Genetic diversity in rice (*Oryza sativa*) and estimation of outcrossing rate using morphological markers

by

Saidu Bah

A thesis submitted in accordance with

the academic requirements for the degree

of

Philosophiae Doctor

in the

Department of Plant Sciences (Plant Breeding)

Faculty of Natural and Agricultural Sciences at the

University of the Free State

Bloemfontein, South Africa

January 2015

Supervisor: Prof Maryke T Labuschagne (PhD)

Co-supervisors: Dr Kayode Sanni (PhD)

Dr Rouxléne van der Merwe (PhD)

Declaration

I declare that the thesis submitted in fulfilment of the requirement for the degree Philosophiae Doctor in the Faculty of Natural and Agricultural Sciences (Plant Breeding), University of the Free State, Bloemfontein represents my own original, independent work and that I have not previously submitted the same work for a qualification at another university. I further cede copyright of the thesis to the University of the Free State

Saidu Bah

Acknowledgements

Praise be to Allah!

I wish to extend my sincere gratitude to Prof Maryke of the Plant Breeding division of the University of the Free State (UFS) for the excellent tutoring supervision. I learnt so much from you; both academic knowledge and values of modesty and perseverance. I also wish to thank Dr Rouxléne van der Merwe for the invaluable comments and guidance in the thesis write-up. I am also highly grateful to Ms Sadie Geldenhuys for all the support in especially making this study and my life at UFS and South Africa very comfortable. Please keep doing what you do best; bringing happiness into peoples' lives.

My gratitude equally goes to the Africa Rice Center (AfricaRice) for the funding and provision of research facilities for this work. I wish to thank my co-supervisor Dr Sanni Kayode. I also say thanks to Dr Marie Noëlle Ndjiondjop and her team at the AfricaRice Biotechnology laboratory for assisting with the genotyping of the collection. I am grateful to Dr Mounirou Sow for assistance with molecular analysis and for exposing me to relevant molecular software. I also say thanks to Mr Arnaud, Ms Marie Goueté, and all staff of the AfricaRice biotechnology laboratory for hosting me and helping out. I highly appreciate the contribution of Mr Nacisse Dossou-Yovo and his team on conducting the field trials. Your contribution helped provide the data on which this work was based.

Sincere thanks to Elhadj Pape Abdoulaye Seck for making this possible. May Allah bless you? I also thank Drs Samuel Bruce-Oliver and Marco Wopereis for the encouragement and advice. I also appreciate the support rendered by Dr Moussa Sie.

Special thanks go to my wife Yassin for the encouragement and keeping of the fort whilst I was away. I wish to thank the kids for their patience. I wish to thank all the people who contributed towards the realisation of this project.

Dedication

A dedication to

Elhadj Dr Papa Abdoulaye Seck,

a visionary and a Pan-Africanist with a golden heart

Table of contents

List of tables

List of figures

Abbreviations

AFLP:	Amplified fragment length polymorphism
AfricaRice:	Africa Rice Center
AHC:	Agglomerative hierarchical clustering
AMOVA:	Analysis of molecular variance
ANOVA:	Analysis of variance
BAC:	Bacterial artificial chromosome
bp:	Base pair(s)
CTAB:	Cetyltrimethylammonium bromide
cp:	Chloroplast
ΔK:	Delta K
D:	Gene diversity
DArT:	Diversity array technology
d*ij*:	Dissimilarity among genotypes *i* and *j*
DNA:	Deoxyribonucleic acid
DUS:	Distinctiveness, uniformity and stability
EDTA:	Ethylene-diaminetetra acetate
F_1:	First filial generation
F_2:	Second filial generation
FAO:	Food and Agriculture Organisation of the United Nations
FAOSTAT:	FAO statistical database
F_{ST}:	Fixation index of sub-population relative to the total population/total fixation index
GxE:	Genotype by environment
GM:	Genetically modified
GRIN:	Germplasm resources information network
GRiSP:	Global Rice Science Partnership
H':	Shannon-Weaver diversity index
H^2:	Broad sense heritability

He:	Gene diversity
Indel:	Insertion-deletion
IRGSP:	International rice genome sequencing project
K:	Potassium
m:	Migration (gene flow) rate
MAF:	Major allele frequency
MAS:	Marker-assisted selection
MASL:	Metres above sea level
MCMC:	Markov Chain Monte Carlo
MT:	Metric tonnes
N:	Nitrogen
Na:	Number of alleles per locus
N_e:	Effective population size
N_em:	Effective immigrants per generation
NERICA:	New Rice for Africa
NJ:	Neighbour joining
P:	Phosphorus
PCA:	Principal component analysis
NTYSYSpc:	Numerical Taxonomy and Multivariate Analysis System
PCR:	Polymerase chain reaction
PIC:	Polymorphic information content
p-VATPase:	V-ATPase B-subunit
QTL:	Quantitative trait loci
R^2:	Coefficient of determination
RAPD:	Random amplified polymorphic DNA
RFLP:	Restriction fragment length polymorphism
RNA:	Ribonucleic acid
rpm:	Revolutions per minute
SAHN:	Sequential agglomerative hierarchical nested

SAM:	*S*-adenosyl methionine synthetase
SNP:	Simple nucleotide polymorphisms
SSR:	Simple sequence repeat(s)
STS:	Sequence tagged site(s)
TE:	Tris-EDTA buffer
UNJM:	Unweighted neighbour joining method
UPGMA:	Unweighted pair group method with arithmetic mean
UV:	Ultraviolet
var.:	Variety
WARDA:	West Africa Rice Development Association
$\sigma^2 g$:	Genotypic variance
$\sigma^2 ph$:	Phenotypic variance

SI units

cm:	Centimetre(s)
g:	Gramme(s)
h:	Hour(s)
ha:	Hectare(s)
kg:	Kilogramme(s)
m:	Metre(s)
µg:	Microgram(s)
µl:	Microlitre(s)
mg:	Milligram(s)
min:	Minute(s)
ml:	Millilitre(s)
mM:	Millimolar(s)
M:	Molar
pH:	Power of hydrogen
s:	Second(s)
t:	Tonne(s)
V:	Volt(s)
v/v:	Volume per volume
w/v:	Weight per volume
°:	Degrees
°C:	Degrees Celsius

Chapter 1

General introduction

Rice (*Oryza sativa* L.*)* is a tropical cereal plant belonging to the family Gramineae (Poaceae) and genus *Oryza*. It consists of two cultivated species and many wild relatives (Khush, 1997; Sweeney and McCouch, 2007; Vaughan *et al*., 2008). However, the classification of the genus is not definite. Present consensus is that there are two cultivated species and some 30 wild relatives of rice (GRIN, 2015).The two cultivated species of rice are *O. sativa* referred to as Asian rice and *O. glaberrima* Steud. known as African rice. The cultivated species are diploid and could have originated from a common ancestor with an AA genome through parallel evolutionary pathways but each having a unique domestication history. There is ongoing debate about the domestication history of rice. One hypothesis depicted the pathway for *O. sativa* as: *O. rufipogon* Griff.→ *O.nivara* S.D. Sharma & Shastry→ *O. sativa* (Londo *et al*., 2006) and that for *O. glaberrima* as: *O. longistaminata* A. Chev. & Roehr→ *O. barthii* A. Chev.→ *O. glaberrima*.

Rice has many varietal forms and has adapted to a diversity of soils and climates within latitudes 55°N and 36°S of the equator (Khush, 1997). Asian cultivated rice (*O. sativa*) is grown all over the world. Domestication might have occurred in the Yangtze River Valley in China some 10000 to 14000 years ago (Vaughan *et al*., 2008; GRiSP, 2010). *Oryza sativa* has two prominent variety groups namely; Indica (which is predominant in tropical regions) and Japonica (which is prevalent in subtropical and temperate regions). There is debate on whether the variety groups originated from single or independent domestication events (Molina *et al*., 2011). African rice, *O. glaberrima,* was probably domesticated separately in the basin of the Upper Niger River in West Africa more than 3500 years ago (Jones *et al*., 1997). *O. glaberrima* is cultivated only in Africa and has been found to be tolerant to prevailing abiotic stresses and resistant to diseases that affect rice in Africa.

Rice is a major food crop for humankind. It is the most important food crop for people in the developing world and a staple food for about half of the world's population. Globally 3.5 billion people depend on it for more than 20% of their daily calorie intake. Rice production and consumption is dominated by Asia (Dawe *et al*., 2010). Rice is also an important crop in Africa. It is becoming increasingly popular as a staple food for most people, especially in urban areas. Rice per capita consumption in Africa has been increasing over the years (Dawe *et al*., 2010). In 2009 a milled equivalent of 19.6 million tonnes were consumed (FAOSTAT, 2013). Rice production has increased rapidly in Africa. This is largely due to expansion in cultivated area and the adoption of

new and improved varieties with accompanying complementary technologies. From 2001–2005 rice production in Africa showed a growth rate of 5.81% with an annual average of 8.1 million tonnes of milled rice. West Africa and East Africa accounted for more than 95% of total rice production (WARDA, 2008). In sub-Saharan Africa, West Africa is the leading rice producing region. Notwithstanding, rice consumption has increased at an even faster rate, thus making the continent a net importer of rice.

Rice is a semi-aquatic plant which is suitably adapted to saturated soil conditions. However it can grow under varying conditions along the topo sequence from unsaturated soil conditions typical of the upland ecosystem to semi-saturated typical of hydromorphic conditions to the saturated valley bottom. It is a self-pollinated plant. Pollen shedding usually coincides with the opening of the flower and most pollen will fertilise a stigma in the same flower. The flower opens only once for about one hour. Pollen grains are viable for about 10–20 min whilst the stigma remains receptive for several days. The viability of pollen is dependent on humidity and exposure to ultraviolet (UV) radiation (Khush, 1997; Gealy *et al*., 2003). Temperature is also critical to the survival of rice pollen (Matsui *et al*., 2000; 2001; Jagadish *et al*., 2007; Weerakoon *et al*., 2008). Khush (1997) identified four major rice agro-ecosystems namely (a) Irrigated lowland, (b) Rainfed lowland, (c) Upland and (d) Flood-prone. The relative importance of each of these systems varies according to production region.

Self-pollination is an advantage in rice seed production of non-hybrid varieties as it reduces chances of outcrossing with an undesirable donor. It helps maintain genetic purity and by implication eases the production of quality seed. However, some level of natural outcrossing does occur in rice. This happens when the stigma receives pollen from a different flower on the same plant or from a different plant. Outcrossing in rice is mainly due to the movement of pollen by wind. When outcrossing occurs with an undesirable donor, contamination occurs and off-types are produced (Da Silva *et al*., 2005; Nuijten *et al*., 2009; Lee *et al*., 2013). This could lead to undesirable gene flow between the off-types and other plants in subsequent generations. This is a critical issue in the production of quality rice seed as it reduces the genetic purity of the seed crop. Outcrossing can also lead to the introduction of genes into non-target plants. For instance, the introduction of herbicide resistant genes into weeds will make the latter resistant to herbicides and more difficult to control (Ellstrand *et al*., 1999; Rong *et al*., 2007; Chandler and Dunwell 2008; Kumar *et al*., 2008). It is important to understand the nature and levels of rice to rice outcrossing under field and environmental conditions that prevail in Africa. Research results will contribute to

formulation of strategies aimed at avoiding undesirable outcrossing and gene exchange (Song *et al*., 2003; Kuroda *et al*., 2005).

In recent years there has been an expansion in the production of rice in sub-Saharan Africa. This has been largely due to the introduction of appropriate varieties like NERICA (New Rice for Africa) and other improved technologies aimed at improving productivity. Increase in area is equally true for other ecologies as several governments have put in place policies that are conducive to the production of rice. NERICA varieties have been found to be compatible with *O. sativa* (Ikeda *et al*., 2009). This compatibility could be an advantage as NERICA and other interspecific varieties could serve as bridge cultivars to transfer useful gene(s) from *O. glaberrima* to *O. sativa* and *vice versa*. It is equally a disadvantage as it is possible to receive pollen from unwanted sources.

The issue of undesirable outcrossing has become salient with the recent introduction of genetically modified (GM) crops. There is a consistent increase in the use of GM crops worldwide. Global coverage by GM crops in 2012 was 170.3 million ha in 28 countries. This is predicted to be 200 million ha in 2015 (ISAAA Brief, 2012). GM rice is also gaining increased prominence in terms of research and coverage. GM rice can contribute to increased food security. It has also led to reduced usage of pesticides, therefore reducing the hazards related to the release of poisonous chemicals into the environment (Brookes and Barfoot, 2012). But there is also the potential for gene flow between GM rice and non-GM rice like conventional varieties and other wild relatives of rice. Therefore, it is necessary to quantify gene flow between rice varieties as part of routine biosafety risk assessment.

Various studies have indicated that outcrossing occurs between and within different rice species (Chen *et al*., 2004; Endo *et al*., 2009; Shivrain *et al*., 2009; Somaratne *et al*., 2012). Outcrossing is mainly effected by movement of pollen by wind. Insects have also been found to influence levels of outcrossing in rice (Gealy *et al*., 2003). The rate of outcrossing is generally determined by the proximity in time and space of the pollen donor and the receptive female. Other factors that influence outcrossing rate in rice are environmental conditions at the time of anthesis (Hoyle and Cresswell, 2007; Jagadish *et al*., 2007).

Undesirable outcrossing during rice seed production results in the appearance of off-types and weedy rice. This invariably leads to contamination, which reduces seed yield and quality. Knowledge on the nature and extent of contamination by natural outcrossing is important for plant breeding and seed programmes. There is a need to periodically quantify its impact over locations and seasons. Such information will be useful for gene banks, research institutes, seed

multiplication and quality control programmes and other bodies whose mandate is to maintain and preserve the genetic purity of rice germplasm (Da Silva *et al*., 2005; Jiang *et al*., 2012).

Whilst several studies have been carried out on levels of outcrossing in rice, all of them were based on genotypes of Asian rice *O. sativa*. They did not include rice of African origin like *O. glaberrima.* Nor did they include interspecific hybrids (to which the NERICA varieties belong); and landraces in Africa. At present, there is no information available on the outcrossing rate of interspecific hybrids. Environments and genotypes in other studies are different from those in Africa.

The goal of this research was to estimate levels of natural outcrossing in interspecific (*O. sativa* x *O. glaberrima*) and intraspecific (*O. sativa*) rice under field conditions in Africa using morphological markers. In order to identify suitable markers and select appropriate genotypes for the outcrossing trial, an evaluation of diversity of a collection of accessions was carried out. Based on results from the diversity study, accessions were divided into groups and optimum planting dates for the pollen donor and recipient plants with suitable traits were identified. Planting dates were selected with the aim to synchronise flowering between the pollen donor and recipient plants.

Knowledge on the nature and rate of outcrossing will help in reducing undesirable outcrossing and the production of off-types, which will result in good quality rice seed. It will also provide information on biosafety risk assessment and what can be done to alleviate related problems.

The objectives of the study were to:

1. Assess the genetic variability of 36 rice genotypes to identify suitable genotypes and morphological markers for evaluating outcrossing
2. Estimate the rate of natural outcrossing in rice under field conditions in Africa using agro-morphological markers
3. Determine the influence of distance and genotype on outcrossing

References

Brookes G, Barfoot P. 2012. Global impact of biotech crops: Environmental effects, 1996–2010. GM crops and food: Biotechnology in Agriculture and the Food Chain 3:129–137.

Chandler S, Dunwell JM. 2008. Gene flow, risk assessment and the environmental release of transgenic plants. Critical Reviews in Plant Sciences 27: 25–49.

Chen LJ, Lee DS, Song ZP, Suh HS, Lu BR. 2004. Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Annals of Botany 93:67–73.

Da Silva EF, Da Silva LM, Montalvan R. 2005. Crossing rate and distance in upland rice. Bragantia, Campinas 64:197–201.

