgPabg-596655 | 0 | 0 | 53 | 0 | 8 | 12 | 63.48 | 1 | ******* | [a+h+d:b] |
| bgPabg-601086 | 0 | 0 | 53 | 0 | 7 | 13 | 66.15 | 1 | ******* | [a+h+d:b] |
| bgPabg-595038 | 0 | 0 | 60 | 0 | 10 | 3 | 69.43 | 1 | ******* | [a+h+d:b] |
| bgPt-596221 | 39 | 0 | 0 | 27 | 0 | 7 | 13.13 | 1 | ****** | [a:h+b+c] |
| bgPt-423527 | 7 | 0 | 0 | 42 | 0 | 24 | 11.27 | 1 | ***** | [a:h+b+c] |
| bgPt-597457 | 11 | 0 | 0 | 53 | 0 | 9 | 11.27 | 1 | ***** | [a:h+b+c] |
| bgPabg-422515 | 0 | 0 | 35 | 0 | 29 | 9 | 8.07 | 1 | **** | [a+h+d:b] |
| bgPt-598223 | 12 | 0 | 0 | 49 | 0 | 12 | 8.27 | 1 | **** | [a:h+b+c] |
| bgPt-596528 | 0 | 0 | 11 | 0 | 47 | 15 | 8.5 | 1 | **** | [a+h+d:b] |
| bgPt-602242 | 10 | 0 | 0 | 47 | 0 | 16 | 9.69 | 1 | **** | [a:h+b+c] |

Appendix 23 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| AG81 | 39 | 8 | 26 | 0 | 0 | 0 | 10.76 | 2 | **** | [a:h:b] |
| bgPabg-423727 | 0 | 0 | 35 | 0 | 31 | 7 | 6.79 | 1 | *** | [a+h+d:b] |
| bgPabg-597688 | 13 | 0 | 0 | 48 | 0 | 12 | 6.82 | 1 | *** | [a:h+b+c] |
| bgPt-598698 | 11 | 0 | 0 | 44 | 0 | 18 | 7.19 | 1 | *** | [a:h+b+c] |
| bgPabg-423557 | 0 | 0 | 34 | 0 | 29 | 10 | 7.29 | 1 | *** | [a+h+d:b] |
| bgPt-594258 | 0 | 0 | 16 | 0 | 47 | 10 | 3.94 | 1 | ** | [a+h+d:b] |
| bgPabg-594494 | 0 | 0 | 18 | 0 | 52 | 3 | 4.15 | 1 | ** | [a+h+d:b] |
| bgPabg-423556 | 0 | 0 | 32 | 0 | 32 | 9 | 4.27 | 1 | ** | [a+h+d:b] |
| bgPt-594658 | 0 | 0 | 16 | 0 | 48 | 9 | 4.27 | 1 | ** | [a+h+d:b] |
| bgPt-595005 | 0 | 0 | 32 | 0 | 32 | 9 | 4.27 | 1 | ** | [a+h+d:b] |
| bgPabg-596205 | 33 | 0 | 0 | 33 | 0 | 7 | 4.4 | 1 | ** | [a:h+b+c] |
| bgPabg-423121 | 0 | 0 | 33 | 0 | 33 | 7 | 4.4 | 1 | ** | [a+h+d:b] |
| bgPt-598296 | 14 | 0 | 0 | 44 | 0 | 15 | 4.42 | 1 | ** | [a:h+b+c] |
| bgPabg-594288 | 0 | 0 | 18 | 0 | 53 | 2 | 4.47 | 1 | ** | [a+h+d:b] |
| bgPt-598407 | 13 | 0 | 0 | 42 | 0 | 18 | 4.51 | 1 | ** | [a:h+b+c] |
| bgPt-598598 | 15 | 0 | 0 | 47 | 0 | 11 | 4.68 | 1 | ** | [a:h+b+c] |
| bgPt-597688 | 32 | 0 | 0 | 31 | 0 | 10 | 4.75 | 1 | ** | [a:h+b+c] |
| bgPabg-423378 | 0 | 0 | 32 | 0 | 31 | 10 | 4.75 | 1 | ** | [a+h+d:b] |
| bgPabg-423708 | 0 | 0 | 32 | 0 | 31 | 10 | 4.75 | 1 | ** | [a+h+d:b] |
| bgPt-598212 | 14 | 0 | 0 | 45 | 0 | 14 | 4.77 | 1 | ** | [a:h+b+c] |
| bgPt-600914 | 16 | 0 | 0 | 50 | 0 | 7 | 4.95 | 1 | ** | [a:h+b+c] |
| bgPabg-597113 | 34 | 0 | 0 | 33 | 0 | 6 | 5.02 | 1 | ** | [a:h+b+c] |
| bgPt-600845 | 15 | 0 | 0 | 48 | 0 | 10 | 5.04 | 1 | ** | [a:h+b+c] |
| bgPt-602273 | 14 | 0 | 0 | 46 | 0 | 13 | 5.14 | 1 | ** | [a:h+b+c] |
| bgPabg-422619 | 0 | 0 | 31 | 0 | 29 | 13 | 5.14 | 1 | ** | [a+h+d:b] |
| bgPt-598513 | 15 | 0 | 0 | 49 | 0 | 9 | 5.4 | 1 | ** | [a:h+b+c] |
| bgPabg-423430 | 0 | 0 | 34 | 0 | 32 | 7 | 5.53 | 1 | ** | [a+h+d:b] |
| bgPt-594435 | 31 | 0 | 0 | 28 | 0 | 14 | 5.7 | 1 | ** | [a:h+b+c] |
| bgPabg-594180 | 0 | 0 | 33 | 0 | 30 | 10 | 5.95 | 1 | ** | [a+h+d:b] |
| bgPabg-595799 | 15 | 0 | 0 | 51 | 0 | 7 | 6.15 | 1 | ** | [a:h+b+c] |
| bgPt-597377 | 30 | 0 | 0 | 26 | 0 | 17 | 6.17 | 1 | ** | [a:h+b+c] |
| bgPt-601401 | 12 | 0 | 0 | 44 | 0 | 17 | 6.17 | 1 | ** | [a:h+b+c] |
| bgPt-600790 | 32 | 0 | 0 | 28 | 0 | 13 | 6.42 | 1 | ** | [a:h+b+c] |
| bgPabg-595118 | 15 | 0 | 0 | 52 | 0 | 6 | 6.53 | 1 | ** | [a:h+b+c] |
| bgPabg-600828 | 0 | 0 | 33 | 0 | 29 | 11 | 6.54 | 1 | ** | [a+h+d:b] |
| PRIMER2 | 17 | 25 | 31 | 0 | 0 | 0 | 6.91 | 2 | ** | [a:h:b] |
| PRIMER45 | 18 | 27 | 28 | 0 | 0 | 0 | 7.42 | 2 | ** | [a:h:b] |
| bgPabg-423527 | 16 | 0 | 0 | 43 | 0 | 14 | 2.71 | 1 | * | [a:h+b+c] |
| bgPabg-422458 | 0 | 0 | 30 | 0 | 33 | 10 | 2.75 | 1 | * | [a+h+d:b] |
| bgPt-601852 | 14 | 0 | 0 | 39 | 0 | 20 | 2.78 | 1 | * | [a:h+b+c] |
| bgPabg-423147 | 0 | 0 | 31 | 0 | 34 | 8 | 2.88 | 1 | * | [a+h+d:b] |
| bgPt-597010 | 17 | 0 | 0 | 46 | 0 | 10 | 2.97 | 1 | * | [a:h+b+c] |
| bgPt-597832 | 17 | 0 | 0 | 46 | 0 | 10 | 2.97 | 1 | * | [a:h+b+c] |
| bgPabg-423618 | 0 | 0 | 29 | 0 | 31 | 13 | 3 | 1 | * | [a+h+d:b] |
| bgPabg-422294 | 0 | 0 | 30 | 0 | 32 | 11 | 3.14 | 1 | * | [a+h+d:b] |
| bgPabg-422618 | 0 | 0 | 30 | 0 | 32 | 11 | 3.14 | 1 | * | [a+h+d:b] |
| bgPabg-422656 | 0 | 0 | 30 | 0 | 32 | 11 | 3.14 | 1 | * | [a+h+d:b] |
| bgPabg-422657 | 0 | 0 | 30 | 0 | 32 | 11 | 3.14 | 1 | * | [a+h+d:b] |
| bgPabg-423395 | 0 | 0 | 31 | 0 | 33 | 9 | 3.27 | 1 | * | [a+h+d:b] |
| bgPt-595011 | 0 | 0 | 17 | 0 | 47 | 9 | 3.27 | 1 | * | [a+h+d:b] |
| bgPt-598091 | 16 | 0 | 0 | 45 | 0 | 12 | 3.31 | 1 | * | [a:h+b+c] |
| bgPabg-423122 | 0 | 0 | 30 | 0 | 31 | 12 | 3.55 | 1 | * | [a+h+d:b] |
| bgPt-598289 | 0 | 0 | 17 | 0 | 48 | 8 | 3.57 | 1 | * | [a+h+d:b] |
| bgPt-598809 | 15 | 0 | 0 | 44 | 0 | 14 | 3.67 | 1 | * | [a:h+b+c] |
| bgPabg-423390 | 0 | 0 | 32 | 0 | 33 | 8 | 3.82 | 1 | * | [a+h+d:b] |
| DSRB16 | 24 | 13 | 36 | 0 | 0 | 0 | 4.64 | 2 | * | [a:h:b] |
| PRIMER37 | 21 | 26 | 26 | 0 | 0 | 0 | 4.84 | 2 | * | [a:h:b] |
| PRIMER98 | 36 | 11 | 26 | 0 | 0 | 0 | 5.67 | 2 | * | [a:h:b] |
| BN145 | 35 | 10 | 28 | 0 | 0 | 0 | 5.87 | 2 | * | [a:h:b] |