Dawe D, Pandey S, Nelson A. 2010. Emerging trends and spatial patterns of rice production. In: Pandey S, Byerlee D, Dawe D, Doberman A, Mohanta S, Rozelle S, Hardy B. (Eds.). Rice in the Global Economy: Strategic Research and Policy Issues for Food Security. International Rice Research Institute (IRRI). pp 15–35.

Ellstrand NC, Prentice HC, Hancock JFE. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539–563.

Endo T, Sato H, Yamaguchi M, Kataoka T, Nakagomi K, Ito T, Mori K. 2009. Estimate of outcrossing rates in a rice plant (*Oryza sativa* L.) under field conditions using a purple grain rice cultivar, *Okunomurasaki*. Breeding Science 59:195–202.

FAOSTAT. 2013. http://faostat3.fao.org/faostat-gateway/go/to/download/C/CC/E (Accessed 15/07/2014)

Gealy D, Mitten D, Rutger N. 2003. Gene flow between red rice and herbicide resistant rice: implications for weed management. Weed Technology 17:627–645.

GRIN (Germplasm Resources Information Network). 2015. USDA, ARS, National Genetic Resources Program. (Online Database). National Germplasm Resources Laboratory, Beltsville, Maryland. http://www.ars-grin.gov.4/cgi-bin/npgs/html/taxdumplist.pl (Accessed 11/09/2015).

GRiSP (Global Rice Science Partnership). 2010. CGIAR Thematic Area 3: Sustainable crop productivity increase for global food security. A CGIAR Research Program on Rice-Based Production Systems. November 2010.

Hoyle M, Cresswell, JE. 2007. The effect of wind direction on cross-pollination in wind-pollinated GM crops. Ecological Applications 17:1234–1243.

Ikeda R, Sokei Y, Akintayo I. 2009. Seed fertility of F_1 hybrids between upland rice NERICA cultivars and *Oryza sativa* L. or *O. glaberrima* Steud. Breeding Science 59:27–35.

ISAAA Brief. 2012 Executive Summary. Global Status of Commercialized Biotech/GM Crops: 2012. http://www.isaaa.org/resources/publications/briefs/44/executivesummary/ (Accessed 02/08/2014)

Jiang Z, Xia H, Basso B, Lu BR. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. Theoretical and Applied Genetics 124:309–322.

Jagadish SVK, Craufurd PQ, Wheeler TR. 2007. High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). Journal of Experimental Botany 58:1627–1635.

Jones MP, Dingkuhn M, Aluko GK, Semon M. 1997. Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. Euphytica 92:237–246.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25–34.

Kumar V, Bellinder RR, Brainard DC, Malik RK, Gupta RK. 2008. Risks of herbicide-resistant rice in India: A review. Crop Protection 27:320–329.

Kuroda Y, Sato YI, Bounphanousay C, Kono Y, Tanaka K. 2005. Gene flow from cultivated rice (*Oryza sativa* L.) to wild Oryza species (*O. rufipogon* Griff. and *O. nivara* Sharma and Shastry) on the Vientiane plain of Laos. Euphytica 142:75–83.

Lee J, Park JH, Koh HJ. 2013. Morphological and genetic characterization of off-type rice plants collected from farm fields in Korea. Journal of Plant Biology 56:160–167.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. PNAS 103:9578–9583.

Matsui T, Omasa K, Horie T. 2000. High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (*Oryza sativa* L.). Plant Production Science 3:430–434.

Matsui T, Omasa K, Horie T. 2001. The difference in sterility due to high temperatures during the flowering period among *japonica* rice varieties. Plant Production Science 4:90–93.

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. PNAS 108: 8351–8356.

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P. 2009. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers‘ fields. PLoS ONE 4:1–9.

Rong J, Lu BR, Song Z, Su J, Snow AA, Zhang X, Sun S, Chen R, Wang F. 2007. Dramatic reduction of crop-to-crop gene flow within a short distance from transgenic rice fields. New Phytologist 173:346–353.

Shivrain VK, Burgos NR, Sales MA, Mauromoustakos A, Gealy DR, Smith KL, Black, HW, Jia M. 2009. Factors affecting the outcrossing rate between Clearfield™ rice and red rice (*Oryza sativa*). Weed Science 57:394–403.

Somaratne LHMYK, Abayawickrama ASMT, Wickramasinghe IP, Samarasinghe WLG. 2012. Estimating out-crossing rate of Bg 379-2 using morphological markers and confirmation by molecular markers. Rice Science 19:166–168.

Song ZP, Lu BR, Zhu YG, Chen JK. 2003. Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657–665.

Sweeney M, McCouch S. 2007. The complex history of the domestication of rice. Annals of Botany 100:951–957.

Vaughan DA, Lu BR, Tomooka N. 2008. The evolving story of rice evolution. Plant Science 174:394–408.

WARDA. 2008. Africa Rice Trends, 2007.

Weerakoon WMW, Abeywickrama T, de Costa WAJM, Maruyama A. 2008. Out-crossing of heat stress affected spikelets of lowland rice in the sub-humid zone of Sri Lanka and its long-term implications. http://www.niaes.affrc.go.jp/marco/marco2009/english/W2-05_Weerakoon_WMW_P.pdf. (Accessed 21/06/2012)

Chapter 2

Rice as a crop and an outcrossing self-pollinating plant: associated risks and benefits

2.1 Abstract

Rice (*O. sativa*), is a tropical cereal crop belonging to the family Gramineae which is adapted to both upland and lowland ecologies and is grown worldwide. It is an important food crop for developing countries. There are two cultivated species and several wild and weedy relatives. The two cultivated species are *O. sativa* and *O. glaberrima*. *Oryza sativa* is believed to have originated from Asia and *O. glaberrima* from Africa. Rice is a self-pollinating plant that has been shown to exhibit outcrossing under field conditions.

This review focused on rice as a plant and crop. Its origin, taxonomy, diversity, distribution and domestication are presented. Several methods of estimating genetic diversity are also presented and discussed. The role and influence of rice as a food crop for humankind is also highlighted. The status of *O. sativa* as a self-pollinating plant that can outcross with close relatives is discussed. Implications of rice outcrossing with regards to gene flow and the impact on rice seed quality and the environment are also discussed.

Keywords: Diversity, domestication, gene flow, *O. glaberrima*, *O. sativa*, origin, outcrossing, rice.

2.2 Introduction

Rice is a tropical cereal plant belonging to the family Gramineae (Poaceae). It is one of the major food crops of the world and the most important food crop for people in developing countries. It is the staple food for about half of the world's population comprising mainly of the poor and vulnerable. Globally about 3.5 billion people (Figure 2.1) depend on rice for more than 20% of daily calorie intake. Global rice demand is expected to be 555 million tonnes in 2035 (GRiSP, 2010). In addition rice has become a model system for plant biology as the entire genome has been sequenced by the International Rice Genome Sequencing Project (IRGSP) and is now available (IRGSP, 2005).

Global production of rice was estimated at 741 million tonnes in 2013, covering an area of 165 million ha (FAOSTAT, 2015). Asia dominates cultivation, production and consumption of rice (Dawe *et al.,* 2010). Rice is also important in Africa, the Americas and the Caribbean nations. It is the staple food for many countries in Africa (e.g. Gambia, Guinea, Sierra Leone, Madagascar); and is rapidly replacing other foods in non-traditional rice eating countries. For instance, in Ivory

Coast, rice calorie intake rose from 12% in 1961 to 22% in 2004. In 2013 rice coverage in Africa was estimated at 10.9 million ha with a production of 28.7 million tonnes (FAOSTAT, 2015).

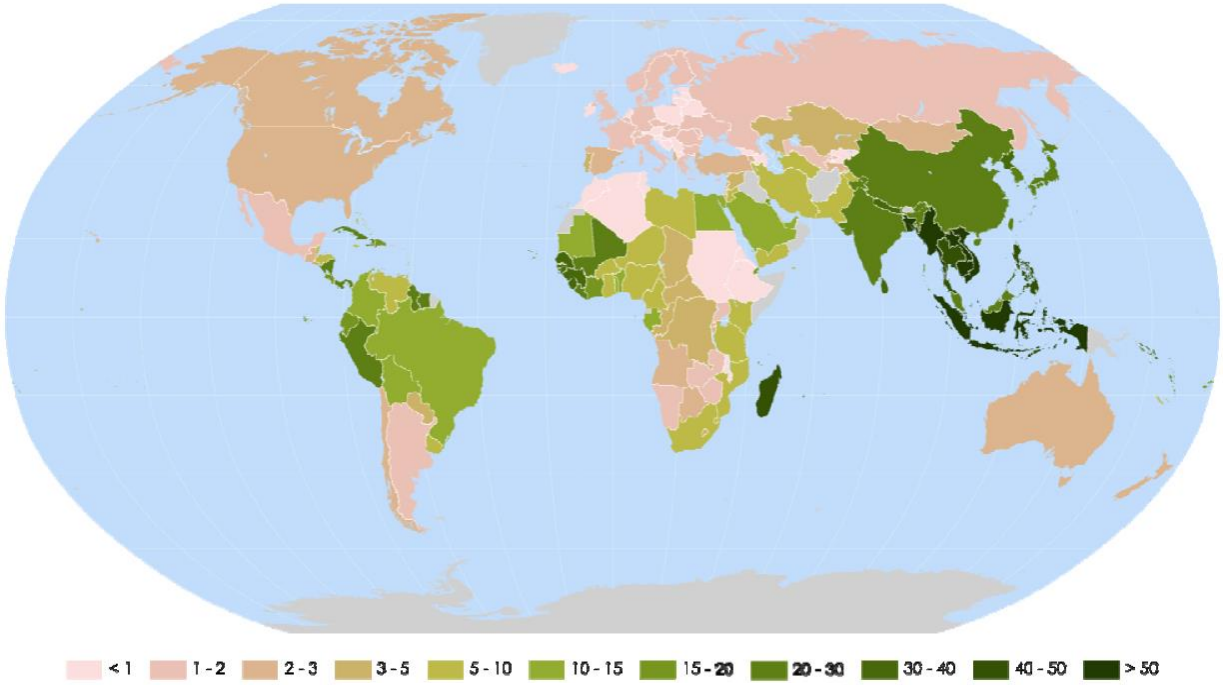


Figure 2.1 Share of rice total calories consumed per country (GRiSP, 2010)

2.3 Taxonomy and geographical distribution of rice

The genus *Oryza* comprises of two cultivated species and a range of crop wild relatives (Khush, 1997; Sweeney and McCouch, 2007; Vaughan *et al.,* 2008; GRIN, 2015). Cultivated species are different, each with a distinct history of domestication. Wild species are distributed in the humid tropics and subtropics that are found in Africa, Asia, the Americas and Australia. Wild species have both diploid (2n = 2x = 24) and tetraploid (2n = 4x = 48) forms whilst cultivated species are all diploid. The taxonomy of the AA genome *Oryza* species has divergent views by contributing scholars. Species differentiation has been based on origin or geography, whether the species is annual or perennial in habit and whether it is in a cultivated or wild habitat (Vaughan *et al*., 2008; Table 2.1). The nomenclature has changed over time, depending on the proponent(s). Generally four species complexes are recognised: (i) *O. sativa,* (ii) *O. officinalis,* (iii) *O. ridleyi* and (iv) *O. granulata*. Interspecific crossing is possible between species, but it is difficult for such crossings to result in fertile progeny (Morishima, 2001; Vaughan *et al*., 2003; 2008; Sweeney and McCouch, 2007).

The *O. sativa* complex contains the two cultivated species (Morishima, 2001; Sweeney and McCouch, 2007; GRIN, 2015). GRIN (2015) recognises two cultivated species namely *O. sativa* and *O. glaberrima* also known as Asian rice and African rice respectively with seven primary wild relatives namely*; O. barthii*, *O. glaberrima* (wild types) *O. glumaepatula*, *O. longistaminata*, *O. meridionalis* Ng, *O. nivara*, and *O. rufipogon*. Secondary wild relatives are: *O. alta* Swallen, *O. australiensis* Domin, *O. brachyantha* A. Chev. & Roehr., *O.* eichingeri Peter, *O. grandiglumis* (Döll) Prodoehl, *O latifolia* Desv., *O. malampuzhaensis* Krishnasw. & Chandrasekh., *O. minuta* J. Presl, *O. officinalis* Wall. ex G. Watt, *O. punctata* Kotschy ex Steud., *O. rhizomatis* D. A. Vaughan, and *O. schweinfurthiana* Prodoehl. Tertiary wild relatives are: *O. coarctata* Roxb., *O. longiglumis* Jansen, *O. meyeriana* (Zoll. & Moritzi) Baill., *O. meyeriana* (Zoll. & Moritzi) Baill. var. *granulata* (Nees & Arn. ex G. Watt) Duist., *O. meyeriana* (Zoll. & Moritzi) Baill. var. *inandamanica* (J. L. Ellis) Veldkamp, *O. meyeriana* (Zoll. & Moritzi) Baill. var. *meyeriana*, *O. neocaledonica* Morat, *O. ridleyi* Hook. f. and *O. schlechteri* Pilg. (GRIN, 2015).

Table 2.1 Distribution, life cycle and cultivation status of AA genome *Oryza* species (Vaughan *et al*., 2008)

Life cycle	Cultivation status	Africa	Latin America	Asia	Australia and New Guinea
Perennial	Wild	*O. longistaminata* A. Chev. et Roehr.	*O. glumaepatula* Steud.		*O. rufipogon* Griff
Annual	Wild	*O. barthii* A. Chev		*O. nivara* Sharma and Shastry	*O. meridionalis* Ng
Annual	Cultivated	*O. glaberrima* Steud.		*O. sativa* L. (worldwide)	

Vaughan *et al*. (2003) collated numerous classifications by several rice scientists and presented a clear and concise nomenclature (Table 2.2).

Oryza sativa is cultivated throughout the tropics, subtropics and warm temperate regions of the world. *Oryza glaberrima* is cultivated in West Africa and West Central tropical Africa (GRIN, 2015). *Oryza rufipogon* (a perennial) is found in tropical continental Asia and Oceania. *Oryza barthii* (an annual) is predominant in West Africa but also found in East, North East and Southern Africa. *Oryza longistaminata* (a perennial) is found in mainland Africa and in the Western Indian Ocean

island of Madagascar. *Oryza barthii* is distinguished from other species by its short ligule and *O. longistaminata* by its rhizomatous habit. *Oryza meridionalis* is native to Australia and *O. glumaepatula* is endemic in Central and South America (Morishima, 2001; Sweeney and McCouch, 2007). There are no distinct morphological traits that distinguish *O. glumaepatula* from AA-genome wild rice species found in Asia and Australia. However, there are sterility barriers between them.

Table 2.2 *Oryza* species: chromosome number, genome group and usual habitat (Vaughan *et al*., 2003)

Section complex	Chromosome number	Genome group	Habitat
***Oryza sativa* complex**			
O. sativa L.	24	AA	Upland to deep water; open
O. rufipogon Griff.	24	AA	Annual: seasonally dry; open Perennial: seasonally deep water and wet year round; open
O. glaberrima Steud.	24	AA	Upland to deep water; open
O. barthii A. Chev.	24	AA	Seasonally dry; open
O. longistaminata Chet et Roehr.	24	AA	Seasonally dry to deep water; open
O. meridionalis Ng.	24	AA	Seasonally dry; open
O. glumaepatula Steud.	24	AA	Seasonally dry inundated areas; open
***O. officinalis* complex**			
O. officinalis Wall ex Watt	24	CC	Seasonally dry; open
O. minuta JS Presl. ex CB Presl	48	BBCC	Stream sides; semi shade
O. rhizomatis Vaughan.	24	CC	Seasonally dry; open
O. eichingeri Peter.	24	CC	Stream sides; forest floor; semi shade
O. malapuzhaensis Krishnaswamy and Chandrasakaran	48	BBCC	Seasonally dry forest pools; shade
O. punctata Kotschy ex Steud.	24	BB	(Diploid) seasonally dry; open
	48	BBCC	(Tetraploid) forest floor; shade
O. latifolia Desv.	48	CCDD	Seasonally dry; open
O. alta Swallen.	48	CCDD	Seasonally inundated; open

Table 2.2 Continued. *Oryza* species: chromosome number, genome group and usual habitat (Vaughan *et al*., 2003)

Section complex	Chromosome number	Genome group	Habitat
O. grandiglumis (Doell.) Prod.	48	CCDD	Seasonally inundated; open
O. australiensis Domin.	24	EE	Seasonally dry; open
Ridleyanae Tateoka			
O. schlechteri Pilger	48	Unknown	River banks; open
***O. ridleyi* complex**			
O. ridleyi Hook.	48	HHJJ	Seasonally inundated forest floor; shade
O. longiglumis Jansen.	48	HHJJ	Seasonally inundated forest floor; shade

The close similarity between *O. rufipogon* and *O. nivara* has resulted in *O. nivara* being considered as an ecotype of *O. rufipogon* (Morishima, 2001; Vaughan *et al*., 2003; Kwon *et al*., 2006; Zhou *et al*., 2008; Zheng and Ge, 2010). The two species are cross compatible with fertile progeny and the variation between them is continuous. However, some rice scientists consider *O. rufipogon* and *O. nivara* as distinct species (Duan *et al*., 2007; Kuroda *et al*., 2007; Xu *et al*., 2012; Banaticla-Hilario *et al*., 2013). Early taxonomic studies looked at gene analysis to point out interspecies relationships (Sano and Sano, 1990; Hirano *et al.,* 1994; Kanazawa *et al*., 2000). However, this approach is not conclusive as gene lineage does not imply speciation. Later studies, using molecular markers have been carried out to elucidate phylogenetic relationships between taxa (Kawakami *et al*., 2007; Gao and Innan, 2008; Molina *et al.,* 2011; Huang *et al.,* 2012). Findings were however, not consistent between the different studies.

Morishima (2001) pointed out four directions for differentiation within the gene pool of *O. sativa* namely; (i) differentiation from wild to cultivated types, (ii) differentiation from perennial to annual types in wild races, (iii) geographical differentiation in wild races and (iv) varietal differentiation toward Indica and Japonica groups. Asian rice comprises two major groups, Indica and Japonica, traditionally referred to as *hsien* and *keng* in China. Glaszmann (1987) in his classification of Asian germplasm using isozymes identified two major groups; Indica and Japonica and four other groups that were affiliated to the main ones. The Indica group is typically lowland rice grown in tropical Asia, while Japonica is grown in South Asia. A subsequent study by Garris *et al*. (2005) using simple sequence repeat (SSR) markers, confirmed earlier isozyme findings.

Oryza glaberrima is mainly found in West Africa and shows less diversity than *O. sativa*. Studying the population structure of 198 accessions of *O. glaberrima*, Semon *et al.* (2005) utilised 93 nuclear SSR markers and then used a combination of Bayesian clustering, principal coordinate analysis and a Mantel test to assess the hypothesis of genetic isolation by genetic distance. They identified three distinct groups namely, I, II and III and reported a gene diversity (He) value of 0.27. Another study by Li *et al*. (2011) assessed genetic diversity in 40 accessions consisting of *O. glaberrima* and *O. barthii* by investigating nucleotide variation in 14 unlinked nuclear genes. They reported a low level of nucleotide diversity for both species (θ_{sil} = 0.0007 for *O. glaberrima*; θ_{sil} = 0.0024 for *O. barthii*). The level of diversity could have been higher if a larger sample size had been used.

2.4 Origin and domestication of cultivated species

The two cultivated species, *O. sativa* and *O. glaberrima*, are believed to have been developed through parallel evolutionary pathways. There is a general consensus with regards to the ancestry of *O. glaberrima*, which is depicted as *O. longistaminata* → *O. barthii* → *O. glaberrima*. *Oryza longistaminata* is a wild perennial whilst *O. barthii* is a wild annual. However, there is still debate over that of the much studied *O. sativa.* There are different opinions as to whether *O. rufipogon* or *O. nivara* or possibly both were direct ancestors of *O. sativa.* The debate on *O. sativa* could as well have been the case with *O. glaberrima*, had the latter been subjected to intense studies. Looking at the ancestral pattern of *O. glaberrima* and that which is common with other grasses (i.e. perennial wild → annual wild → annual cultigen) one is inclined to believe that the ancestral history of *O. sativa* is most likely: *O. rufipogon* (perennial wild) → *O. nivara* (annual wild) → *O. sativa* (annual cultigen).

There are two variety groups of *O. sativa* namely; Indica, which is predominant in tropical regions and Japonica, which is prevalent in subtropical and temperate regions. Researchers generally agree that *O. sativa* originated from China and that *O. rufipogon* is the ancestor. However, they disagree on the number of domestication events and the origin within China. Different studies have proposed either single or multiple domestication models (Figure 2.2). A single origin model suggests that the two variety groups, Indica and Japonica were domesticated from a single source of wild rice relative *O. rufipogon,* with differentiation into Indica and Japonica occurring later. In contrast, proponents of the multiple independent domestication model argue that the two variety groups were domesticated separately and from different sources of *O. rufipogon* (Molina *et al*., 2011).

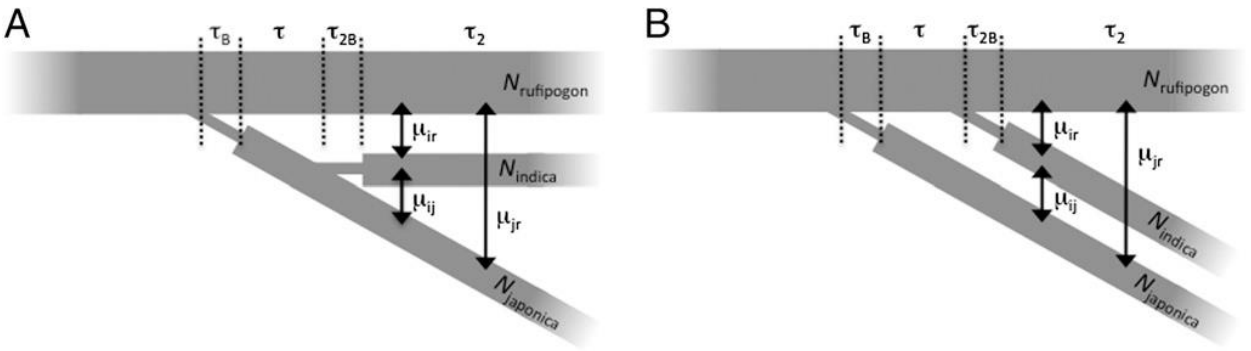


Figure 2.2 Schematics of the single- (A) vs. double- (B) founder models. (A) In the single domestication event, both domesticated variety groups originated from the same *O. rufipogon* ancestral population. (B) In the double-founder model, *indica* and tropical *japonica* were domesticated independently from different *O. rufipogon* populations: *O. rufipogon, O. sativa indica*, and *O. sativa*. tropical *japonica* are indicated by the subscripts *r*, *i*, and *j*, respectively. The times τ_B and τ represent the length of the bottleneck and time thereafter during the two-population epoch. Likewise, τ_{2B} and τ_2 represent the length of time of the bottleneck and time thereafter for the three-population epoch. Symmetric migration (μ) between populations is represented by arrows, and N is the population size (Molina *et al*., 2011).

Huang *et al*. (2012) in characterising the phylogeny of *O. rufipogon* populations surveyed sequence variation at 42 genome-wide sequence tagged sites (STS) in 108 *O. rufipogon* accessions. Using Bayesian clustering, principal component analysis and analysis of molecular variance (AMOVA), they identified two distinct *O. rufipogon* groups namely Ruf-I and Ruf-II. The two groups exhibit clinal variation from north-east to south-west. Ruf-I is similar to *O. sativa* group Indica and it is found in China and the Indochinese Peninsula. Ruf-II is not similar to any cultivated species and it is found in South Asia and the Indochinese Peninsula. *Oryza sativa* group Japonica was found not to be similar to either Ruf-I or Ruf-II. They propose the single model of rice domestication. Another study by Molina *et al.* (2011), determining the evolutionary history of domesticated rice, resequenced 630 gene fragments on chromosomes 8, 10 and 12 from a set of 36 accessions consisting of *O. rufipogon* and *O. sativa*. Using single nucleotide polymorphism (SNP) patterns they identified 20 putative selective sweeps on the chromosomes. They support a single domestication model based on results from demographic modelling of SNP data. Similarly Gao and Innan (2008) proposed a single origin model for *O. sativa*. They used 60 SSR markers to genotype 92 rice accessions comprising *O. sativa* groups Indica and Japonica and wild relative *O. rufipogon*. Bayesian clustering analysis suggested partial sharing of ancestor populations of domesticated rice or a significant level of outcrossing between them.

The alternative hypothesis of multiple domestication of rice leading to *O. sativa* groups Indica and Japonica has been suggested by mainly molecular studies. A study by Londo *et al*. (2006)

examined deoxyribonucleic acid (DNA) sequence variation in three regions from one chloroplast (*atpB-rbcL*) and two nuclear (*p-VATPase* and *SAM*) gene regions to determine the evolutionary history of cultivated rice. For all gene regions variation was observed within and between *O. sativa* and *O. rufipogon* samples. Study results indicated that there were at least two independent domestication events of domesticated rice from different *O. rufipogon* populations and the products are *O. sativa* group Indica and *O. sativa* group Japonica. Another study by Kawakami *et al*. (2007) determined evolution of rice using 275 accessions of cultivated Asian rice and 44 accessions of AA genome *Oryza* species. Accessions were classified into eight chloroplast (cp) genome types (A-H) based on insertion-deletion (indel) events at three regions (8K, 57K and 76K) of the cp genome. Results supported a polyphyletic origin of cultivated Asian rice from at least four principal lineages in the *O. rufipogon – O. nivara* complex. Earlier studies also indicated multiple domestication of cultivated rice *O. sativa* (Bautista *et al*., 2001; Yamanaka *et al*., 2003; Cheng *et al*., 2003; Vitte *et al*., 2004).

Asian rice has many varietal forms with the two main forms being Indica and Japonica. Studies have shown the two groups to be genetically different (Rakshit *et al*., 2007; Kumagai *et al*., 2010). Rice has adapted to a diversity of agro-ecologies within latitudes 55°N and 36°S of the equator (Khush, 1997). Varieties of *O. sativa* (Asian rice) are classified into six groups that are also invariably referred to as subspecies. *Oryza sativa* group Indica corresponds to group I and *O. sativa* group Japonica to group VI (Glaszmann, 1987; Garris *et al*., 2005). *Oryza sativa* group Japonica is further subdivided into temperate and tropical Japonica. *Oryza sativa* group Indica was probably domesticated in India at the foot of the Himalayas whilst *O. sativa* group Japonica was domesticated in China (Khush, 1997).

Asian rice was probably domesticated in the Yangtze River Valley in China some 10000 to 14000 years ago (Vaughan *et al.,* 2003; Sweeney and McCouch, 2007; GRiSP, 2010). On the contrary, studies by Wei *et al.* (2012) indicated that Asian rice originated in southern China along the Pearl River Basin near the Tropic of Cancer. *Oryza glaberrima* is believed to have been domesticated separately in the basin of the Upper Niger River in West Africa more than 3500 years ago (Portères (1970; Jones *et al*., 1997; Linares, 2002; Sarla and Swamy, 2005; Wang *et al*., 2014) and then spread to two secondary centres along the Sahelian rivers (the rivers of Niger, Senegal, Gambia and their tributaries).

Oryza glaberrima originated from wild relative *O. barthii*. This single origin model for *O. glaberrima* has been supported by more recent genetic studies of the genus. Wang *et al*. (2014) investigated

the genome sequence of *O. glaberrima* (var. CG14) using minimum tilling path of 3485 bacterial artificial chromosomes (BAC) clones. They also investigated the relationship between *O. glaberrima* and wild relatives by resequencing 94 *O*. *barthii* accessions. Results from neighbour joining (NJ) tree constructed from combined *O. glaberrima* and *O. barthii* SNP data indicated that *O. glaberrima* was domesticated from *O. barthii*. Single domestication of *O. glaberrima* from wild relative *O. barthii* was also proposed by the study of Li *et al*. (2011), which evaluated nucleotide variation in 14 unlinked nuclear genes in accessions of *O. glaberrima* and wild relative *O. barthii*.

Evidence suggesting that African and Asian rice farmers selected similar or different traits during the process of domestication has been suggested by different studies. Gross *et al*. (2010) reported the presence of red pericarp gene *Rc* in both *O. glaberrima* and *O. sativa*. However, each species shows a distinct and independent *Rc* gene profile. This suggests a convergent but independent domestication of the two species. In another study Sanyal *et al*. (2010) indicated the flowering gene *Hd1* which is prominent in *O. sativa* and *O. barthii* but is absent in *O. glaberrima*.

2.5 Diversity of cultivated rice

Asian cultivated rice has two main varietal groups with several other subgroups. Glaszmann (1987) evaluated genetic structure of 1688 Asian rice accessions by looking at variation of 15 polymorphic loci coding for eight enzymes. Multivariate analysis of the data revealed six varietal groups: I-VI. There were two major groups: I and VI which correspond to Indica and Japonica groups respectively of *O. sativa*. Group I is found in the whole of tropical Asia and is dominant in the south of the Indian subcontinent. Group VI is found in temperate areas and high elevation areas in South Asia. Huang *et al*. (2012) indicated that one of the *O. rufipogon* groups identified in their study (Ruf-1) was genetically similarly to *O. sativa* group Indica. Group II is found exclusively in South and West Asia. Group II varieties are early maturing and can be grown in irrigated as well as strictly upland conditions. Group III consists of varieties that are early maturing, photoperiod insensitive and adapted to deep water conditions. Group III is found exclusively in Bangladesh and parts of India (Glaszmann 1987).

Group IV consists of tall and late maturing rice genotypes. Accessions in this group generally elongate to 6 m and mature in 12 months. They are tolerant to cold and flooding and are photoperiod sensitive. Group V consists of high quality aromatic and long grain rice. High value Sadri and Basmati rice belong to this group. This classification of Asian rice was largely supported by the study of Garris *et al*. (2005) who evaluated genetic structure of rice by genotyping 234

accessions of rice using 169 nuclear SSR markers and two chloroplast loci. They identified five groups namely: indica, aus, aromatic, temperate japonica, and tropical japonica rice.

Oryza glaberrima and wild relative *O. barthii* have been reported to have low genetic diversity (Li *et al*., 2011). African cultivated rice has so far not been subdivided into variety groups. However, the study of Semon *et al*. (2005) which used SSR markers and a model based approach to evaluate population structure of 198 *O. glaberrima* and nine *O. sativa* accessions identified five genetically distinct groups. Three of the groups (Groups 1, 2 and 3) had characteristics of *O. glaberrima* whilst two groups (Groups 4 and 5) had characteristics of Indica and Japonica groups respectively of *O. sativa*.

2.6 Diversity estimation methods

Several techniques have been developed to evaluate variability in plants. These techniques have been divided into three broad groups (Semagn *et al*., 2006).

- Phenotypic markers based on agro-morphological trait differences between plants

- Biochemical markers based on proteins or their activity as detected through gene expression

- Molecular markers based on the detection of DNA polymorphisms between individuals

2.6.1 Phenotypic markers

Phenotypic markers, also known as agro-morphological markers, were the earliest known markers to be used in plant diversity studies. They are based on the fact that constituent genes of an individual are usually expressed in the phenotype. Tateoka (1963) used spikelet and awn length to differentiate wild forms within the *Oryza* complex based on their geographical origin. Agro-morphological traits were also used to estimate the genetic diversity in African rice species (Chang *et al*., 1977; Second *et al*., 1977). They are still widely used as markers in rice diversity studies (Nuijten and Almekinders, 2008; Sanni *et al.*, 2008; Babaei *et al*., 2011; Willocquet *et al*., 2012; Dong *et al*., 2013; Sow *et al*., 2014). Agro-morphological markers are usually visually assessed using direct observation of the trait of interest. Over the years many agro-morphological marker traits with standardised methods of recording have been identified for evaluating diversity in rice (IRRI, 2002; Bioversity International *et al*., 2007).