**Appendix 24: The locus genotype frequency for wide cross map of phase determination**

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| P1b251AAG+CTT4 | 25 | 0 | 0 | 73 | 0 | 0 | 0.01 | 1 | - | [a:h+b+c] |
| P2b83AAC+CTA1 | 26 | 0 | 0 | 72 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P2b213AAC+CTA2 | 26 | 0 | 0 | 72 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P3b87AGC+CTC2 | 0 | 0.00 | 23 | 0 | 75.00 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P3b99AGC+CTC3 | 25 | 0.00 | 0 | 73 | 0.00 | 0 | 0.01 | 1 | - | [a:h+b+c] |
| P3b155AGC+CTC4 | 18 | 0.00 | 0 | 80 | 0.00 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P3b157AGC+CTC5 | 0 | 0 | 25 | 0 | 73.00 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P3b264AGC+CTC8 | 0 | 0 | 26 | 0 | 72 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P4b224AGC+CAG | 0 | 0 | 23 | 0 | 75 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P5b93ATC+CTT1 | 0 | 0 | 27 | 0 | 71 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P5b107ATC+CTT2 | 24 | 0 | 0 | 74 | 0 | 0 | 0.01 | 1 | - | [a:h+b+c] |
| P5b191ATC+CTT3 | 0 | 0 | 19 | 0 | 79 | 0 | 1.65 | 1 | - | [a+h+d:b] |
| P6b140AAC+CAG1 | 22 | 0 | 0 | 76 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P6b187AAC+CAG2 | 26 | 0 | 0 | 72 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P6b217AAC+CAG3 | 27 | 0 | 0 | 71 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P7b333AAG+CAG1 | 21 | 0 | 0 | 77 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P7b357AAG+CAG2 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P8b270ACG+CTA1 | 0 | 0 | 30 | 0 | 68 | 0 | 1.65 | 1 | - | [a+h+d:b] |
| P9b153AGA+CTA2 | 0 | 0 | 22 | 0 | 76 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P9b201AGA+CTA3 | 0 | 0 | 20 | 0 | 78 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P9b205AGA+CTA4 | 22 | 0 | 0 | 76 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P9b221AGA+CTA5 | 0 | 0 | 21 | 0 | 77 | 0 | 0.67 | 1 | - | [a+h+d:b] |
| P9b331AGA+CTA6 | 27 | 0 | 0 | 71 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P10b130AAC+CTT1 | 0 | 0 | 23 | 0 | 75 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P10b296AAC+CTT3 | 23 | 0 | 0 | 75 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P12b374AAG+CTC2 | 19 | 0 | 0 | 79 | 0 | 0 | 1.65 | 1 | - | [a:h+b+c] |
| P12b420AAG+CTC4 | 22 | 0 | 0 | 76 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P14b305ATC+CTA2 | 0 | 0 | 22 | 0 | 76 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P15b351AGT+CAA2 | 28 | 0 | 0 | 70 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P15b646AGT+CAA3 | 0 | 0 | 21 | 0 | 77 | 0 | 0.67 | 1 | - | [a+h+d:b] |
| P16b121AGA+CTC1 | 22 | 0 | 0 | 76 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P16b329AGA+CTC2 | 19 | 0 | 0 | 79 | 0 | 0 | 1.65 | 1 | - | [a:h+b+c] |
| P17b180ACG+CAA1 | 0 | 0 | 19 | 0 | 79 | 0 | 1.65 | 1 | - | [a+h+d:b] |
| P17b240ACG+CAA2 | 20 | 0 | 0 | 78 | 0 | 0 | 1.1 | 1 | - | [a:h+b+c] |
| P17b242ACG+CAA3 | 0 | 0 | 28 | 0 | 70 | 0 | 0.67 | 1 | - | [a+h+d:b] |
| P18b186ACA+CAT1 | 0 | 0 | 28 | 0 | 70 | 0 | 0.67 | 1 | - | [a+h+d:b] |
| P18b221ACA+CAT2 | 0 | 0 | 25 | 0 | 73 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P18b225ACA+CAT3 | 0 | 0 | 20 | 0 | 78 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P18b226ACA+CAT4 | 26 | 0 | 0 | 72 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P19b102AGC+CTA1 | 0 | 0 | 27 | 0 | 71 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P19b105AGC+CTA2 | 27 | 0 | 0 | 71 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P20b127ACG+CAG1 | 22 | 0 | 0 | 76 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P20b245ACG+CAG2 | 23 | 0 | 0 | 75 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P21b132AGC+CA1 | 0 | 0 | 27 | 0 | 71 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P21b154AGC+CA2 | 0 | 0 | 26 | 0 | 72 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P21b224AGC+CA4 | 0 | 0 | 30 | 0 | 68 | 0 | 1.65 | 1 | - | [a+h+d:b] |
| P21b370AGC+CA5 | 0 | 0 | 18 | 0 | 80 | 0 | 2.3 | 1 | - | [a+h+d:b] |
| P22b101GCC+CA1 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P22b166GCC+CA2 | 0 | 0 | 25 | 0 | 73 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P22b243GCC+CA4 | 23 | 0 | 0 | 75 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P22b47GCC+CA5 | 0 | 0 | 26 | 0 | 72 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P22b252GCC+CA6 | 0 | 0 | 30 | 0 | 68 | 0 | 1.65 | 1 | - | [a+h+d:b] |
| P22b285GCC+CA7 | 21 | 0 | 0 | 77 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P22b297GCC+CA8 | 29 | 0 | 0 | 69 | 0 | 0 | 1.1 | 1 | - | [a:h+b+c] |
| P22b307GCC+CA9 | 0 | 0 | 25 | 0 | 73 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P23b217AAC+CA1 | 27 | 0 | 0 | 71 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P23b308AAC+CA3 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P24b79AAC+AG1 | 27 | 0 | 0 | 71 | 0 | 0 | 0.34 | 1 | - | [a:h+b+c] |
| P24b159AAC+AG8 | 31 | 0 | 0 | 67 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P24b221AAC+AG9 | 31 | 0 | 0 | 67 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P24b289AAC+AG13 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P24b291AAC+AG14 | 18 | 0 | 0 | 80 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P25b93AAG+AG1 | 0 | 0 | 25 | 0 | 73 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P25b115AAG+AG3 | 25 | 0 | 0 | 73 | 0 | 0 | 0.01 | 1 | - | [a:h+b+c] |
| P25b199AAG+AG6 | 21 | 0 | 0 | 77 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P25b270AAG+AG7 | 0 | 0 | 20 | 0 | 78 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P25b320AAG+AG9 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| 597705 | 17 | 0 | 0 | 43 | 0 | 38 | 0.36 | 1 | - | [a:h+b+c] |