2.6.2 Biochemical markers

Biochemical markers have been widely used in plant diversity studies. There are two groups of biochemical markers namely metabolic-based markers and protein-based markers. Metabolic-based markers are related to products of secondary metabolism of plants, such as polyphenols, flavonoids, carbohydrates and oils (De Vicente *et al*., 2004). Karthigeyan *et al*. (2008) successfully used biochemical markers of catechins in combination with morphological and amplified fragment length polymorphism (AFLP) markers to evaluate genetic diversity of tea (*Camellia senensis*). Protein-based markers are commonly used in rice diversity studies. Zhu *et al*. (2014) demonstrated the effectiveness of a protein-based marker to differentiate between genotypes of Indica and Japonica groups of *O. sativa*. Glazmann (1987) successfully used isozymes which are protein-based biochemical markers to evaluate genetic diversity of Asian cultivated rice and subsequently identified six variety groups.

2.6.3 Molecular markers

The discovery and use of molecular markers in diversity evaluation studies is a significant breakthrough in the field of molecular genetics. Molecular markers are based on the principle that the genetic profile of living organisms, including plants, is encoded in their DNA. Since the discovery of DNA, there has been a proliferation of molecular marker techniques. Some of the most widely used by order of year of discovery include the following: restriction fragment length polymorphism (RFLP; Grodzicker *et al*., 1975 then popularised by Botstein *et al*., 1980), random amplified polymorphic DNA (RAPD; Welsh and McClelland 1990; Williams *et al*., 1990), SSR (Akkaya *et al*., 1992), SNP (Jordan and Humphries, 1994), AFLP (Vos *et al*., 1995) and diversity arrays technology (DArT; Jaccoud *et al*., 2001).

As indicated by Semagn *et al*. (2006) molecular markers can be classified into three groups based on:

- Mode of transmission (biparental nuclear inheritance, maternal nuclear inheritance, maternal organelle inheritance, or paternal organelle inheritance)
- Mode of gene action (dominant or codominant)
- Method of analysis (hybridisation-based or polymerase chain reaction (PCR) based markers)

Hybridisation-based markers

RFLP is the most popular hybridisation-based marker technique. The technique uses probes that are hybridised to filters containing DNA which has been digested with restriction enzymes. The enzymes consistently cut DNA at specific base pair sequences which are called recognition sites. Restriction enzymes recognise four, six or eight base pair (bp) sequences in DNA and cleave double-stranded DNA whenever these sequences are encountered. This results in a series of fragments of varying lengths. Fragments are separated using gel electrophoresis. Separated fragments are denatured to single strands transferred from the gel to a membrane or filter via the Southern blot procedure (Southern, 1975). Detection of individual fragments is done by nucleic acid hybridisation with labelled probes and finally visualised by autoradiography (Semagn *et al*., 2006). Advantages of RFLP markers include high reproducibility, codominant inheritance and repeatability. Several studies have used RFLP analysis to evaluate genetic diversity and differentiation in groups of Indica and Japonica of Asian cultivated rice (Lu *et al*., 2002; Zhang *et al*., 2003; Jeong and Kim, 2005; Wang *et al*., 2007). Results demonstrated that the variety group Indica was genetically more diverse than Japonica. Analysis using RFLP is still being used in studies related to rice (García de Salamone *et al*., 2012; Fang *et al*., 2013; Cano-Calle *et al*., 2015). However, RFLP markers have several limitations which include the following:

- Requirement for the presence of high quality and quantity of DNA
- The technique is time consuming, laborious and not amendable to automation
- May involve the use of radioactively labeled probes, which is now being replaced by fluorescently labelled probes

PCR-based markers

PCR and advances in genome sequencing have resulted in PCR-based marker techniques currently being the most widely used by plant scientists. It involves using an enzyme to amplify (replicate) exponentially, small initial quantities of DNA. The versatility and high polymorphism of PCR-based markers have made them popular in molecular studies. The development of several thousand SSRs for rice (Temnykh *et al*., 2000; 2001) has made SSR a widely used genotyping tool for rice. SSR markers were used for major studies of the two cultivated rice species (Garris *et al*., 2005; Semon *et al*., 2005; Agnoun *et al*., 2012; Kumar *et al*., 2012; Choudhury *et al*., 2013).

There are several other PCR-based techniques. RAPD was the first PCR-based technique to amplify DNA without prior sequence information. Sequences for RAPD primers are arbitrarily chosen, however, for optimum results there should be a minimum of 40% GC content and absence of palindromic sequences (Williams *et al*., 1990). The biggest limitation of RAPD is the lack of repeatability. It is also a dominant marker technique which is incapable of distinguishing homozygotes from heterozygotes. AFLP is another PCR-based marker that combines the power of RFLP and flexibility of PCR-based technology. Advantages of AFLP include reliability and reproducibility. Several studies have used the technique to evaluate genetic diversity in rice (Maheswaran *et al*., 1997; Zhu *et al*., 1998; Chakanda *et al*., 2012). Principal limitations to AFLP are that it is labourious and may involve handling radioactive reagents. Like RAPD most AFLP loci are dominant and incapable of identifying homozygotes from heterozygotes.

Another commonly used PCR-based technique is SNP. This marker type is based on the ability to detect single base pair changes in DNA between species or paired chromosomes. SNPs are abundant and widely distributed in the *O. sativa* genome. In a study of the rice genome, Liu and Zhang (2006) identified 80127 SNP sites and found one SNP for every 154 bp between variety groups Indica and Japonica. They further reported a SNP rate of 0.65% for closely related cultivars for which it has been difficult to find polymorphic sites by other conventional methods. Pioneering studies revealed that SNPs and indels are abundant and widely distributed in the plant genome (Drenkard *et al*., 2000; Batley *et al*., 2003) including rice (Nasu *et al*., 2002; Yu *et al*., 2002).

Latest generation molecular marker techniques are microarrays. DArT is a microarray hybridisation-based high throughput technique, that allows simultaneous genotyping of hundreds of polymorphic loci spread over the genome (Jaccoud *et al*., 2001; Wenzl *et al*., 2004). The technique has great potential for genome wide association studies in rice (Phung *et al*., 2014). Several methods have been developed based on microarrays. It has been used as a ribonucleic acid (RNA) expression assay technique that permits quantitative analysis of RNAs transcribed from both known and unknown genes, thus providing diagnostic fingerprints and inference and prediction about complex cell control systems (Xiang and Chen, 2000). Microarray genotyping was used by Edwards *et al*. (2008) to reveal high levels of polymorphisms in diverse rice populations. Microarrays have huge potential in predicting gene expression under conditions of stress (Ma *et al*., 2012; Liu *et al*., 2013; Chen *et al*., 2014).

2.7 Rice as a crop plant

Worldwide, rice is a major food staple and one of the world's most important cereal crops. It serves as a major source of nutrition, supplying daily calorie intake for more than half the global population. It is the most important food source for people in the developing countries (Vaughan *et al*., 2003; Dawe *et al.,* 2010; GRiSP, 2010). Global rice production in 2013 was 741 million tonnes with Asia accounting for more than 89% of global area committed to rice and over 90% of total production. Africa accounted for about 7% of global rice area and 4% of global production in 2013 (FAOSTAT, 2015). Although Asia dominates rice production and consumption, the crop is also important in Africa and in other regions of the world. It is the most rapidly growing food source for most people in urban areas in sub-Saharan Africa.

Rice consumption across Africa has shown a sustained increase over the years (GRiSP, 2010). In the "Mano River Union" countries, notably Liberia, Guinea and Sierra Leone, rice accounts for more than 30% of daily calorie intake. In Ivory Coast rice calorie intake increased from 12% in 1961 to 22% in 2004. In Senegal it increased from 20–31%; and for Nigeria rice calorie increased from 1–8% within the same period (Dawe *et al.,* 2010). In the Gambia rice (milled equivalent) per capita consumption per annum is 57.6 kg (FAOSTAT, 2015).

In sub-Saharan Africa, West Africa is the leading regional producer, accounting for 51% of total continental production in 2013 (FAOSTAT, 2015). Increased production is due to expansion in cultivated area, the adoption of new and improved varieties and the application of complementary technologies needed for improved productivity. Between 2006 and 2015 rice production in sub-Saharan Africa increased significantly, however, over the same period consumption has outpaced production, making Africa a net importer of rice (Figure 2.3).

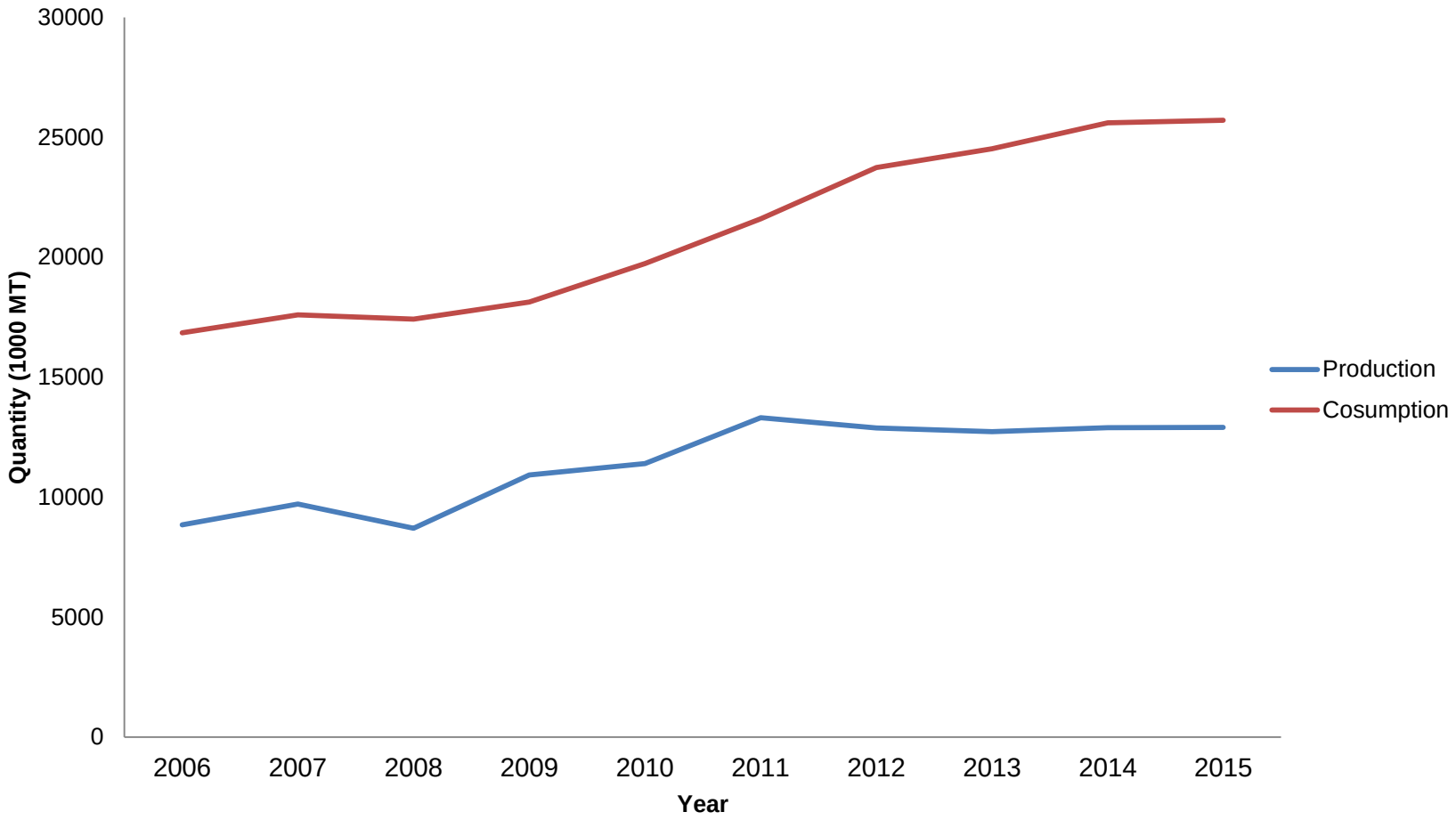
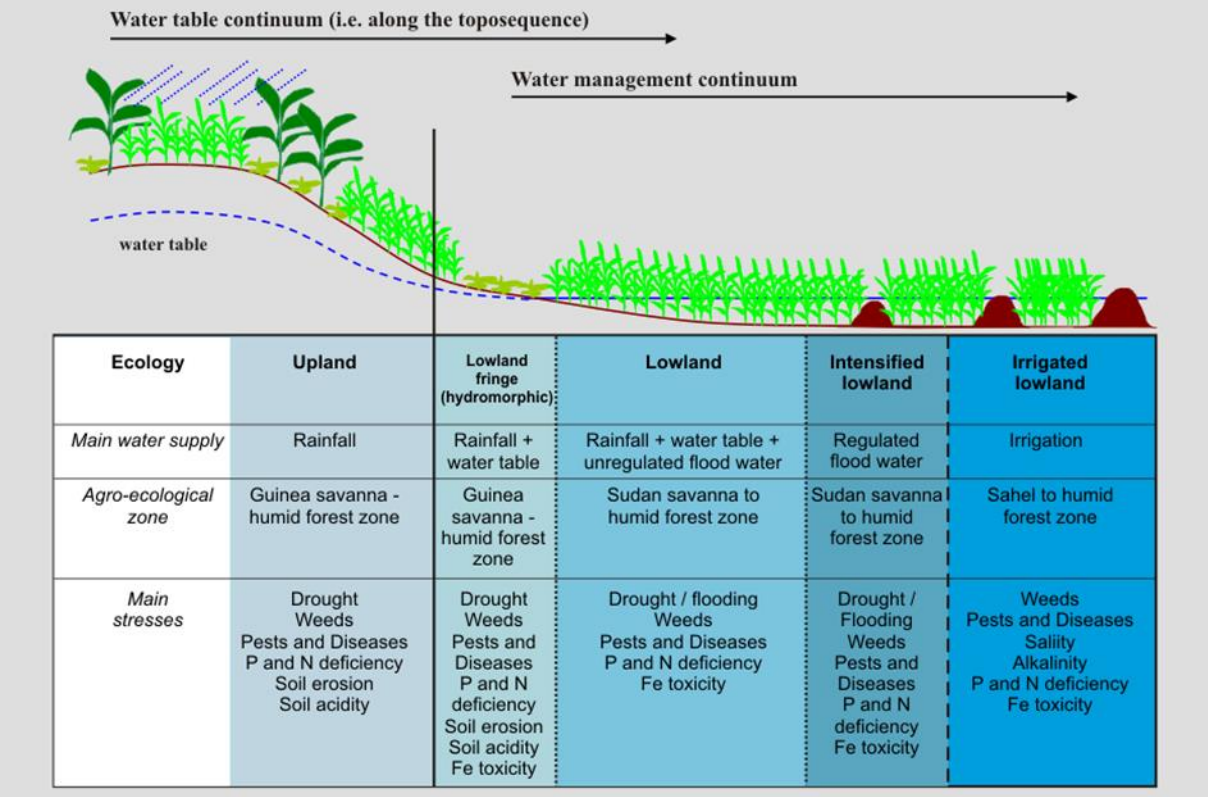


Figure 2.3 Rice production and consumption trends in Africa for 2006-2015 (FAOSTAT, 2015)

2.8 Rice production systems

Rice is a semi-aquatic plant which is adapted to a wide range of ecosystems under varying conditions of moisture and temperature. Figure 2.4 shows the major rice ecosystems in Africa. The irrigated lowland relies on irrigation facilities and Infrastructure for water supply. The intensified lowland, the lowland and the lowland fringe, depend on rainfall and the water table for water supply. The upland relies solely on rainfall for water supply (Figure 2.4). Generally, four major agro-ecosystems are recognised namely: irrigated lowland, rainfed lowland, rainfed upland and flood prone (Khush, 1997; Saito *et al*., 2013). In Africa, flood prone rice is subdivided into deep water and mangrove swamp areas (Saito *et al*., 2013). The relative importance of each of these systems varies according to the production region. The irrigated ecology accounts for 26% of rice area in Africa. The upland and lowland ecologies account for 32% and 38% respectively (Diagne *et al*., 2013). The flood prone area is of low importance in terms of surface area (Balasubramanian *et al*., 2007; Seck *et al*., 2012).