295

Appendix 24 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| P25b482AAG+AG12 | 30 | 0 | 0 | 68 | 0 | 0 | 1.65 | 1 | - | [a:h+b+c] |
| P26b113ACA+AG2 | 26 | 0 | 0 | 72 | 0 | 0 | 0.12 | 1 | - | [a:h+b+c] |
| P26b116ACA+AG3 | 21 | 0 | 0 | 77 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P26b136ACA+AG4 | 0 | 0 | 31 | 0 | 67 | 0 | 2.3 | 1 | - | [a+h+d:b] |
| P26b142ACA+AG5 | 0 | 0 | 25 | 0 | 73 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P26b193ACA+AG8 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P26b280ACA+AG9 | 0 | 0 | 29 | 0 | 69 | 0 | 1.1 | 1 | - | [a+h+d:b] |
| P26b282ACA+AG10 | 0 | 0 | 24 | 0 | 74 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P26b374ACA+AG12 | 0 | 0 | 23 | 0 | 75 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P27b167AGA+CA1 | 0 | 0 | 27 | 0 | 71 | 0 | 0.34 | 1 | - | [a+h+d:b] |
| P27b329AGA+CA3 | 18 | 0 | 0 | 80 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P27b353AGA+CA4 | 0 | 0 | 31 | 0 | 67 | 0 | 2.3 | 1 | - | [a+h+d:b] |
| P27b383AGA+CA5 | 0 | 0 | 23 | 0 | 75 | 0 | 0.12 | 1 | - | [a+h+d:b] |
| P27b414AGA+CA6 | 30 | 0 | 0 | 68 | 0 | 0 | 1.65 | 1 | - | [a:h+b+c] |
| P28b70AGT+CA1 | 0 | 0 | 21 | 0 | 77 | 0 | 0.67 | 1 | - | [a+h+d:b] |
| P28b99AGT+CA2 | 21 | 0 | 0 | 77 | 0 | 0 | 0.67 | 1 | - | [a:h+b+c] |
| P28b154AGT+CA3 | 25 | 0 | 0 | 73 | 0 | 0 | 0.01 | 1 | - | [a:h+b+c] |
| P28235+237AGT+CA4+5 | 23 | 46 | 29 | 0 | 0 | 0 | 1.1 | 2 | - | [a:h:b] |
| P28b241AGT+CA6 | 18 | 0 | 0 | 80 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P28b255AGT+CA7 | 0 | 0 | 24 | 0 | 74 | 0 | 0.01 | 1 | - | [a+h+d:b] |
| P28b286AGT+CA8 | 31 | 0 | 0 | 67 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| P28b351AGT+CA10 | 31 | 0 | 0 | 67 | 0 | 0 | 2.3 | 1 | - | [a:h+b+c] |
| AG81 | 17 | 59 | 22 | 0 | 0 | 0 | 4.59 | 2 | - | [a:h:b] |
| BGSRRCol7ssrCol7 | 21 | 46 | 28 | 0 | 0 | 3 | 1.13 | 2 | - | [a:h:b] |
| B63DominantssrB63 | 0 | 0 | 28 | 0 | 64 | 6 | 1.45 | 1 | - | [a+h+d:b] |
| B63Codominant | 23 | 38 | 31 | 0 | 0 | 6 | 4.17 | 2 | - | [a:h:b] |
| PRIMER32 | 20 | 51 | 15 | 0 | 0 | 12 | 3.56 | 2 | - | [a:h:b] |
| Bm2coL33 | 22 | 46 | 20 | 0 | 0 | 10 | 0.27 | 2 | - | [a:h:b] |
| BN145 | 21 | 52 | 15 | 0 | 0 | 10 | 3.73 | 2 | - | [a:h:b] |
| PRIMER98 | 18 | 40 | 24 | 0 | 0 | 16 | 0.93 | 2 | - | [a:h:b] |
| mBam3co7 | 16 | 45 | 27 | 0 | 0 | 10 | 2.8 | 2 | - | [a:h:b] |
| PRIMER15 | 26 | 44 | 19 | 0 | 0 | 9 | 1.11 | 2 | - | [a:h:b] |
| Bam2col80 | 20 | 26 | 15 | 0 | 0 | 37 | 2.15 | 2 | - | [a:h:b] |
| GH-19-B2-D9 | 14 | 41 | 27 | 0 | 0 | 16 | 4.12 | 2 | - | [a:h:b] |
| PRIMER65 | 27 | 40 | 19 | 0 | 0 | 12 | 1.91 | 2 | - | [a:h:b] |
| PRIMER16 | 23 | 46 | 16 | 0 | 0 | 13 | 1.73 | 2 | - | [a:h:b] |
| PRIMER38 | 24 | 41 | 24 | 0 | 0 | 9 | 0.55 | 2 | - | [a:h:b] |
| PRIMER73 | 19 | 49 | 20 | 0 | 0 | 10 | 1.16 | 2 | - | [a:h:b] |
| PRIMER82 | 22 | 41 | 25 | 0 | 0 | 10 | 0.61 | 2 | - | [a:h:b] |
| PRIMER10 | 17 | 48 | 23 | 0 | 0 | 10 | 1.55 | 2 | - | [a:h:b] |
| PRIMER19 | 27 | 37 | 19 | 0 | 0 | 15 | 2.52 | 2 | - | [a:h:b] |
| 423624 | 0 | 0 | 19 | 0 | 50 | 29 | 0.24 | 1 | - | [a+h+d:b] |
| 594856 | 17 | 0 | 0 | 52 | 0 | 29 | 0 | 1 | - | [a:h+b+c] |
| 594857 | 11 | 0 | 0 | 56 | 0 | 31 | 2.63 | 1 | - | [a:h+b+c] |
| 595170 | 17 | 0 | 0 | 50 | 0 | 31 | 0 | 1 | - | [a:h+b+c] |
| 596903 | 0 | 0 | 20 | 0 | 46 | 32 | 0.99 | 1 | - | [a+h+d:b] |
| 594400 | 0 | 0 | 19 | 0 | 43 | 36 | 1.05 | 1 | - | [a+h+d:b] |
| 596605 | 0 | 0 | 13 | 0 | 53 | 32 | 0.99 | 1 | - | [a+h+d:b] |
| 596233 | 12 | 0 | 0 | 47 | 0 | 39 | 0.68 | 1 | - | [a:h+b+c] |
| 597717 | 13 | 0 | 0 | 45 | 0 | 40 | 0.21 | 1 | - | [a:h+b+c] |
| 596311 | 0 | 0 | 14 | 0 | 43 | 41 | 0.01 | 1 | - | [a+h+d:b] |
| 596059 | 16 | 0 | 0 | 52 | 0 | 30 | 0.08 | 1 | - | [a:h+b+c] |
| 422614 | 14 | 0 | 0 | 46 | 0 | 38 | 0.09 | 1 | - | [a:h+b+c] |
| 594987 | 0 | 0 | 16 | 0 | 47 | 35 | 0.01 | 1 | - | [a+h+d:b] |
| 594981 | 14 | 0 | 0 | 44 | 0 | 40 | 0.02 | 1 | - | [a:h+b+c] |
| 601441 | 0 | 0 | 15 | 0 | 43 | 40 | 0.02 | 1 | - | [a+h+d:b] |
| 597860 | 15 | 0 | 0 | 45 | 0 | 38 | 0 | 1 | - | [a:h+b+c] |
| 423415 | 0 | 0 | 16 | 0 | 44 | 38 | 0.09 | 1 | - | [a+h+d:b] |
| 597289 | 15 | 0 | 0 | 43 | 0 | 40 | 0.02 | 1 | - | [a:h+b+c] |
| 594656 | 0 | 0 | 16 | 0 | 42 | 40 | 0.21 | 1 | - | [a+h+d:b] |
| 423049 | 0 | 0 | 16 | 0 | 42 | 40 | 0.21 | 1 | - | [a+h+d:b] |
| 595013 | 0 | 0 | 15 | 0 | 41 | 42 | 0.1 | 1 | - | [a+h+d:b] |
| 596646 | 15 | 0 | 0 | 42 | 0 | 41 | 0.05 | 1 | - | [a:h+b+c] |
| 423257 | 15 | 0 | 0 | 43 | 0 | 40 | 0.02 | 1 | - | [a:h+b+c] |
| 422333 | 15 | 0 | 0 | 41 | 0 | 42 | 0.1 | 1 | - | [a:h+b+c] |
| 597343 | 0 | 0 | 16 | 0 | 39 | 43 | 0.49 | 1 | - | [a+h+d:b] |
| 595707 | 17 | 0 | 0 | 44 | 0 | 37 | 0.27 | 1 | - | [a:h+b+c] |
| 423152 | 0 | 0 | 18 | 0 | 43 | 37 | 0.66 | 1 | - | [a+h+d:b] |