Ecology	Upland	Lowland fringe (hydromorphic)	Lowland	Intensified lowland	Irrigated lowland
Main water supply	Rainfall	Rainfall + water table	Rainfall + water table + unregulated flood water	Regulated flood water	Irrigation
Agro-ecological zone	Guinea savanna - humid forest zone	Guinea savanna - humid forest zone	Sudan savanna to humid forest zone	Sudan savanna to humid forest zone	Sahel to humid forest zone
Main stresses	Drought Weeds Pests and Diseases P and N deficiency Soil erosion Soil acidity	Drought Weeds Pests and Diseases P and N deficiency Soil erosion Soil acidity Fe toxicity	Drought / flooding Weeds Pests and Diseases P and N deficiency Fe toxicity	Drought / Flooding Weeds Pests and Diseases P and N deficiency Fe toxicity	Weeds Pests and Diseases Salility Alkalinity P and N deficiency Fe toxicity

Figure 2.4 Major rice ecosystems in Africa (Adapted from WARDA, 2004)

i. Irrigated lowland: Rice is grown in lowland areas with controlled and assured supply of water. The crop is grown in bunded fields with irrigation canals for water control. Irrigated systems are the most productive with the highest yields and are usually associated with the use of improved varieties, high input use and relatively improved management.

ii. Rainfed lowland: Rice is grown on lowlands where water supply is not always assured and there is no effective water control. The system depends on rainfall and the water table for water supply. Rice is grown on level to slightly sloping fields in lower parts of the topo sequence and in inland valleys (Saito *et al*., 2013). The type of varieties grown range from modern improved varieties like those in irrigated ecologies to traditional varieties. Choice of variety depends on the type of rainfed lowland.

iii. Upland ecology: Rice is grown on free-draining soils where the water table is below the rooting depth. Generally there is no assured supply of water and all moisture comes from rainfall. Upland systems in Africa are the most extensive but the least productive. They are usually associated with traditional varieties. Recently there has been a widespread adoption of improved varieties (like the NERICAs) by farmers to replace traditional ones.

iv. Flood-prone ecology: Rice production is done in (a) deep-water or (b) mangrove swamp areas. For deep-water rice production systems the crop is grown in low-lying areas and flood plains close to rivers, inland valleys and coastal wetlands (Kawano and Sakagami, 2008). In West Africa rice that is grown in coastal tidal swamps and estuaries where the dominant vegetation is mangrove, is called mangrove swamp rice (WARDA, 1994). Tall varieties that are photoperiod sensitive, tolerant to salinity and with an ability to elongate are most suitable and are widely used for this ecology.

2.9 Rice as a self-pollinating plant

Rice is basically a self-pollinated plant. This phenomenon referred to as autogamy, is aided by the non-exposure of the stigma during anthesis. The stigma remains within the palea and the lemma. The shedding of pollen often coincides with the opening of the flower and most pollen will fertilise a stigma in the same flower. The flower opens only once, which lasts for about one hour. The viability of pollen is influenced by humidity, temperature and exposure to UV radiation. The stigma, which is viable for a longer period, remains receptive for several days (Gealy *et al*., 2003).

Rice pollen is fragile and short lived, viable for about 10–20 min after shedding. Exposure to above or below critical temperatures has been shown to kill pollen. Rice floret sterility was observed when plants were exposed to night temperatures below 16ºC at 10–16 days before heading.

Jagadish *et al.* (2007) reported that exposure of rice pollen to temperatures ≥ 33.7ºC at anthesis for about one hour caused sterility. In general, species and cultivars differ in pollen tolerance towards heat (Satake and Yoshida, 1978; Matsui *et al*., 2000). When spikelets reach a temperature of 31ºC, sterility increases and at 36ºC there is complete sterility (Weerakoon *et al*., 2008). Self-pollination is an advantage, as it reduces undesirable outcrossing and by implication eases the production of quality rice seed. However, some level of natural outcrossing does occur in rice. This is where the stigma receives pollen from a different flower on the same plant or from a different plant. When outcrossing occurs with an undesirable donor, contamination occurs and off-types are produced. This is a critical issue in the production of quality rice seed as it reduces the genetic purity of the seed crop. Progeny from such seed stock will segregate and adversely affect the accepted uniformity and the genetic purity of the crop. A seed crop of below the accepted standards would fetch less money and thereby cause economic loss (Lee *et al*., 2013).

2.10 Outcrossing in the rice plant

Outcrossing may be defined as the successful fertilisation and hybridisation between two plants. For outcrossing to occur there must be cross pollination, and for cross pollination to occur there must be a pollen vector and plants must be close enough to each other. Flowering periods of donor and recipient plants must also overlap and pollen of the donor must be viable and come into contact with a receptive female/ovule. Rate of outcrossing is generally determined by the proximity in time and space of the pollen donor and the receptive female. Other factors that influence rate of outcrossing in rice are prevailing environmental conditions at time of anthesis. Humidity, temperature, wind speed and wind direction have been found to be factors that influence rates of outcrossing in rice. Extreme temperatures that kill pollen, but not the stigma, may facilitate outcrossing of the floret by viable foreign pollen (Weerakoon *et al.*, 2008). Hokanson (2006) showed that duration of pollen viability is influenced by prevailing relative humidity and exposure to UV radiation. With the current global climate change, the world is experiencing increasing temperatures and therefore the outcrossing rate in rice is likely to increase (Jagadish *et al*., 2010; Ziska *et al*., 2012).

Several studies have indicated occurrence of outcrossing in rice (Langevin, 1990; Ellstrand *et al.,* 1999; Chen *et al*., 2004; Da Silva *et al*., 2005; Endo *et al*., 2009; Nuijten *et al*., 2009; Somaratne *et al*., 2012). Investigators have used several methods and different markers in outcrossing studies. Da Silva *et al.* (2005) evaluated outcrossing rates in two cultivars of upland rice. Using the morphological marker of leaf pubescence, they recorded variable rates of outcrossing that

were less than 1% at 0.5 m from the donor. However, with only two genotypes (one donor and one recipient), the number was too low to be representative. It was not sufficient to reflect outcrossing rates that are known to be influenced by type of genotype. They reported on rainfall, temperature and humidity at flowering time, but there was nothing on prevailing wind speed. This is a drawback, as wind speed also influences outcrossing rates. In another study Endo *et al*. (2009) used morphological markers of purple grain colour of the donor and the leaf colour of F_1 plants. The recipient plant had white kernel colour and was glutinous. Outcrossing rates were generally low with values of less than 0.05% within 30 m from the donor. However, wind speed during the flowering period influenced outcrossing rates. This study also used an insufficient number of genotypes as only one donor and one recipient was used. Plot sizes were also relatively small. Whilst outcrossing rates were small, pollen dispersal to over 20 m from the donor was observed.

With the advent of molecular markers, several studies have used them to evaluate outcrossing rates in rice. Somaratne *et al*. (2012) initially used morphological markers and then molecular markers for confirmation to measure outcrossing in rice. Short round grain and purple colour of the stem were used as morphological markers. Then the STS marker pTA248 was used as the molecular marker. In their study involving one donor and one recipient, a potential outcrossing rate of 3.41% with an average of 1.29% was recorded. Similarly, the study had the drawback of too few genotypes. Rong *et al*. (2007) used molecular markers to study outcrossing rates of three insect resistant GM rice (*Bt/CpTI*) and non-GM isogenic lines. The marker (*hpt*) which is linked to the hygromycin-resistance transgene was used. Low frequencies of outcrossing were recorded and there was a dramatic reduction with distance. Outcrossing ranged from 0.28% at 0.2 m to less than 0.01% at 6.2 m. Different plot sizes (400 m^2, 380 m^2, 200 m^2 and 100 m^2) were used for the donor whilst 640 m^2 and 1216 m^2 were used for the recipient cultivars. Difference in plot size of the donor was found not to influence outcrossing rate. However, this could be due to the size of the recipient plots being larger than the donor plots. Therefore there was always much more pollen from the recipient to annul the effect of the pollen load from the donor. Results could have been different if donor and recipient plot sizes were similar or if the donor plot size was larger. The study did not include environmental conditions which could have added credibility to the findings.

2.11 Genetically modified rice: potential and associated risks

Globally there has been a sustained increase in the use of GM crops over recent decades. This is in response to increased demand for food consumption and industrial uses. The reaction of the public in accepting GM crops has been mixed. The risks and benefits of GM crops continue to be debated. The global area of biotech crops increased from 1.7 million ha in 1996 to 181.5 million ha in 2014. Africa is also witnessing increasing hectarage for GM crops where 2.9 million ha were planted in 2011 (ISAAA, 2014). Conventional breeding approaches contributed tremendously to the success of the Green Revolution in Asia and lifted millions out of hunger and poverty, however, it is unlikely that conventional breeding alone can meet the challenge of increased food demand of a rapidly growing global population in the 21st century.

Crop research is presently looking at novel techniques like biotechnology and GM rice to attain yields that are high and stable in order to satisfy global demand. Research on GM rice, which is broad and diverse, is presented in Table 2.3. It is focused on increased yields, resistance to pests and diseases, high protein and vitamin content and the pharmaceutical industry. Other areas of research are on nutrient-use efficiency and tolerance to abiotic stresses like drought and salinity (Lu and Snow, 2005; Chen *et al.,* 2009). GM rice can contribute to increased food security due to increased yields that will be realised from alleviating specific constraints. Insect resistant GM rice could also lead to reduced use of pesticides, therefore limiting the release of hazardous chemicals into the environment (Brookes and Barfoot, 2012). But there is the potential for gene flow between GM rice and non-GM rice (conventional varieties, other wild relatives and weedy rice). This has raised concern over the impact of GM rice on the environment. There are cases of negative impact associated with GM crops (Ellstrand *et al*., 1999).

There are also reports of gene flow between cultivated herbicide resistant rice and weedy rice and wild relatives of rice (Messeguer *et al*., 2001; 2004). Messeguer *et al*. (2004) assessed the frequency of pollen-mediated gene flow from a transgenic rice line that harbours genes which encode for β-glucuronidase and phosphinothricin acetyl transferase to red rice weed and a conventional rice cultivar. Pollen recipient and donor plants were planted in concentric circles to investigate the effect of the wind on the frequency of pollen flow. Based on the detection of herbicide resistant seedlings from progeny tests of recipient plants they reported frequencies of gene flow averaged over all wind directions of 0.036% and 0.086% for red rice and conventional rice, respectively.

Table 2.3 Transgenic traits and other information on GM rice cultivars (Lu and Snow, 2005)

Transgenic trait	Additional details	Gene donor	Reference
Herbicide resistance			
Glufosinate resistance	Commercial names: LibertyLink, Basta	Bacterium	ISB (2004)
Glyphosate resistance	Commercial name: Roundup Ready	Bacterium, CBI	ISB (2004)
Imidazolinone resistance	Nontransgenic variety also exists	Maize	ISB (2004)
Sulfonylurea resistance		Human	Inui *et al*. (2001)
CBI		CBI	ISB (2004)
Pest resistance			
Fungal disease resistance	*Rhizoctonia solani* (sheath blight)	Rice, *Arabidopsis thaliana*, bean, tobacco, barley	ISB (2004)
	Pyricularia oryzae (blast)	Unknown	ISB (2004)
Bacterial disease resistance	*Xanthomonas oryzae* (bacterial blight)	Rice, alfalfa	ISB (2004)
Viral disease resistance	Several types	Virus coat proteins	OECD (1999)
Insect resistance	Lepidopteran	Bacterium, cowpea	Xu *et al*. (1996), ISB (2004)
	Coleopteran	Bacterium, cowpea	Ghoshal *et al*. (2001), ISB (2004)
	Homopteran (brown planthopper)	Snowdrop lectin	Wu *et al*. (2002)
Stress tolerance			
Salinity tolerance		CBI	ISB (2004)
Drought tolerance		CBI	ISB (2004)
Cold tolerance		Rice	Hoshida *et al*. (2000)
Yield increase			
Yield increased		CBI, maize	ISB (2004)
Photosynthesis and yield enhanced by C4 enzymes		CBI, maize	Ku *et al*. (2001), ISB (2004)

Table 2.3 Continued Transgenic traits and other information on GM rice cultivars (Lu and Snow, 2005)

Transgenic trait	Additional details	Gene donor	Reference
Other traits			
Altered storage protein		Pea, rice	ISB (2004)
Pharmaceutical proteins		Human, CBI, *Forsythia intermedia*	ISB (2004)
Altered carbohydrate metabolism		CBI	ISB (2004)
Novel protein		Human, CBI	ISB (2004)
Altered polyamine metabolism		CBI	ISB (2004)
Increased starch level		CBI	ISB (2004)
Altered morphology		CBI	ISB (2004)
Male sterility	For hybrid seed production	CBI	ISB (2004)
Value-added protein		Human	ISB (2004)
Heavy metal bioremediation		Bacterium	ISB (2004)
Altered seed composition		CBI	ISB (2004)
Enhanced betacarotene		Daffodil (*Narcissus pseudonarcissus*)	Burkhardt *et al*. (1997)
Enhanced iron and zinc		Soybean	Vasconcelos *et al*. (2003)
Low allergen		Unknown	OECD (1999)
Low protein	For sake brewing	Unknown	OECD (1999)

*CBI, confidential business information

In another study, Rong *et al*. (2004) used SSR markers to study gene flow between a traditional glutinous rice variety and hybrid rice variety that were grown in close proximity in China. The two varieties were intercropped using different combinations of alternative lines from a ratio of 1:1 up to 1:10. SSR analysis was used for the detection of hybrids. They recorded gene flow rates of 0.04% in the traditional cultivar and 0.18% in the hybrid cultivar. Rong *et al*. (2007) also assessed crop-to-crop gene flow using three insect resistant GM rice and non GM isogenic lines. The GM rice line was used as pollen donor and the hygromycin-resistance transgene was used to screen seeds from non GM rice rows at different distances ranging from 0.2–10 m from donor plants.

Identification of transgenic plants was done by culturing germinated seeds in a liquid medium that contained 50 µg ml^{-1} hygromycin B. surviving seedlings were considered as transgenic plants containing the hygromycin-resistance gene. They examined about 2.1 million seeds and reported gene flow of 0.28% at 0.2 m to < 0.01% at 6.2 m.

Chen *et al*. (2004) assessed gene flow under field conditions between cultivated *O. sativa* and two of its wild relatives; *O. sativa f spontanea* (weedy rice) and *O. rufipogon* (wild rice). Multilocation trials that mimicked natural conditions were carried out in China and Korea. Cultivated rice which served as the pollen donor was intercropped with the wild and weedy taxa which served as pollen recipients. They detected gene flow frequency of between 0.011 and 0.046% from cultivated rice to weedy rice; and about 1.21 and 2.19% from cultivated rice to wild rice. Somaratne *et al*. (2012) estimated the outcrossing rate of rice variety Bg 379-2 using morphological markers of short-round grain and purple culm colour of pollen donor Bg 450. The pollen donor was planted on either side of the recipient. They detected a potential outcrossing rate of 3.41% with an average of 1.29%. Studies using GM rice have also reported outcrossing between rice and wild relatives. Shivrain *et al*. (2009) studied the effect of cultivar type (Clearfield® rice) and red rice biotype on outcrossing rates. They used a split-split-plot design with date of planting as main factor, rice cultivar as subplot factor and red rice biotype as sub-subplot factor. They detected outcrossing rates that ranged from 0.004–0.006%. Most outcrossing was observed within 1 m of donor pollen, but outcrossed plants were observed as far as 6 m from donors.

Outcrossing between cultivated rice and wild rice is frequent and often higher than between cultivated rice and weedy rice (Song *et al.,* 2003; Chen *et al*., 2004). These rates, though seemingly low, should be viewed with caution and within context. Seeds obtained from sources of even low levels of outcrossing could result in high numbers of outcrossed progenies when planted. Outcrossing by a transgene may result in a hybrid with fitness advantage with regards to competitors. Thus it may spread rapidly through weedy, wild or conventional rice varieties. Langevin (1990) observed that hybrids resulting from outcrossing between cultivated rice and red rice were more robust and fitter than either of the parental sources. For instance, in cases where herbicide resistant rice outcrosses with weedy rice, there will be production of many weed-crop hybrid seeds per ha. The scenario becomes alarming when it is a seed crop where even larger areas will be infested in subsequent seasons with undesirable weed plants.

In recent years there has been a rapid expansion in the production of rice in sub-Saharan Africa. This has been largely due to the introduction of appropriate varieties and associated technologies

by research, like the introduction of the upland NERICA varieties by the Africa Rice Center (WARDA, 2000; 2006). The rice crisis in 2008 also led to a further increase in rice cultivation as several governments in sub-Saharan Africa implemented policies that were conducive to the production of rice. Total cultivated area for rice in sub-Saharan Africa increased from 8.5 million ha in 2007 to 10.9 million ha in 2013. Production increased from 21 million tonnes in 2007 to 28.7 million tonnes in 2015 (FAOSTAT, 2015). Rice cultivation has recently been introduced and adopted in hitherto non-traditional rice producing and consuming countries in sub-Saharan Africa. This is the case in Benin in West Africa and Uganda and Ethiopia in East Africa. In Ethiopia the area under rice cultivation increased from 6000 ha in 2005 to 222 000 ha in 2010; and rice production increased from 15 460 tonnes in 2005 to 887 400 tonnes in 2010 (Mohapatra, 2012). Increase in cultivated area has resulted in a corresponding increase in demand for quality rice seed. Demand for rice seed in Africa increased from 715 579 tonnes in 2007 to 824 968 tonnes in 2011 (FAOSTAT, 2015). The situation has been compounded by a corresponding increase in the number of rice farmers. For example in Ethiopia the number of rice farmers increased from 18 000 to more than 565 000 (Mohapatra, 2012).

Research is in progress to exploit new sources of diversity. The Africa Rice Center has developed promising lines resulting from crosses between *O. sativa* and *O. barthii* (AfricaRice, 2011). Interspecific rice hybrids like the NERICAs have been found to be compatible with *O. sativa* (Ikeda *et al.,* 2009). Therefore interspecific varieties could serve as bridge cultivars for the transfer of useful genes between *O. glaberrima* and *O. sativa*. On the other hand, interspecific varieties could also serve as a conduit for the introgression of undesirable traits into rice genotypes. It is not uncommon in sub-Saharan Africa to observe rice fields with high levels of admixtures and impurities. This might have been due to outcrossing from an undesirable donor. Rice fields with different varieties are often grown in close proximity. Red rice and other wild relatives also grow in similar habitats and are usually found in close proximity to cultivated rice. The combination of different varieties, red rice and other wild relatives of rice in close proximity and flowering synchronously, would eventually result in outcrossing. Many farmers in sub-Saharan Africa without an alternative, resort to traditional varieties and landraces and use indigenous knowledge for seed production and variety maintenance. That is to say farmers generally use the informal sector to cater for seed demand. It is believed that less than 10% of seed used in Africa is obtained from the formal sector. Within the informal seed sector, no strict regulations are observed. Growing different varieties at close proximity is not uncommon. Grain is also often used as seed and *vice versa*. Nuijten *et al*. (2009) used AFLP analysis to assess genetic diversity of rice

germplasm collected from farmers' fields in West Africa. They reported progenies of interspecific hybridisation between African and Asian rice occurring in farmers' fields.

Production of quality seed is relatively easy for a self-pollinating plant like rice. Outcrossing in rice, which is effected by wind, is limited. Levels of outcrossing are influenced by weather conditions and by spatial and temporal proximity between donor and recipient. Species and cultivar differences have also been shown to influence levels of outcrossing. It is higher between species than within species. Higher outcrossing rates have been observed between *O. sativa* and *O. rufipogon* than within *O. sativa*. It is also higher in *indica* cultivars and wild species than in *japonica* cultivars of *O. sativa* (Messeguer *et al.,* 2001). The characteristics of the pollen donor also influence the level of outcrossing. The production of many florets per unit area and the presence of large extruded anthers that release large numbers of pollen per floret (high pollen load) facilitate outcrossing (Matsui *et al*., 2000).

It is often necessary to prevent or limit undesirable outcrossing. Isolation by time or by space is the recommended method to counter outcrossing from undesirable donors. Temporal isolation is defined here as sowing a cultivar at a given point in time that its flowering period does not coincide with the time of flowering of any undesirable pollen donor. The cultivar begins and completes flowering either before or after any undesirable donor. Isolation by time is effective against outcrossing in rice (Gealy *et al.,* 2003; Endo *et al.,* 2009). However, use of isolation by time is limited as it is constrained by the length of the growing season, which is definite. In addition, other related essential conditions like adequate radiation, day length and temperature are only available at specific times of the year and must be exploited within that period.

Spatial isolation is defined here as sowing a cultivar at a distance far enough to prevent pollen flow and outcrossing from an undesirable donor. This is the most widely used method to prevent undesired outcrossing. Da Silva *et al.* (2005) recommended an isolation distance of 2 m between neighbouring rice cultivars. Some studies have indicated that outcrossing due to pollen flow does not occur beyond 10 m (Messeguer *et al*., 2001; Gealy *et al*., 2003; Zhang *et al*., 2003). However, Endo *et al*. (2009) detected donor pollen at a distance of 20 m from the source and Kanya *et al*. (2009) recommended an isolation distance of 250 m from the donor. Devos *et al*. (2006) recommended a physical pollen barrier to restrict gene flow due to pollen dispersal. Seed certifying agencies recommend specific isolation distances, considered adequate for different classes of seed. The distance differs according to the objective. In general, the isolation distance for hybrid seed is much higher than for non-hybrid seed.

It is important to quantify gene flow between rice varieties in order to assure the production of good quality seed. Knowledge gained is equally important for plant breeding programmes and for genetic maintenance and conservation. There has been widespread adoption of NERICA varieties and other interspecific and high yielding varieties across Africa. This raises the issue of effective and efficient maintenance of varieties to avoid undesirable gene flow in order to guarantee genetic purity and their genetic potential. Gene banks as well as international and national agricultural research institutions have as a major objective, the maintenance and preservation of the genetic purity of accessions in their custody. Seed multiplication and certification programmes also have maintenance and preservation of genetic purity as a major objective. This objective underpins all procedures in the management of seed lots during conservation and regeneration. In addition, the occurrence of undesirable outcrossing which could lead to gene flow during the evaluation of breeding lines in a breeding programme can delay and/or complicate the fixing of such materials for release as cultivars. Segregating material cannot meet the condition for the distinctiveness, uniformity and stability (DUS) test that is required for variety release. Material that is not fixed before release will segregate in subsequent generations in farmers' fields. The quantification of gene flow could also serve as part of routine biosafety risk assessment with regards to transgene escape and its impact on the environment (Messeguer *et al*., 2004; Rong *et al*., 2004; 2007; Da Silva *et al*., 2005; Chandler and Dumwell, 2008; Shivrain *et al*., 2009).

Many studies have been done to evaluate outcrossing rates in rice. Different kinds of markers, ranging from morphological traits to herbicide tolerance and other molecular traits, have been used in studies to detect hybridisation (Song *et al*., 2003; Messeguer *et al*., 2004; Da Silva *et al*., 2005; Lu and Snow, 2005; Rong *et al*., 2007; Endo *et al*., 2009; Shivrain *et al*., 2009; Phan *et al*., 2012; Somaratne *et al*., 2012). However, studies were generally small-scale with plots not large enough to produce a pollen load that can reliably affect levels of outcrossing. Furthermore, studies have been on cultivars of *O. sativa* and they did not include *O. glaberrima* and other local land races of African rice. In addition, there is no information on outcrossing rates of interspecific hybrids (to which the NERICA varieties belong). Presently across Africa, large numbers of farmers have adopted interspecific hybrids and at the same time, the Africa Rice Center is focusing efforts for the release of more interspecific hybrid varieties in the near future. Therefore, there exists a need for a deeper understanding of the nature of outcrossing in these varieties (Nuijten *et al*., 2009; AfricaRice 2010; 2011).

Natural outcrossing of rice is dependent on environment and genotype. Information on pollen flow and outcrossing in rice that is available and being used in Africa was obtained in distant places and cannot be extrapolated to African conditions. Environments and genotypes in those studies are different from those in Africa. It is important to understand the nature and levels of rice to rice outcrossing under field and environmental conditions that prevail in Africa. Such an understanding will contribute to formulating strategies to avoid undesirable outcrossing and gene exchange.

2.12 Methods of estimating outcrossing

Outcrossing can be referred to here as pollination of a plant by foreign pollen from a different variety or species. It invariably results in the introduction of genetic material from one genotype to another. Therefore, high levels of outcrossing could lead to increased gene flow between genotypes. Techniques used to measure gene flow fall into two categories namely, direct and indirect methods. Direct methods utilise pollen, gene or seed dispersal measurements to estimate levels of gene flow. They have been widely used in rice gene flow studies (Song *et al*., 2003; Messeguer *et al*., 2004; Rong *et al*., 2004; Da Silva *et al*., 2005; Lu and Snow, 2005; Endo *et al*., 2009; Shivrain *et al*., 2009; Phan *et al*., 2012; Somaratne *et al*., 2012). Indirect methods use a specific parameter to characterise the distribution of genotypes and then apply population genetics theory to estimate gene flow (Neigel, 1997; Ellstrand, 2003).

2.12.1 Direct methods of estimating gene flow

2.12.1.1 Pollen dispersal

Pollen dispersal can be used to estimate the distance over which gene flow can occur in one generation. A genotype of interest is planted and then devices are put in place to monitor pollen dispersal distances. For entomophilous species, pollinator foraging behaviour is also monitored. Inferences on gene flow are made based on these observations (Chapman *et al*., 2003; Cresswell and Osborne, 2004). For anemophilous plants, pollen traps are placed at specific distances to capture dispersed pollen at predetermined heights. Rice is mainly wind-pollinated, but exhibits a limited amount of entomophily (Gealy *et al*., 2003).

Kanya *et al*. (2009) in a study carried out in Kenya, assessed pollen dispersal distance in rice *O. sativa* cultivar Basmati. Pollen traps were placed at different distances up to 300 m from the source cultivar. Samples were drawn at different heights for each distance. Rice pollen was found up to 250 m from the source. Maximum pollen count was observed between 11:00-12:00 hours local time at a temperature of 28ºC. They recommended an isolation distance of at least 250 m

to minimise contamination between different rice genotypes. However, the recommendation would have been more appropriate if more genotypes had been included in the study, then differences due to genotype could have been captured.

In a similar study carried out in China, Song *et al*. (2004) monitored the pollen flow pattern of rice using a universal paternal line for hybrid seed production. They observed that pollen density decreased with distance from pollen source and that density was positively correlated with size of the pollen source. They indicated that pollen flow patterns were influenced by climatic conditions, especially wind speed and wind direction. This is in accordance with other studies that monitored pollen flow in rice (Messeguer *et al*., 2004; Hoyle and Cresswell, 2007; Endo *et al*., 2009).

2.12.1.2 Gene dispersal measures

This technique involves the use of a genotype with a unique trait as a pollen donor to another genotype that lacks the trait. This implies a unidirectional monitoring of gene flow via pollen from donor to recipient. The recipient plants are placed at various distances from the donor. A marker associated with the unique trait of the donor is used to detect and confirm hybridisation. The marker could be morphological, molecular or a combination of both. Progeny from recipients are tested for the presence of the pollen donor marker. The proportion of plants with the marker gives the hybridisation rate and hence the extent of gene flow between the two genotypes.

In a study utilising leaf pilosity as a morphological marker, Da Silva *et al*. (2005) recorded outcrossing rates of 0.30% and 0.35% in rice in two sites at 0.5 m from the pollen donor. Similarly Rong *et al*. (2006) detected outcrossing rates of less than 0.30% at 0.2 m from the donor which reduced with increased distance from the donor. At 6.2 m from the donor the rate of outcrossing was less than 0.01%. Messeguer *et al*. (2001) in their studies of gene flow in rice using transgenic rice, detected outcrossing rates of less than 1% at 1 m from the donor. Endo *et al*. (2009) used a non-glutinous cultivar with purple grain and ligule colour as a pollen donor in their study that investigated the relation between outcrossing rates and climatic conditions. Two phenotypic markers: the xenia grain and the purple colour of leaves and ligules of F_1 plants were used to detect hybridisation. The study combined pollen and gene dispersal approaches to infer potential gene flow between two rice cultivars. They observed hybrid plants up to 25 m from the donor and recorded donor pollen at a distance of 30 m from the donor. Pollen flow patterns and outcrossing rates were found to be influenced by wind speed and direction, the length of the flowering period and the overlap in flowering.

In a study estimating outcrossing rate in rice, Somaratne *et al*. (2012) combined morphological and molecular markers to detect hybrid progeny. The traits of short-round grains and purple colour of culm of the pollen donor were used as morphological markers. Outcrossing rates were confirmed by STS molecular marker pTA248. In the study a potential outcrossing rate of 3.41% at a distance of 0.8 m from the pollen donor was detected. Song *et al*. (2003), using SSR markers recorded high outcrossing rates between cultivated rice *O. sativa* and the wild species *O. rufipogon*. Hybrid plants were detected as far as 43.2 m from the pollen donor.

2.12.1.3 Estimates of gene flow through paternity analysis

Paternity analysis is based on the principle that progenies derive their DNA in equal proportions from each of the parents. Therefore, an examination of progeny or seeds of known maternal parent using appropriate genetic markers can reveal the paternal parent. The method involves use of multilocus fingerprints to infer male parents of seed offspring from known female parents. Progeny with multilocus fingerprints different from the parents within the population are assumed to have resulted from pollen outside of the population and the two parents.

Jiang *et al*. (2012) used SSR markers in paternity analysis of 24 weedy rice (*O. sativa f. spontanea*) populations and their coexisting rice cultivars from northern Italy to study genetic differentiation, outcrossing rates and introgression based on microsatellite polymorphisms. They recorded genetic differentiation (F_{ST}) of 0.26 in weedy populations and identified 28% of private alleles in most populations that exhibited multiple cluster assignments. Outcrossing rates were positively correlated with private alleles of corresponding popuations and 15% of paternal specific alleles were acquired from introgression of coexisting rice cultivars. They concluded that private alleles of feral populations resulted from outcrossing with nearby cultivars of a crop and that allele introgression from a crop may play an important role in the adaptive evolution of feral crops.

2.12.2 Indirect methods of estimating gene flow

2.12.2.1 Fixation index based estimates

Assuming neutral genes, population structure in equilibrium, a low rate of gene flow between subpopulations and uniform and constant population size and gene flow, then gene flow in relation to the average number of effective immigrants per generation ($N_e m$) in a population can be used to infer population structure. Species exhibit clinal variation in both morphology and gene frequency. The amount of variation is influenced by mitigating forces of genetic differentiation and homogeneity.