Appendix 24 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| 594814 | 17 | 0 | 0 | 43 | 0 | 38 | 0.36 | 1 | - | [a:h+b+c] |
| 597279 | 0 | 0 | 18 | 0 | 45 | 35 | 0.43 | 1 | - | [a+h+d:b] |
| 597746 | 17 | 0 | 0 | 41 | 0 | 40 | 0.57 | 1 | - | [a:h+b+c] |
| 597971 | 0 | 0 | 18 | 0 | 41 | 39 | 0.95 | 1 | - | [a+h+d:b] |
| 423386 | 16 | 0 | 0 | 42 | 0 | 40 | 0.21 | 1 | - | [a:h+b+c] |
| 595948 | 0 | 0 | 18 | 0 | 40 | 40 | 1.13 | 1 | - | [a+h+d:b] |
| 597945 | 0 | 0 | 18 | 0 | 42 | 38 | 0.8 | 1 | - | [a+h+d:b] |
| 597087 | 0 | 0 | 16 | 0 | 39 | 43 | 0.49 | 1 | - | [a+h+d:b] |
| 595380 | 0 | 0 | 18 | 0 | 40 | 40 | 1.13 | 1 | - | [a+h+d:b] |
| 595797 | 0 | 0 | 19 | 0 | 38 | 41 | 2.11 | 1 | - | [a+h+d:b] |
| 423742 | 17 | 0 | 0 | 41 | 0 | 40 | 0.57 | 1 | - | [a:h+b+c] |
| 596776 | 0 | 0 | 18 | 0 | 35 | 45 | 2.27 | 1 | - | [a+h+d:b] |
| 594685 | 0 | 0 | 20 | 0 | 39 | 39 | 2.49 | 1 | - | [a+h+d:b] |
| 595006 | 0 | 0 | 20 | 0 | 42 | 36 | 1.74 | 1 | - | [a+h+d:b] |
| 598173 | 0 | 0 | 19 | 0 | 41 | 38 | 1.42 | 1 | - | [a+h+d:b] |
| 423258 | 18 | 0 | 0 | 37 | 0 | 43 | 1.75 | 1 | - | [a:h+b+c] |
| 601723 | 0 | 0 | 10 | 0 | 52 | 36 | 2.6 | 1 | - | [a+h+d:b] |
| 594707 | 9 | 0 | 0 | 47 | 0 | 42 | 2.38 | 1 | - | [a:h+b+c] |
| 593909 | 11 | 0 | 0 | 54 | 0 | 33 | 2.26 | 1 | - | [a:h+b+c] |
| 596231 | 12 | 0 | 0 | 52 | 0 | 34 | 1.33 | 1 | - | [a:h+b+c] |
| 595901 | 12 | 0 | 0 | 55 | 0 | 31 | 1.8 | 1 | - | [a:h+b+c] |
| 602199 | 12 | 0 | 0 | 54 | 0 | 32 | 1.64 | 1 | - | [a:h+b+c] |
| 600900 | 0 | 0 | 13 | 0 | 48 | 37 | 0.44 | 1 | - | [a+h+d:b] |
| 597681 | 12 | 0 | 0 | 54 | 0 | 32 | 1.64 | 1 | - | [a:h+b+c] |
| 595924 | 12 | 0 | 0 | 53 | 0 | 33 | 1.48 | 1 | - | [a:h+b+c] |
| 601918 | 12 | 0 | 0 | 48 | 0 | 38 | 0.8 | 1 | - | [a:h+b+c] |
| 595496 | 12 | 0 | 0 | 52 | 0 | 34 | 1.33 | 1 | - | [a:h+b+c] |
| 423805 | 14 | 0 | 0 | 52 | 0 | 32 | 0.51 | 1 | - | [a:h+b+c] |
| 422423 | 11 | 0 | 0 | 45 | 0 | 42 | 0.86 | 1 | - | [a:h+b+c] |
| 598804 | 13 | 0 | 0 | 47 | 0 | 38 | 0.36 | 1 | - | [a:h+b+c] |
| 598647 | 14 | 0 | 0 | 50 | 0 | 34 | 0.33 | 1 | - | [a:h+b+c] |
| 601358 | 14 | 0 | 0 | 50 | 0 | 34 | 0.33 | 1 | - | [a:h+b+c] |
| 595882 | 13 | 0 | 0 | 52 | 0 | 33 | 0.87 | 1 | - | [a:h+b+c] |
| 600972 | 14 | 0 | 0 | 49 | 0 | 35 | 0.26 | 1 | - | [a:h+b+c] |
| 598750 | 14 | 0 | 0 | 49 | 0 | 35 | 0.26 | 1 | - | [a:h+b+c] |
| 596721 | 14 | 0 | 0 | 48 | 0 | 36 | 0.19 | 1 | - | [a:h+b+c] |
| 601257 | 14 | 0 | 0 | 47 | 0 | 37 | 0.14 | 1 | - | [a:h+b+c] |
| 601845 | 15 | 0 | 0 | 49 | 0 | 34 | 0.08 | 1 | - | [a:h+b+c] |
| 601050 | 15 | 0 | 0 | 49 | 0 | 34 | 0.08 | 1 | - | [a:h+b+c] |
| 597434 | 0 | 0 | 15 | 0 | 49 | 34 | 0.08 | 1 | - | [a+h+d:b] |
| 598337 | 14 | 0 | 0 | 45 | 0 | 39 | 0.05 | 1 | - | [a:h+b+c] |
| 596097 | 15 | 0 | 0 | 46 | 0 | 37 | 0.01 | 1 | - | [a:h+b+c] |
| 595754 | 13 | 0 | 0 | 45 | 0 | 40 | 0.21 | 1 | - | [a:h+b+c] |
| 596331 | 14 | 0 | 0 | 45 | 0 | 39 | 0.05 | 1 | - | [a:h+b+c] |
| 594898 | 14 | 0 | 0 | 45 | 0 | 39 | 0.05 | 1 | - | [a:h+b+c] |
| 601965 | 15 | 0 | 0 | 44 | 0 | 39 | 0.01 | 1 | - | [a:h+b+c] |
| 423710 | 15 | 0 | 0 | 43 | 0 | 40 | 0.02 | 1 | - | [a:h+b+c] |
| 601384 | 16 | 0 | 0 | 46 | 0 | 36 | 0.02 | 1 | - | [a:h+b+c] |
| 601470 | 17 | 0 | 0 | 47 | 0 | 34 | 0.08 | 1 | - | [a:h+b+c] |
| 598182 | 17 | 0 | 0 | 45 | 0 | 36 | 0.19 | 1 | - | [a:h+b+c] |
| 594803 | 15 | 0 | 0 | 43 | 0 | 40 | 0.02 | 1 | - | [a:h+b+c] |
| 594994 | 15 | 0 | 0 | 42 | 0 | 41 | 0.05 | 1 | - | [a:h+b+c] |
| 596047 | 19 | 0 | 0 | 47 | 0 | 32 | 0.51 | 1 | - | [a:h+b+c] |
| 601748 | 17 | 0 | 0 | 43 | 0 | 38 | 0.36 | 1 | - | [a:h+b+c] |
| 594224 | 16 | 0 | 0 | 44 | 0 | 38 | 0.09 | 1 | - | [a:h+b+c] |
| 593937 | 15 | 0 | 0 | 41 | 0 | 42 | 0.1 | 1 | - | [a:h+b+c] |
| 595761 | 16 | 0 | 0 | 44 | 0 | 38 | 0.09 | 1 | - | [a:h+b+c] |
| 594391 | 17 | 0 | 0 | 44 | 0 | 37 | 0.27 | 1 | - | [a:h+b+c] |
| 597175 | 0 | 0 | 20 | 0 | 46 | 32 | 0.99 | 1 | - | [a+h+d:b] |
| 596042 | 0 | 0 | 19 | 0 | 43 | 36 | 1.05 | 1 | - | [a+h+d:b] |
| 597962 | 16 | 0 | 0 | 42 | 0 | 40 | 0.21 | 1 | - | [a:h+b+c] |
| 597832 | 16 | 0 | 0 | 41 | 0 | 41 | 0.29 | 1 | - | [a:h+b+c] |
| 595583 | 17 | 0 | 0 | 41 | 0 | 40 | 0.57 | 1 | - | [a:h+b+c] |
| 594712 | 18 | 0 | 0 | 43 | 0 | 37 | 0.66 | 1 | - | [a:h+b+c] |
| 597285 | 18 | 0 | 0 | 43 | 0 | 37 | 0.66 | 1 | - | [a:h+b+c] |
| 597010 | 17 | 0 | 0 | 42 | 0 | 39 | 0.46 | 1 | - | [a:h+b+c] |
| 602110 | 17 | 0 | 0 | 42 | 0 | 39 | 0.46 | 1 | - | [a:h+b+c] |
| 595093 | 17 | 0 | 0 | 42 | 0 | 39 | 0.46 | 1 | - | [a:h+b+c] |