Gene flow can be inferred from the fixation index of population subdivisions relative to the total population (F_{ST}). The approach is based on Wright's island model of population genetic structure (Wright, 1931). The model assumes a population at equilibrium with regards the homogenising effects of gene flow and the disruptive effects of genetic drift. It also assumes that population size and migration rate are uniform and constant over space and time. F_{ST} is the correlation between random gametes within subpopulations relative to gametes in total population and is a function of the average effective number of migrants per generation ($N_e m$). For subpopulations at migration-drift equilibrium:

$$F_{ST} = \frac{1}{4N_e m+1} \qquad (1)$$

Where $N_e = \text{effective subpopulation size;} \quad m = \text{proportion of migrants per generation}$

Thus, the effective number of migrants can be estimated from equation (1) as:

$$N_e m = \frac{1}{4}\left(\frac{1}{F_{ST}} - 1\right) \qquad (2)$$

This has been extended to develop other methods for measuring population differentiation (Nei, 1973; Weir and Cockerham 1984; Slatkin, 1985). Slatkin (1985) proposed the parameter R_{ST}, which uses frequency of rare alleles to derive population genetic differentiation. Weir and Cockerham (1984) proposed the parameter θ and Nei (1973) proposed the parameter G_{ST} to measure population genetic differentiation.

Zhao *et al*. (2013) used this approach to investigate genetic variation in 11 Chinese *O. rufipogon* populations. They used 79 SSR loci to infer the effects of habitat fragmentation on genetic structure and concluded that isolation by distance due to historical rather than recent fragmentation, followed by local adaptation, has driven population divergence in *O. rufipogon*. In another study Das *et al*. (2013) also used fixation index to examine genetic distances and the population structure among 91 Indian rice accessions using 23 SSR markers and model-based clustering approach respectively. Similarly, Chung and Park (2009) used fixation index to investigate genetic diversity, linkage disequilibrium and population structure of weedy and cultivated rice accessions using 63 SSR markers. He *et al*. (2014) used the approach to study genetic diversity and structure of 21 weedy rice populations from Sri Lanka. They reported high levels of within-population genetic diversity and low levels among-population differentiation.

Because of the relative ease of its application, F_{ST} was widely used in estimating gene flow within sub-populations. However, the index is criticised due to unrealistic assumptions associated with it (Whitlock and McCauley, 1999). F_{ST} is a summary statistic and does not reflect the effects of landscape-level processes that affect spatial patterns of genetic structure and rates of gene flow. In addition, there is the assumption that all populations are equal sources of migrants and estimates are inferred on a historical perspective. Therefore, contemporary variation in gene exchange among populations or changes to dispersal processes is not reflected by the model. (Sork *et al.,* 1999).

Other F_{ST} based methods for estimating gene flow are the stepping stone model (Kimura and Weiss, 1964) and the isolation by distance model (Wright, 1969). In these models, gene flow is expected to decline monotically with increasing distance between spatially discrete and continuous populations respectively. Both models depict decrease in gene flow with increasing geographic distance between populations. The advantage of these models over Wright's island model is that they incorporate a spatial dimension in estimating gene flow. In recent years the use of molecular markers to complement earlier models has increased the reliability of population genetics studies (Garris *et al.,* 2005; Karasawa *et al*., 2007).

2.13 Conclusions

Rice is a tropical cereal that belongs to the family Gramineae. It comprises of two cultivated species and some 28 wild relatives. The cultivated species namely *O. sativa* and *O*. *glaberrima* were probably domesticated in the Yangtze River Valley in China some 10000–14000 years ago, and in the basin of the Upper Niger River in West Africa more than 3500 years ago. *Oryza sativa* is grown worldwide and has more genetic diversity with several variety groups and is believed to have been domesticated from the wild relative *O. rufipogon*. The less diversified *O. glaberrima* is grown only in Africa and was probably domesticated from wild relative *O. barthii*.

Rice is a staple food providing more than 20% of daily calorie intake for 3.5 billion people in many parts of Asia, Africa and other parts of the world. Rice is also becoming an increasingly important food crop in Africa where it is replacing traditional staple foods. World rice consumption is projected to rise from 439 million tonnes in 2010 to 555 million tonnes in 2035. This implies an increase in production is required to meet the demand. The consistent production increase over the years has resulted in a corresponding increase in demand for quality rice seed. Demand for rice seed in Africa increased from 7.0 million tonnes in 2007 to 7.9 tonnes in 2013 (FAOSTAT, 2015). Therefore, the provision of good quality seed to meet production demand is vital. However,

one of the constraints to provide quality seed is undesirable outcrossing in rice that occurs under field conditions. This leads to genetic impurities that are manifested by the appearance of off-types and non-uniform crop stands. Rice is a self-pollinated crop, but it is known to exhibit some level of outcrossing. Progeny from such seed stock will segregate and adversely affect the accepted uniformity and the genetic purity of the crop, thereby causing economic loss.

In order to meet the challenge in food security, rice research has introduced a range of promising technologies including interspecific rice (like the NERICAs) and GM rice varieties. However these varieties have been shown to outcross under field conditions. The assessment of the nature and extent of natural outcrossing is important for plant breeding and seed programmes and for the effective maintenance and conservation of rice germplasm. Information on outcrossing is particularly lacking for African rice germplasm evaluated under prevailing conditions in Africa.

2.14 References

AfricaRice (Africa Rice Center). 2010. New Breeding Directions at AfricaRice: Beyond NERICA. Cotonou, Benin: 24pp.

AfricaRice (Africa Rice Center). 2011. Boosting Africa's Rice Sector: A Research for Development Strategy 2011-2020. Cotonou, Benin: ii + 77 pp.

Agnoun Y, Sié M, Djedatin G, Dramé KN, Toulou B, Ogunbayo SA, Sanni KA, Tia D, Ahanchédé A, Vodouhè RS, Ndjiondjop MN. 2012. Molecular profiling of interspecific lowland rice progenies resulting from crosses between TOG5681 and TOG5674 (*Oryza glaberrima*) and IR64 (*Oryza sativa*). International Journal of Biology 4:19–28.

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139.

Babaei A, Nematzadeh GA, Hashemi H. 2011. An evaluation of genetic differentiation in rice mutants using semi-random markers and morphological characteristics. Australian Journal of Crop Science 5:1715–1722.

Balasubramanian V, Sié M, Hijmans RJ, Otsuka K. 2007. Increasing rice production in sub-Saharan Africa: challenges and opportunities. Advances in Agronomy 94:55–133.

Banaticla-Hilario MCN, Sosef MSM, McNally KL, Hamilton NRS, van den Berg RG. 2013. Ecogeographic variation in the morphology of two Asian wild rice species, *Oryza nivara* and *Oryza rufipogon*. International Journal of Plant Science 174:896–909.

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D. 2003. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiology 132:84–91.

Bautista NS, Solis R, Kamijima O, Ishii T. 2001. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. Genes and Genetic Systems 76:71–79.

Biodiversity International, IRRI, WARDA. 2007. Descriptors for Wild and Cultivated Rice (*Oryza* spp.), Bioversity International, Rome, Italy; International Rice Research Institute, Los Baños, Philippines; WARDA, Africa Rice Center, Cotonou, Benin. 63 pp.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314–331.

Brookes G, Barfoot P. 2012. Global impact of biotech crops: Environmental effects, 1996–2010. GM crops and food: Biotechnology in Agriculture and the Food Chain 3:129–137.

Burkhardt PK, Beyer P, Wunn J, Kloti A, Armstrong GA, Schledz M, Lintig VJ, Potrykus I. 1997. Transgenic rice (*Oryza sativa*) endosperm expressing daffodil (*Narcissus pseudonarcissus*) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. The Plant Journal 11:1071–1078.

Cano-Calle D, Arango-Isaza RE, Saldamando-Benjumea C. 2015. Molecular identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) corn and rice strains in Colombia by using a PCR-RFLP of the mitochondrial gene cytochrome oxydase I (*COI*) and a PCR of the gene *FR* (for rice). Annals of the Entomological Society of America 108:172–180.

Chakanda R, van Treuren R, Visser B, van den Berg R. 2012. Analysis of genetic diversity in farmers' rice varieties in Sierra Leone using morphological and AFLP® markers. Genetic Resources and Crop Evolution 60:1237-1250.

Chandler S, Dunwell JM. 2008. Gene flow, risk assessment and the environmental release of transgenic plants. Critical Reviews in Plant Sciences 27: 25-49.

Chang TT, Marciano AP, Loresto GC. 1977. Morpho-agronomic variousness and economic potentials of *Oryza glaberrima* and wild species in the genus *Oryza*. In: Meeting on African Rice Species. IRAT-ORSTOM, Paris. pp 67–76.

Chapman RE, Wang J, Bourke AFG. 2003. Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. Molecular Ecology 12:2801–2808.

Chen H, Lin YJ, Zhang QF. 2009. Review and prospect of transgenic rice research. Chinese Science Bulletin 54:4049–4068.

Chen JS, Lin SC, Chen CY, Hsieh YT, Pai PH, Chen LK, Lee S. 2014. Development of a microarray for two rice subspecies: characterization and validation of gene expression in rice tissues. BMC Research Notes 7:1–18.

Chen LJ, Lee DS, Song ZP, Suh HS, Lu BR. 2004. Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Annals of Botany 93:67–73.

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E. 2003. Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67–75.

Choudhury B, Khan ML, Dayanandan S. 2013. Genetic structure and diversity of indigenous rice (*Oryza sativa*) varieties in the Eastern Himalayan region of Northeast India. SpringerPlus 2:228. http://www.springerplus.com/content/2/1/228. (Accessed 26/9/2014)

Chung JW, Park YJ. 2009. Population structure analysis reveals the maintenance of isolated sub-populations of weedy rice. Weed Research 50:606–620.

Cresswell JE, Osborne JL. 2004. The effect of patch size and separation on bumblebee foraging in oilseed rape: implications for gene flow. Journal of Applied Ecology 41:539–546.

Da Silva EF, Da Silva LM, Montalvan R. 2005. Crossing rate and distance in upland rice. Bragantia, Campinas 64:197–201.

Das B, Sengupta S, Parida SK, Roy B, Ghosh M, Prasad M, Ghose TK. 2013. Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. 14:71 http://www.biomedcentral.com/1471-2156/14/71. (Accessed 25/05/2015)

Dawe D, Pandey S, Nelson A. 2010. Emerging trends and spatial patterns of rice production. In: Pandey S, Byerlee D, Dawe D, Doberman A, Mohanta S, Rozelle S, Hardy B. (Eds.). Rice in the Global Economy: Strategic Research and Policy Issues for Food Security. International Rice Research Institute (IRRI). pp 15–35.

De Vicente C, Metz T, Alercia A. 2004. Descriptors for Genetic Markers Technologies. International Plant Genetic Resources Institute (IPGRI), Rome. 24 pp.

Devos Y, Reheul D, Schrijver AD. 2006. Considerations of cross-fertilization between GM and non-GM maize. ISB News Report March 2006. http://www.isb.vt.edu/articles/mar0603.htm. (Accessed 27/05/2014)

Diagne A, Assagba EA, Futakuchi K, Wopereis MCS. 2013. Estimation of cultivated area, number of farming households and yield for major rice-growing environments in Africa. In: Wopereis MCS, Johnson DE, Ahmadi N, Tollens E, Jalloh A. (Eds). Realizing Africa's Rice Promise. CAB International Oxfordshire, UK. pp 35–45.

Dong CF, Ding CL, Xu NX, Cheng YH, Shen YX, Gu HR. 2013. Double-purpose rice (*Oryza sativa* L.) variety selection and their morphological traits. Field Crops Research 149:276–282.

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM. 2000. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. Plant Physiology 124:1483–1492.

Duan S, Lu B, Li Z, Tong J, Kong J, Yao W, Li S, Zhu Y. 2007. Phylogenetic analysis of AA-genome *Oryza* species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. Biochemical Genetics 45:113–129.

Edwards JD, Janda J, Sweeney MT, Gaikwad AB, Liu B, Leung H, Galbraith DW. 2008. Development and evaluation of a high-throughput, low-cost genotyping platform based on oligonucleotide microarrays in rice. Plant Methods 4:13 doi:10.1186/1746-4811-4-13. http://www.plantmethods.com/content/4/1/13. (Accessed 27/05/2015)

Ellstrand NC. 2003. Current knowledge of gene flow in plants: implications for transgene flow. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 358:1163–1170.

Ellstrand NC, Prentice HC, Hancock JFE. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539–563.

Endo T, Sato H, Yamaguchi M, Kataoka T, Nakagomi K, Ito T, Mori K. 2009. Estimate of outcrossing rates in a rice plant (*Oryza sativa* L.) under field conditions using a purple grain rice cultivar, *Okunomurasaki*. Breeding Science 59:195–202.

Fang C, Zhuang Y, Xu T, Li Y, Li Y, Lin W. 2013. Changes in rice allelopathy and rhizosphere microflora by inhibiting rice phenylalanine ammonia-lyase gene expression. Journal of Chemical Ecology 39:204–212.

FAOSTAT. 2015. http://faostat3.fao.org/download/Q/QC/E. (Accessed 27/05/2015)

Gao L, Innan H. 2008. Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japónica* demonstrated by multilocus microsatellites. Genetics 179:965–976.

García de Salamone IE, Funes JM, Di Salvo LP, Escobar-Ortega JS, D'Auria F, Ferrando L, Fernandez-Scavino A. 2012. Inoculation of paddy rice with *Azospirillum brasilense* and *Pseudomonas fluorescens*: Impact of plant genotypes on rhizosphere microbial communities and field crop production. Applied Soil Ecology 61:196–204.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638.

Gealy D, Mitten D, Rutger N. 2003. Gene flow between red rice and herbicide resistant rice: Implications for weed management. Weed Technology 17:627–645.

Ghoshal D, Sen SK, Goyal A. 2001. Introduction and expression of cowpea trypsin inhibitor (*CpTI*) gene in transgenic tobacco. Journal of Plant Biochemistry and Biotechnology 10:19–24.

Glaszmann JC. 1987. Isozymes and classification of Asian rice varieties. Theoretical and Applied Genetics 74:21–30.

GRIN (Germplasm Resources Information Network). 2015. USDA, ARS, National Genetic Resources Program. (Online Database). National Germplasm Resources Laboratory, Beltsville, Maryland. http://www.ars-grin.gov.4/cgi-bin/npgs/html/taxdumplist.pl (Accessed 11/09/2015)

GRiSP. 2010. CGIAR Thematic Area 3: Susceptible crop productivity increase for global food security. A CGIAR Research Program on Rice-Based Production Systems. November 2010.

Grodzicker T, Williams J, Sharp P, Sambrook J. 1975. Physical mapping of temperature-sensitive mutantions of adenoviruses. Cold Spring Harbor Symposia on Quantitative Biology 39:439–446.

Gross BL, Steffen FT, Olsen KM. 2010. The molecular basis of white pericarps in African domesticated rice: novel mutations at the *Rc* gene. Journal of Evolutionary Biology 23:2747–2753.

He Z, Jiang X, Ratnasekera D, Grassi F, Perera U, Lu BR. 2014. Seed-mediated gene flow promotes genetic diversity of weedy rice within populations: Implications for weed management. PLoS ONE 9:1-21.

Hirano HY, Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E, Sano Y. 1994. Retrotransposition of a plant SINE into the *wx* locus during evolution of rice. Journal of Molecular Evolution 38:132–137.

Hokanson K. 2006. Confinement during field testing of self-pollinated plant made pharmaceutical and plant made industrial crops using rice as a model. In: Rose R, McCammon S, Lively S. (Eds.). Proceedings: Workshop on Confinement of Genetically Engineered Crops During Field Testing, September 13–14, 2004. pp 23–34.

Hoshida H, Tanaka Y, Hibino T, Hayashi Y, Tanaka A, Takabe T, Takabe T. 2000. Enhanced tolerance to salt stress in transgenic rice that overexpresses chloroplast glutamine synthetase. Plant Molecular Biology 43:103–111.