297

Appendix 24 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| 595373 | 17 | 0 | 0 | 40 | 0 | 41 | 0.71 | 1 | - | [a:h+b+c] |
| 596726 | 17 | 0 | 0 | 38 | 0 | 43 | 1.02 | 1 | - | [a:h+b+c] |
| 594531 | 18 | 0 | 0 | 39 | 0 | 41 | 1.32 | 1 | - | [a:h+b+c] |
| 596188 | 18 | 0 | 0 | 39 | 0 | 41 | 1.32 | 1 | - | [a:h+b+c] |
| 422352 | 19 | 0 | 0 | 39 | 0 | 40 | 1.86 | 1 | - | [a:h+b+c] |
| 596017 | 20 | 0 | 0 | 41 | 0 | 37 | 1.97 | 1 | - | [a:h+b+c] |
| 596545 | 18 | 0 | 0 | 38 | 0 | 42 | 1.52 | 1 | - | [a:h+b+c] |
| 594387 | 20 | 0 | 0 | 39 | 0 | 39 | 2.49 | 1 | - | [a:h+b+c] |
| 597365 | 19 | 0 | 0 | 37 | 0 | 42 | 2.38 | 1 | - | [a:h+b+c] |
| 597037 | 19 | 0 | 0 | 37 | 0 | 42 | 2.38 | 1 | - | [a:h+b+c] |
| 596417 | 20 | 0 | 0 | 40 | 0 | 38 | 2.22 | 1 | - | [a:h+b+c] |
| 598056 | 19 | 0 | 0 | 38 | 0 | 41 | 2.11 | 1 | - | [a:h+b+c] |
| 595470 | 19 | 0 | 0 | 36 | 0 | 43 | 2.67 | 1 | - | [a:h+b+c] |
| 595859 | 19 | 0 | 0 | 38 | 0 | 41 | 2.11 | 1 | - | [a:h+b+c] |
| P1b67AAG+CTT1 | 0 | 0 | 42 | 0 | 56 | 0 | 16.67 | 1 | ******* | [a+h+d:b] |
| P8b315ACG+CTA2 | 0 | 0 | 46 | 0 | 52 | 0 | 25.16 | 1 | ******* | [a+h+d:b] |
| P12b105AAG+CTC1 | 0 | 0 | 43 | 0 | 55 | 0 | 18.63 | 1 | ******* | [a+h+d:b] |
| P23b287AAC+CA2 | 44 | 0 | 0 | 54 | 0 | 0 | 20.69 | 1 | ******* | [a:h+b+c] |
| P24b271AAC+AG11 | 0 | 0 | 43 | 0 | 55 | 0 | 18.63 | 1 | ******* | [a+h+d:b] |
| P24b392AAC+AG15 | 0 | 0 | 50 | 0 | 48 | 0 | 35.39 | 1 | ******* | [a+h+d:b] |
| P24b499AAC+AG16 | 0 | 0 | 48 | 0 | 50 | 0 | 30.05 | 1 | ******* | [a+h+d:b] |
| P25b94AAG+AG2 | 0 | 0 | 48 | 0 | 50 | 0 | 30.05 | 1 | ******* | [a+h+d:b] |
| P25b299AAG+AG8 | 0 | 0 | 44 | 0 | 54 | 0 | 20.69 | 1 | ******* | [a+h+d:b] |
| P25b444AAG+AG10 | 0 | 0 | 45 | 0 | 53 | 0 | 22.87 | 1 | ******* | [a+h+d:b] |
| P26b108ACA+AG1 | 42 | 0 | 0 | 56 | 0 | 0 | 16.67 | 1 | ******* | [a:h+b+c] |
| PRIMER66 | 12 | 31 | 43 | 0 | 0 | 12 | 29.05 | 2 | ******* | [a:h:b] |
| 598450 | 33 | 0 | 0 | 32 | 0 | 33 | 23.02 | 1 | ******* | [a:h+b+c] |
| 595671 | 0 | 0 | 0 | 68 | 0 | 30 | 22.67 | 1 | ******* | [a:h+b+c] |
| P14b264ATC+CTA1 | 0 | 0 | 41 | 0 | 57 | 0 | 14.82 | 1 | ****** | [a+h+d:b] |
| P19b494AGC+CTA4 | 0 | 0 | 40 | 0 | 58 | 0 | 13.07 | 1 | ****** | [a+h+d:b] |
| P25b178AAG+AG5 | 0 | 0 | 41 | 0 | 57 | 0 | 14.82 | 1 | ****** | [a+h+d:b] |
| P28b328AGT+CA9 | 0 | 0 | 41 | 0 | 57 | 0 | 14.82 | 1 | ****** | [a+h+d:b] |
| P1b161AAG+CTT3 | 39 | 0 | 0 | 59 | 0 | 0 | 11.44 | 1 | ***** | [a:h+b+c] |
| P24b109AAC+AG4 | 39 | 0 | 0 | 59 | 0 | 0 | 11.44 | 1 | ***** | [a:h+b+c] |
| P24b151AAC+AG7 | 0 | 0 | 39 | 0 | 59 | 0 | 11.44 | 1 | ***** | [a+h+d:b] |
| DSRB20 | 37 | 40 | 12 | 0 | 0 | 9 | 14.96 | 2 | ***** | [a:h:b] |
| P1b88AAG+CTT2 | 0 | 0 | 12 | 0 | 86 | 0 | 8.5 | 1 | **** | [a+h+d:b] |
| P11b80AAG+CAA1 | 37 | 0 | 0 | 61 | 0 | 0 | 8.5 | 1 | **** | [a:h+b+c] |
| P12b403AAG+CTC3 | 37 | 0 | 0 | 61 | 0 | 0 | 8.5 | 1 | **** | [a:h+b+c] |
| P22b207GCC+CA3 | 38 | 0 | 0 | 60 | 0 | 0 | 9.92 | 1 | **** | [a:h+b+c] |
| P24b82AAC+AG2 | 0 | 0 | 12 | 0 | 86 | 0 | 8.5 | 1 | **** | [a+h+d:b] |
| P26b168ACA+AG6 | 0 | 0 | 38 | 0 | 60 | 0 | 9.92 | 1 | **** | [a+h+d:b] |
| P26b534ACA+AG13 | 11 | 0 | 0 | 87 | 0 | 0 | 9.92 | 1 | **** | [a:h+b+c] |
| P27b297AGA+CA2 | 0 | 0 | 11 | 0 | 87 | 0 | 9.92 | 1 | **** | [a+h+d:b] |
| P28b570AGT+CA12 | 38 | 0 | 0 | 60 | 0 | 0 | 9.92 | 1 | **** | [a:h+b+c] |
| DSRB17 | 21 | 32 | 36 | 0 | 0 | 9 | 12.08 | 2 | **** | [a:h:b] |
| DSRB19 | 34 | 31 | 21 | 0 | 0 | 12 | 10.63 | 2 | **** | [a:h:b] |
| P5b204ATC+CTT4 | 0 | 0 | 13 | 0 | 85 | 0 | 7.2 | 1 | *** | [a+h+d:b] |
| P28b389AGT+CA11 | 0 | 0 | 36 | 0 | 62 | 0 | 7.2 | 1 | *** | [a+h+d:b] |
| 596162 | 0 | 0 | 23 | 0 | 35 | 40 | 6.64 | 1 | *** | [a+h+d:b] |
| 597026 | 0 | 0 | 24 | 0 | 37 | 37 | 6.69 | 1 | *** | [a+h+d:b] |
| 598161 | 25 | 0 | 0 | 39 | 0 | 34 | 6.75 | 1 | *** | [a:h+b+c] |
| P3b83AGC+CTC1 | 0 | 0 | 15 | 0 | 83 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P3b211AGC+CTC7 | 0 | 0 | 16 | 0 | 82 | 0 | 3.93 | 1 | ** | [a+h+d:b] |
| P9b84AGA+CTA1 | 34 | 0 | 0 | 64 | 0 | 0 | 4.91 | 1 | ** | [a:h+b+c] |
| P10b203AAC+CTT2 | 0 | 0 | 33 | 0 | 65 | 0 | 3.93 | 1 | ** | [a+h+d:b] |
| P15b319AGT+CAA1 | 0 | 0 | 16 | 0 | 82 | 0 | 3.93 | 1 | ** | [a+h+d:b] |
| P18b275ACA+CAT5 | 16 | 0 | 0 | 82 | 0 | 0 | 3.93 | 1 | ** | [a:h+b+c] |
| P18b320ACA+CAT6 | 0 | 0 | 34 | 0 | 64 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P19b131AGC+CTA3 | 0 | 0 | 34 | 0 | 64 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P21b202AGC+CA3 | 0 | 0 | 34 | 0 | 64 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P24b116AAC+AG5 | 0 | 0 | 34 | 0 | 64 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P24b126AAC+AG6 | 15 | 0 | 0 | 83 | 0 | 0 | 4.91 | 1 | ** | [a:h+b+c] |
| P24b228AAC+AG10 | 0 | 0 | 33 | 0 | 65 | 0 | 3.93 | 1 | ** | [a+h+d:b] |
| P24b272AAC+AG12 | 0 | 0 | 34 | 0 | 64 | 0 | 4.91 | 1 | ** | [a+h+d:b] |
| P24b510AAC+AG17 | 34 | 0 | 0 | 64 | 0 | 0 | 4.91 | 1 | ** | [a:h+b+c] |
| P25b164AAG+AG4 | 33 | 0 | 0 | 65 | 0 | 0 | 3.93 | 1 | ** | [a:h+b+c] |
| P26b182ACA+AG7 | 0 | 0 | 15 | 0 | 83 | 0 | 4.91 | 1 | ** | [a+h+d:b] |

298

Appendix 24 (continued).

| Locus | a | h | b | c | d | - | X2 | Df | Signif. | Classification |
|---|---|---|---|---|---|---|---|---|---|---|
| P26b347ACA+AG11 | 34 | 0 | 0 | 64 | 0 | 0 | 4.91 | 1 | ** | [a:h+b+c] |
| PRIMER48 | 21 | 53 | 12 | 0 | 0 | 12 | 6.53 | 2 | ** | [a:h:b] |
| 594941 | 0 | 0 | 21 | 0 | 35 | 42 | 4.67 | 1 | ** | [a+h+d:b] |
| 595018 | 0 | 0 | 22 | 0 | 34 | 42 | 6.1 | 1 | ** | [a+h+d:b] |
| 596804 | 21 | 0 | 0 | 33 | 0 | 44 | 5.56 | 1 | ** | [a:h+b+c] |
| 422577 | 21 | 0 | 0 | 34 | 0 | 43 | 5.1 | 1 | ** | [a:h+b+c] |
| 601017 | 0 | 0 | 8 | 0 | 56 | 34 | 5.33 | 1 | ** | [a+h+d:b] |
| 601131 | 0 | 0 | 8 | 0 | 56 | 34 | 5.33 | 1 | ** | [a+h+d:b] |
| 598669 | 0 | 0 | 8 | 0 | 53 | 37 | 4.6 | 1 | ** | [a+h+d:b] |
| 595196 | 21 | 0 | 0 | 37 | 0 | 40 | 3.89 | 1 | ** | [a:h+b+c] |
| 595834 | 21 | 0 | 0 | 36 | 0 | 41 | 4.26 | 1 | ** | [a:h+b+c] |
| P3b185AGC+CTC6 | 0 | 0 | 32 | 0 | 66 | 0 | 3.06 | 1 | * | [a+h+d:b] |
| P21b394AGC+CA6 | 0 | 0 | 32 | 0 | 66 | 0 | 3.06 | 1 | * | [a+h+d:b] |
| P24b88AAC+AG3 | 0 | 0 | 17 | 0 | 81 | 0 | 3.06 | 1 | * | [a+h+d:b] |
| P25b459AAG+AG11 | 0 | 0 | 17 | 0 | 81 | 0 | 3.06 | 1 | * | [a+h+d:b] |
| PRIMER85 | 31 | 37 | 19 | 0 | 0 | 11 | 5.25 | 2 | * | [a:h:b] |
| PRIMER26 | 13 | 47 | 28 | 0 | 0 | 10 | 5.52 | 2 | * | [a:h:b] |
| 595354 | 0 | 0 | 21 | 0 | 40 | 37 | 2.89 | 1 | * | [a+h+d:b] |
| 594888 | 0 | 0 | 21 | 0 | 39 | 38 | 3.2 | 1 | * | [a+h+d:b] |
| 595675 | 20 | 0 | 0 | 37 | 0 | 41 | 3.09 | 1 | * | [a:h+b+c] |
| 594513 | 19 | 0 | 0 | 34 | 0 | 45 | 3.33 | 1 | * | [a:h+b+c] |
| 601852 | 22 | 0 | 0 | 41 | 0 | 35 | 3.31 | 1 | * | [a:h+b+c] |
| 598103 | 0 | 0 | 22 | 0 | 42 | 34 | 3 | 1 | * | [a+h+d:b] |
| 600901 | 0 | 0 | 21 | 0 | 38 | 39 | 3.53 | 1 | * | [a+h+d:b] |
| 596920 | 21 | 0 | 0 | 38 | 0 | 39 | 3.53 | 1 | * | [a:h+b+c] |
| 595333 | 20 | 0 | 0 | 35 | 0 | 43 | 3.79 | 1 | * | [a:h+b+c] |
| 596472 | 20 | 0 | 0 | 36 | 0 | 42 | 3.43 | 1 | * | [a:h+b+c] |
| 596481 | 20 | 0 | 0 | 36 | 0 | 42 | 3.43 | 1 | * | [a:h+b+c] |
| 594453 | 20 | 0 | 0 | 36 | 0 | 42 | 3.43 | 1 | * | [a:h+b+c] |

**Appendix 25: Unmapped DNA markers of the 'narrow' map in bambara groundnut derived from the cross of DipC and Tiga necaru.**

| No | DNA markers | Type of marker | No | DNA markers | Type of marker |
|---|---|---|---|---|---|
| 1 | Bam2coL33 | SSR | 17 | bgPt-596650 | DArT |
| 2 | Bam2coL80 | | 18 | bgPt-596869 | |
| 3 | PRIMER65 | | 19 | bgPt-596950 | |
| 4 | PRIMER85 | | 20 | bgPt-597130 | |
| 5 | bgPabg-593922 | DArT | 21 | bgPt-597585 | |
| 6 | bgPabg-595641 | | 22 | bgPt-598164 | |
| 7 | bgPabg-596877 | | 23 | bgPt-598235 | |
| 8 | bgPabg-597557 | | 24 | bgPt-598385 | |
| 9 | bgPabg-598400 | | 25 | bgPt-598669 | |
| 10 | bgPabg-601035 | | 26 | bgPt-598683 | |
| 11 | bgPt-594663 | | 27 | bgPt-601017 | |
| 12 | bgPt-594957 | | 28 | bgPt-601027 | |
| 13 | bgPt-595456 | | 29 | bgPt-601131 | |
| 14 | bgPt-595565 | | 30 | bgPt-601486 | |
| 15 | bgPt-596047 | | 31 | bgPabg-594305 | |
| 16 | bgPt-596476 | | | | |

**Appendix 26: Unmapped DNA markers of the 'wide' map in bambara groundnut derived from the cross of DipC and VSSP11.**

| No | DNA markers | Type of marker | No | DNA markers | Type of marker |
|---|---|---|---|---|---|
| 1 | PRIMER10 | SSR | 55 | P24b221AAC+AG9 | AFLP |
| 2 | PRIMER15 | | 56 | P24b228AAC+AG10 | |
| 3 | PRIMER19 | | 57 | P24b289AAC+AG13 | |
| 4 | PRIMER38 | | 58 | P24b291AAC+AG14 | |
| 5 | PRIMER65 | | 59 | P24b392AAC+AG15 | |
| 6 | PRIMER82 | | 60 | P24b499AAC+AG16 | |
| 7 | PRIMER85 | | 61 | P24b510AAC+AG17 | |
| 8 | PRIMER98 | | 62 | P24b79AAC+AG1 | |
| 9 | D.32937 | | 63 | P24b88AAC+AG3 | |
| 10 | D.48339 | | 64 | P25b115AAG+AG3 | |
| 11 | D.51646 | | 65 | P25b199AAG+AG6 | |
| 12 | AG81ssrAG81 | | 66 | P25b270AAG+AG7 | |
| 13 | 422577 | DArT | 67 | P25b299AAG+AG8 | |
| 14 | 423049 | | 68 | P25b320AAG+AG9 | |
| 15 | 423258 | | 69 | P25b444AAG+AG10 | |
| 16 | 423386 | | 70 | P25b459AAG+AG11 | |
| 17 | 594400 | | 71 | P25b482AAG+AG12 | |
| 18 | 594513 | | 72 | P25b93AAG+AG1 | |
| 19 | 594856 | | 73 | P25b94AAG+AG2 | |
| 20 | 594941 | | 74 | P26b116ACA+AG3 | |
| 21 | 595006 | | 75 | P26b282ACA+AG10 | |
| 22 | 595013 | | 76 | P26b534ACA+AG13 | |
| 23 | 595018 | | 77 | P27b383AGA+CA5 | |
| 24 | 595671 | | 78 | P28235+237AGT+CA4+5 | |
| 25 | 595675 | | 79 | P28b241AGT+CA6 | |
| 26 | 595797 | | 80 | P28b570AGT+CA12 | |
| 27 | 595948 | | 81 | P28b70AGT+CA1 | |
| 28 | 596042 | | 82 | P3b155AGC+CTC4 | |
| 29 | 596162 | | 83 | P10b203AAC+CTT2 | |
| 30 | 596311 | | 84 | P12b105AAG+CTC1 | |
| 31 | 596605 | | 85 | P12b403AAG+CTC3 | |
| 32 | 596776 | | 86 | P12b420AAG+CTC4 | |
| 33 | 596903 | | 87 | P16b329AGA+CTC2 | |
| 34 | 597026 | | 88 | P18b221ACA+CAT2 | |
| 35 | 597087 | | 89 | P1b251AAG+CTT4 | |
| 36 | 597175 | | 90 | P1b88AAG+CTT2 | |
| 37 | 597279 | | 91 | P21b132AGC+CA1 | |
| 38 | 597945 | | 92 | P21b154AGC+CA2 | |
| 39 | 598161 | | 93 | P21b202AGC+CA3 | |
| 40 | 598669 | | 94 | P21b224AGC+CA4 | |
| 41 | 601017 | | 95 | P21b370AGC+CA5 | |
| 42 | 601131 | | 96 | P21b394AGC+CA6 | |
| 43 | 601441 | | 97 | P22b243GCC+CA4 | |
| 44 | P12b105AAG+CTC1 | AFLP | 98 | P22b47GCC+CA5 | |
| 45 | P18b186ACA+CAT1 | | 99 | P3b264AGC+CTC8 | |
| 46 | P21b394AGC+CA6 | | 100 | P3b83AGC+CTC1 | |
| 47 | P22b207GCC+CA3 | | 101 | P3b87AGC+CTC2 | |
| 48 | P22b252GCC+CA6 | | 102 | P3b99AGC+CTC3 | |
| 49 | P22b285GCC+CA7 | | 103 | P4b224AGC+CAG | |
| 50 | P23b287AAC+CA2 | | 104 | P5b191ATC+CTT3 | |
| 51 | P23b308AAC+CA3 | | 105 | P5b204ATC+CTT4 | |
| 52 | P24b126AAC+AG6 | | 106 | P8b270ACG+CTA1 | |
| 53 | P24b151AAC+AG7 | | 107 | P8b315ACG+CTA2 | |
| 54 | P24b159AAC+AG8 | | | | |

**Appendix 27: Additive, dominance effect and the residual variance after fitting the QTL for interval mapping analysis, using the F₃ generation data of the 'narrow' cross in the FutureCrop glasshouses**

| Traits | Linkage group | Position (cM) | Locus | mu_A | mu_H | mu_B | Variance | Additive | Dominance |
|---|---|---|---|---|---|---|---|---|---|
| Flower no./plant (FN) | 8 | 2 | - | 86.1 | 108.1 | 119.5 | 978.6 | -16.7 | 5.3 |
| | 8 | 0 | Bam2coL63 | 87.6 | 107.3 | 118.8 | 986.3 | -15.6 | 4.0 |
| Terminal leaflet length (TLL) | 8 | 2 | - | 7.8 | 8.2 | 8.6 | 0.5 | -0.4 | 0.0 |
| | 8 | 0 | Bam2coL63 | 7.8 | 8.2 | 8.6 | 0.5 | -0.4 | 0.0 |
| Terminal leaflet width (TLW) | 3 | 19.7 | bgPt-600935 | 3.8 | 3.3 | 3.4 | 0.2 | 0.2 | -0.3 |
| Leaf area (LA) | 3 | 16.6 | - | 62.8 | 44.9 | 59.1 | 96.8 | 1.8 | -16.1 |
| | 3 | 15.6 | bgPabg-597113 | 62.3 | 44.2 | 59.5 | 96.9 | 1.4 | -16.7 |
| Plant spread (PS) | 4 | 0 | BN6b | 37.5 | 34.3 | 30.1 | 29.9 | 3.7 | 0.5 |
| Stem no./plant (STN) | 4 | 14.2 | - | 9.4 | 10.0 | 11.2 | 2.5 | -0.9 | -0.3 |
| | 4 | 11.2 | bgPt-600898 | 9.5 | 10.0 | 11.1 | 2.5 | -0.8 | -0.3 |
| Node no./stem (NN) | 1 | 33 | bgPabg-596774 | 7.6 | 8.3 | 9.7 | 3.7 | -1.1 | -0.4 |
| | 4 | 11.2 | bgPt-600898 | 9.8 | 8.7 | 7.6 | 3.8 | 1.1 | 0.0 |
| Internode length (IL) | 4 | 3 | bgPabg-596988 | 2.9 | 2.0 | 1.4 | 0.5 | 0.7 | -0.1 |
| Pod no./plant (PN) | 1 | 33 | bgPabg-596774 | 39.8 | 49.9 | 53.8 | 212.7 | -7.0 | 3.1 |
| Double seeded pods/plant (DPN) | 4 | 1 | - | 2.8 | 2.1 | 1.8 | 0.6 | 0.5 | -0.2 |
| | 4 | 0 | BN6b | 2.8 | 2.1 | 1.8 | 0.6 | 0.5 | -0.2 |
| Peduncle length (PEL) | 4 | 1 | - | 3.8 | 3.0 | 2.1 | 0.5 | 0.9 | 0.1 |
| | 4 | 2.4 | bgPt-423527 | 3.8 | 3.0 | 2.1 | 0.5 | 0.9 | 0.0 |
| Pod weight (PWE) | 1 | 33 | bgPabg-596774 | 29.1 | 40.0 | 42.8 | 175.8 | -6.8 | 4.0 |
| Pod length (PLE) | 12 | 15.1 | - | 16.7 | 16.1 | 15.2 | 0.9 | 0.8 | 0.1 |
| | 12 | 12. 9 | bgPt-598767 | 16.7 | 16.0 | 15.2 | 1.0 | 0.8 | 0.0 |
| Pod width (PWD) | 12 | 20.1 | - | 13.5 | 12.9 | 12.3 | 0.4 | 0.6 | 0.0 |
| | 12 | 22.5 | bgPabg-595682 | 13.5 | 12.9 | 12.4 | 0.4 | 0.5 | 0.0 |
| Pod length of double seeded (DPL) | 1 | 0 | bgPabg-597086 | 23.5 | 24.8 | 26.5 | 5.0 | -1.5 | -0.2 |
| | 12 | 10.5 | - | 25.9 | 25.0 | 22.6 | 5.1 | 1.6 | 0.7 |
| Pod width of double seeded (DPW) | 12 | 17.1 | - | 13.7 | 13.3 | 12.7 | 0.5 | 0.5 | 0.1 |
| | 12 | 12.9 | bgPt-598767 | 13.7 | 13.2 | 12.7 | 0.5 | 0.5 | 0.0 |
| Seed length (SEL) | 10 | 49.1 | bgPabg-593983 | 11.4 | 11.3 | 10.7 | 0.6 | 0.4 | 0.3 |
| Seed width (SEW) | 12 | 15.1 | - | 9.4 | 9.2 | 8.9 | 0.3 | 0.3 | 0.0 |
| | 12 | 12.9 | bgPt-598767 | 9.4 | 9.2 | 8.9 | 0.3 | 0.3 | 0.0 |
| Seed no./plant (SEN) | 1 | 34 | - | 42.9 | 52.1 | 58.6 | 265.5 | -7.9 | 1.4 |
| | 1 | 33 | bgPabg-596774 | 43.3 | 52.1 | 58.4 | 265.8 | -7.6 | 1.2 |
| Seed weight (SWT) | 1 | 33 | bgPabg-596774 | 4.7 | 5.5 | 5.7 | 1.0 | -0.5 | 0.4 |
| Biomass dry weight (BDW) | 1 | 33 | bgPabg-596774 | 48.9 | 61.5 | 72.2 | 404.6 | -11.6 | 1.0 |
| Shelling% (SH%) | 7 | 13.3 | bgPabg-594335 | 82.1 | 78.8 | 75.3 | 32.4 | 3.4 | 0.1 |
| 100-seed weight (HSW) | 7 | 9.4 | - | 60.2 | 53.3 | 51.5 | 60.3 | 4.3 | -2.6 |
| | 7 | 10.5 | bgPt-601852 | 60.2 | 53.4 | 51.6 | 60.5 | 4.3 | -2.5 |

**Appendix 28: Additive, dominance effect and the residual variance after fitting the QTL for interval mapping analysis, using the $F_3$ generation data of 'narrow' cross in the field**

| Traits | Linkage group | Position (cM) | Locus | mu_A | mu_H | mu_B | Variance | Additive | Dominance |
|---|---|---|---|---|---|---|---|---|---|
| *Node no./stem (NN)* | 3 | 30.2 | bgPabg-595707 | 16.7 | 11.2 | 14.6 | 12.1 | 1.0 | -4.5 |
| *Internode length (IL)* | 4 | 3 | bgPabg-596988 | 1.4 | 1.0 | 0.7 | 0.1 | 0.3 | -0.1 |
| *Pod no./plant (PN)* | 18 | 3 | - | 2.2 | 2.5 | 3.0 | 0.6 | -0.4 | -0.1 |
| | 18 | 5.1 | PRIMER10 | 2.2 | 2.5 | 3.0 | 0.6 | -0.4 | -0.1 |
| *pod length (PLE)* | 11 | 3 | | 3.8 | 3.6 | 3.6 | 0.0 | 0.1 | -0.1 |
| | 11 | 0 | bgPabg-595822 | 3.8 | 3.7 | 3.6 | 0.0 | 0.1 | 0.0 |
| *Biomass dry weight (BDW)* | 1 | 28.9 | bgPt-602039 | 12.4 | 11.2 | 16.0 | 11.1 | -1.8 | -3.0 |

**Appendix 29: Additive, dominance effect and the residual variance after fitting the QTL for interval mapping analysis, using the $F_2$ generation data of narrow cross in TCRU**

| Traits | Linkage group | Position (cM) | Locus | mu_A | mu_H | mu_B | Variance | Additive | Dominance |
|---|---|---|---|---|---|---|---|---|---|
| *Terminal leaflet length (TLL)* | 8 | 0 | Bam2coL63 | 9.0 | 9.7 | 10.2 | 2.0 | -0.6 | 0.1 |
| *Terminal leaflet width (TLW)* | 5 | 74.2 | bgPt-595387 | 4.4 | 4.5 | 3.7 | 0.6 | 0.3 | 0.5 |
| *Plant spread (PS)* | 4 | 33.5 | bgPabg-597624 | 60.6 | 62.3 | 49.6 | 125.4 | 5.5 | 7.2 |
| *Pod no./plant (PN)* | 1 | 68 | - | 77.2 | 141.6 | 147.7 | 5440.3 | -35.3 | 29.1 |
| | 1 | 72.7 | bgPt-601022 | 86.5 | 134.3 | 149.5 | 5556.3 | -31.5 | 16.4 |
| *Double seeded pods/plant (DPN)* | 4 | 33.5 | bgPabg-597624 | 3.1 | 3.0 | 1.6 | 1.8 | 0.7 | 0.7 |
| *Seed no./plant (SEN)* | 1 | 72.6 | bgPt-601022 | 83.8 | 122.1 | 145.3 | 5340.0 | -30.7 | 7.6 |
| *Biomass dry weight (BDW)* | 1 | 67 | - | 63.5 | 117.5 | 119.0 | 3889.4 | -27.8 | 26.2 |
| | 1 | 59.6 | bgPabg-596618 | 62.5 | 119.0 | 114.8 | 3796.6 | -26.1 | 30.3 |
| *Shelling% (SH%)* | 12 | 47.5 | bgPt-595486 | 69.4 | 73.3 | 77.4 | 33.4 | -4.0 | -0.1 |
| *100-seed weight (HSW)* | 11 | 0 | bgPabg-595822 | 37.5 | 31.7 | 28.3 | 115.4 | 4.6 | -1.2 |
| | 12 | 47.5 | bgPt-595486 | 28.4 | 33.5 | 37.6 | 115.6 | -4.6 | 0.5 |

**Appendix 30: Additive, dominance effect and the residual variance after fitting the QTL for interval mapping analysis, using the $F_2$ generation data derived from the 'wide' cross of DipC x VSSP11**

| Trait | Linkage group | Position (cM) | Locus | mu_A | mu_H | mu_B | Variance | Additive | Dominance |
|---|---|---|---|---|---|---|---|---|---|
| *Leaf area* (LA) | 15 | 83.1 | - | 396.3 | 317.1 | 434.5 | 3586.1 | -19.1 | -98.3 |
| | 15 | 94.6 | P19b105AGC+CTA2 | 384.6 | 318.9 | 429.6 | 3714.9 | -22.5 | -88.3 |
| *Specific leaf area* (SLA) | 10 | 7 | - | 133.0 | 149.5 | 151.6 | 199.4 | -9.3 | 7.2 |
| | 10 | 12.4 | P16b329AGA+CTC2 | 136.5 | 148.4 | 152.1 | 203.8 | -7.8 | 4.1 |
| *Stem no./plant* (STN) | 13 | 1.1 | P17b242ACG+CAA3 | 9.5 | 10.6 | 14.8 | 14.7 | -2.6 | -1.5 |
| | 1 | 20.9 | 595196 | 9.8 | 10.9 | 14.0 | 16.7 | -2.1 | -1.0 |
| *Internode length* (IL) | 9 | 0 | 600900 | 11.7 | 8.5 | 8.2 | 8.2 | 1.8 | -1.4 |
| | 14 | 26.5 | P19b102AGC+CTA1 | 11.2 | 9.5 | 8.0 | 8.4 | 1.6 | -0.1 |
| | 13 | 0 | P17b240ACG+CAA2 | 10.8 | 9.7 | 7.8 | 8.4 | 1.5 | 0.4 |
| *Carbon isotope discrimination* CID | 7 | 93.9 | P19b494AGC+CTA4 | 21.0 | 19.6 | 21.0 | 0.9 | 0.0 | -1.4 |
| *100-seed weight* (HSW) | 7 | 87.6 | - | 49.4 | 37.5 | 54.7 | 120.7 | -2.6 | -14.5 |
| | 7 | 93.9 | P19b494AGC+CTA4 | 43.6 | 41.1 | 51.7 | 122.5 | -4.0 | -6.6 |