Hoyle M, Cresswell, JE. 2007. The effect of wind direction on cross-pollination in wind-pollinated GM crops. Ecological Applications 17:1234–1243.

Huang P, Molina J, Flowers JM, Rubinstein S, Jackson SA, Purugganan MD, Schaal BA. 2012. Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. Molecular Ecology 21:4593–4604.

Ikeda R, Sokei Y, Akintayo I. 2009. Seed fertility of F_1 hybrids between upland rice NERICA cultivars and *Oryza sativa* L. or *O. glaberrima* Steud. Breeding Science 59:27–35.

Inui H, Shiota N, Ido Y, Inoue T, Hirose S, Kawahigashi H, Ohkawa Y, Ohkawa H. 2001. Herbicide metabolism and tolerance in the transgenic rice plants expressing human CYP2C9 and CYP2C19. Pesticide Biochemistry and Physiology 71:156–169.

IRGSP (International Rice Genome Sequencing Projcet). 2005. The map-based sequence of the rice genome. Nature 436:793–800.

IRRI. 2002. Standard Evaluation System (SES) for Rice. International Rice Research Institute (IRRI), Los Baños, Philippines.

ISAAA 2014. ISAAA Brief 49. Executive Summary: Global Status of Commercialized Biotech/GM Crops. http://www.isaaa.org/resources/publications/briefs/44/executivesummary/. (Accessed 17/05/2015)

ISB (Information Systems for Biotechnology) 2004. Field test releases in the United States. (1 June 2005; www.isb.vt.edu). (Accessed 18/09/2013)

Jaccoud D, Peng K, Feinstein D, Kilian A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29:e25.

Jagadish SVK, Craufurd PQ, Wheeler TR. 2007. High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). Journal of Experimental Botany 58:1627–1635.

Jagadish SVK, Muthurajan R, Oane R, Wheeler TR, Heuer S, Bennett J, Craufurd PQ. 2010. Physiological and proteomic approaches to address heat tolerance during anthesis in rice (*Oryza sativa* L.). Journal of Experimental Botany 61:143–156.

Jeong YP, Kim KH. 2005. RFLP marker linked to red pericarp color gene of Korean local rice variety, Jakwangdo. Korean Journal of Breeding 37:96–100.

Jiang Z, Xia H, Basso B, Lu BR. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. Theoretical and Applied Genetics 124:309–322.

Jones MP, Dingkuhn M, Aluko GK, Semon M. 1997. Interspecific *Oryza sativa* L x *O. glaberrima* Steud. progenies in upland rice improvement. Euphytica 92:237–246.

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

Kanazawa A, Akimoto A, Morishima H, Shimamoto Y. 2000. Inter- and intra-specific distribution of stowaway transposable elements in AA genome of wild rice. Theoretical and Applied Genetics 101:327–335.

Kanya JI, Kinyamario JI, Amugune NO, Hauser TP. 2009. Dispersal distance of rice (*Oryza sativa* L.) pollen at the Tana River delta in the Coast Province, Kenya. African Journal of Biotechnology 8:2265–2270.

Karasawa MMG, Vencovsky R, Silva CM, Zucchi MI, Oliveira GCX, Veasey EA. 2007. Genetic structure of Brazilian wild rice (*Oryza glumaepatula* Steud., Poaceae) populations analyzed using microsatellite markers. Genetics and Molecular Biology 30:400–410.

Karthigeyan S, Rajkumar S, Sharma R, Gulati A, Sud R, Ahuja P. 2008. High level of genetic diversity among the selected accessions of tea (*Camellia sinensis*) from abandoned tea gardens in western Himalaya. Biochemical Genetics 46:810–819.

Kawakami S, Ebana K, Nishikawa T, Sato Y, Vaughan DA, Kadowaki K. 2007. Genetic variation in the chloroplast genome suggests multiple domestication of cultivated Asian rice (*Oryza sativa* L.). Genome 50:180–187.

Kawano N, Sakagami J. 2008. Submergence stress to rice plants in West Africa. In: Sakagami J., Ito O. (Eds.). Forefront of Rice Cultivation in Africa. JIRCAS Working Report No. 57. Japan International Research Center for Agricultural Sciences, Ibaraki, Japan, pp 25–36.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25–34.

Kimura M, Weiss GH. 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.

Ku MSB, Cho DH, Li X, Jiao DM, Pinto M, Miyao M, Matsuoka M. 2001. Introduction of genes encoding C4 photosynthesis enzymes into rice plants: Physiological consequences. In: Goode JA, Chadwick D. (Eds.). Rice Biotechnology: Improving Yield, Stress Tolerance, and Grain Quality. New York: Wiley Novartis Foundation Symposium no. 236, pp 100–116.

Kumagai M, Wang L, Ueda S. 2010. Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. Gene 462:44–51.

Kumar R, Singh AK, Kumar A, Radha. 2012. Evaluation of genetic diversity in rice using simple sequence repeats (SSR) markers. African Journal of Biotechnology 11:14956–14995.

Kuroda Y, Sato YI, Bounphanousay C, Kono Y, Tanaka K. 2007. Genetic structure of three *Oryza* AA genome species (*O. rufipogon, O. nivara and O. sativa*) as assessed by SSR analysis on the Vientiane Plain of Laos. Conservation Genetics 8:149–158.

Kwon SJ, Lee JK, Hong SW, Park YJ, McNally KL, Kim NS. 2006. Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by rim2/hipa cacta transposon display. Genes and Genetic Systems 81:93–101.

Langevin SA. 1990. The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa* L.). Evolution 44:1000–1008.

Lee J, Park JH, Koh HJ. 2013. Morphological and genetic characterization of off-type rice plants collected from farm fields in Korea. Journal of Plant Biology 56:160–167.

Li ZM, Zheng XM, Ge S. 2011. Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. Theoretical and Applied Genetics 123:21–31.

Linares OF. 2002. African rice (*Oryza glaberrima*): History and future potential. PNAS 99:16360–16365.

Liu C, Zhang G. 2006. Single nucleotide polymorphism (SNP) and its application in rice. Yi Chuan (Hereditas in Chinese) 6:737–744.

Liu F, Xu W, Songa Q, Tana L, Liu J, Zhu Z, Fu Y, Su Z, Sun C. 2013. Microarray-assisted fine-mapping of quantitative trait loci for cold tolerance in rice. Molecular Plant 6:757–767.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. PNAS 103:9578–9583.

Lu BR, Snow AA. 2005. Gene flow from genetically modified rice and its environmental consequences. BioScience 55:669–678.

Lu BR, Zheng KL, Qian HR, Zhuang JY. 2002. Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics 106:101–106.

Ma TL, Wu WH, Wang Y. 2012. Transcription analysis of rice root responses to potassium deficiency. BMC Plant Biology 12:161 http://www.bimedcentral.com/147-2229/12/161. (Accessed 16/04/2015)

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N. 1997. Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. Theoretical and Applied Genetics 94:39–45.

Matsui T, Omasa K, Horie T. 2000. High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (*Oryza sativa* L.). Plant Production Science 3:430–434.

Messeguer J, Fogher C, Guiderdoni E, Marfa V, Catala MM, Baldi G, Mele E. 2001. Field assessments of gene flow from transgenic to cultivated rice (*Oryza sativa* L.) using a herbicide resistance gene as trace marker. Theoretical and Applied Genetics 103:1151–1159.

Messeguer J, Marfà V, Català MM, Guiderdoni E, Melé E. 2004. A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. Molecular Breeding 13:103–112.

Mohapatra S. 2012. Ethiopia's millennium crop. In: Rice Today. Jan-Mar, 2012.

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. PNAS 108:8351–8356.

Morishima H. 2001. Evolution and domestication of rice. In: Khush GS, Brar DS, Hardy B. (Eds.). Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000, Los Baños, Philippines. Enfield, NH (USA). Science Publishers, Inc., and Los Baños (Philippines): International Rice Research Institute. pp 63–77.

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y. 2002. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment of SNP markers. DNA Research 9:163–171.

Nei M. 1973. Analysis of gene diversity in subdivided populations. PNAS 70:3321–3323.

Neigel JE. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology and Systematics 28:105–128.

Nuijten E, Almekinders DCJM. 2008. Mechanisms explaining variety naming by farmers and name consistency of rice varieties in The Gambia. Economic Botany 62:148–160.

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P. 2009. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers‘ fields. PLoS ONE 4:1–9.

OECD (Organisation for Economic Co-operation and Development). 1999. OECD’s Database of Field Trials. (1 June 2005; http://webdomino1.oecd.org/ehs/biotrack.nsf). (Accessed 14/07/14)

Phan PDT, Kageyama H, Ishikawa R, Ishii T. 2012. Estimation of the outcrossing rate for annual Asian wild rice under field conditions. Breeding Science 62:256–262.

Phung NTP, Mai CD, Mournet P, Frouin J, Droc G, Ta NK, Jouannic S, Lê LT, Do VN, Gantet P, Courtois B. 2014. Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. BMC Plant Biology 14:371-377.

Portères R. 1970. Primary cradles of agriculture in the African continent. In: Fage JD, Oliver RA. (Eds.). Papers in African Prehistory. Cambridge: Cambridge University Press. pp 43–58.

Rakshit S, Rakshit A, Matsumura H, Takahashi Y, Hasegawa Y, Ito A, Ishii T, Miyashita NT, Terauchi R. 2007. Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theoretical and Applied Genetics 114:731–743.

Rong J, Lu BR, Song Z, Su J, Snow AA, Zhang X, Sun S, Chen R, Wang F. 2006. Dramatic reduction of crop-to-crop gene flow within a short distance from transgenic rice fields. New Phytologist 173:346–353.

Rong J, Xia H, Zhu Y, Wang Y, Lu BR. 2004. Asymmetric gene flow between traditional and hybrid rice varieties (*Oryza sativa*) indicated by nuclear simple sequence repeats and implications for germplasm conservation. New Phytologist 163:439–445.

Saito K, Nelson A, Zwart SJ, Niang A, Sow A, Yoshida H, Wopereis MCS. 2013. Towards a better understanding of biophysical determinants of yield gaps and the potential for expansion of the rice area in Africa. In: Wopereis MCS, Johnson DE, Ahmadi N, Tollens E, Jalloh A. (Eds.). Realizing Africa’s Rice Promise. CAB International Oxfordshire, UK. pp 188–203.

Sanni KA, Fawole I, Guei RG, Ojo DK, Somado EA, Tia DD, Ogunbayo SA, Sanchez I. 2008. Geographical patterns of phenotypic diversity in *Oryza sativa* landraces of Cote d'Ivoire. Euphytica 160:389–400.

Sano Y, Sano R. 1990. Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. Genome 33:209–218.

Sanyal A, Ma J, San Miguel P, Jackson SA, Ammiraju JSS, Yu Y, Rambo T, Currie J, Kollura K, Kim HR, Wing RA, Lu F, Chen J, Mingsheng C. 2010. Orthologous comparisons of the *Hd1* region across genera reveal *hd1* gene availability within diploid *Oryza* species and disruptions to microsynteny in sorghum. Molecular Biology and Evolution 27:2487–2506.

Sarla N, Swamy BPM. 2005. *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. Current Science 89:955–963.

Satake T, Yoshida S. 1978. High temperature induced sterility in indica rices at flowering. Japanese Journal of Crop Science 47:6–17.

Seck PA, Diagne A, Mohanty S, Wopereis MCS. 2012 Crops that feed the world 7: Rice. Food Security 4:7–24.

Second G, Bezançon G, Bozza J. 1977. Variabilité d'*Oryza longistaminata* et du complexe sativa des *Oryza* en Afrique: Aspects ecologiques et evolutifs. In: Meeting on African Rice Species. IRAT-ORSTOM, Paris. pp 47–55.

Semagn K, Bjørnstad Å, Ndjiondjop MN. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540–2568.

Semon M, Nielsen R, Jones MP, McCouch SR. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. Genetics 169:1639–1647.

Shivrain VK, Burgos NR, Sales MA, Mauromoustakos A, Gealy DR, Smith KL, Black, HW, Jia M. 2009. Factors affecting the outcrossing rate between Clearfield™ rice and red rice (*Oryza sativa*). Weed Science 57:394–403.

Slatkin M. 1985. Geneflow in natural populations. Annual Review of Ecology and Systematics 16:393-430.

Somaratne LHMYK, Abayawickrama ASMT, Wickramasinghe IP, Samarasinghe WLG. 2012. Estimating out-crossing rate of Bg 379-2 using morphological markers and confirmation by molecular markers. Rice Science 19:166–168.

Song Z, Lu BR, Chen J. 2004. Pollen flow of cultivated rice measured under experimental conditions. Biodiversity and Conservation 13:579–590.

Song ZP, Lu BR, Zhu YG, Chen JK. 2003. Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657–665.

Sork VL, Nason J, Campbell DR, Fernandez JF. 1999. Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology and Evolution 14:219–224.

Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98:503–517.

Sow M, Sido A, Laing M, Ndjiondjop MN. 2014. Agro-morphological variability of rice species collected from Niger. Plant Genetic Resources 12:22–34.

Sweeney M, McCouch S. 2007. The complex history of the domestication of rice. Annals of Botany 100:951–957.

Tateoka T. 1963. Taxonomic Studies of the Genus *Oryza*, In: Chang TT. (Ed.). Rice Genetics and Cytogenetics: Proceedings of the Symposium. International Rice Research Institute (IRRI), Los Banos. pp 15–21.

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 Research 11:1441–1452.

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho Y.G, Ishii T, McCouch SR 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theoretical and Applied Genetics 100:697–712.

Vasconcelos M, Datta K, Oliva N, Khalekuzzaman M, Torrizo L, Krishnan S, Oliveira M, Goto F, Datta SK. 2003. Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. Plant Science 164:371–378.

Vaughan DA, Lu BR, Tomooka N. 2008. The evolving story of rice evolution. Plant Science 174:394–408.

Vaughan DA, Morishima H, Kadowaki K. 2003. Diversity in the *Oryza* genus. Current Opinion in Plant Biology 6:139–146.

Vitte C, Ishii T, Lamy F, Brar D, Pannaud O. 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Molecular Genetics and Genomics 272:504–511.

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.

Wang M, Yu Y, Haberer G, Marri PR, Fan C. 2014. The genome sequence of African rice *(Oryza glaberrima*) and evidence for independent domestication. Nature Genetics 46:982–988.

Wang S, Liu X, Xu C, Shi L, Zhang X**,** Ding D, Wang Y. 2007. Genome polymorphisms between Indica and Japonica revealed by RFLP. Agricultural Sciences in China 6:108–114.

WARDA (West Africa Rice Development Association). 1994. Training in mangrove rice production: Instructors' manual. WARDA Bouake, Ivory Coast.

WARDA (West Africa Rice Development Association). 2000. WARDA annual report. 1999. Bouake, Cote d'Ivoire.

WARDA (West Africa Rice Development Association). 2004. WARDA annual report. 2003. Bouake, Cote d'Ivoire.

WARDA (West Africa Rice Development Association). 2006. WARDA annual report. 2005. Bouake, Cote d'Ivoire.

Weerakoon WMW, Abeywickrama T, de Costa WAJM, Maruyama A. 2008. Out-crossing of heat stress affected spikelets of lowland rice in the sub-humid zone of Sri Lanka and its long-term Implications. http://www.niaes.affrc.go.jp/marco/marco2009/english/W2-05_Weerakoon_WMW_P.pdf. (Accessed 11/04/2013)

Wei X, Qiao WH, CHen YT, Wang RS, Cao LR, Zhang WX, Yuan NN, Li ZC, Zeng HL, Yang QW. 2012. Domestication and geographic origin of *Oryza sativa* in China: insights from multilocus analysis of nucleotide variation of *O. sativa* and *O. rufipogon*. Molecular Ecology 21:5073–5087.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18:7213–7218.

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. PNAS 101:9915–9920.

Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. Heredity 82:117–125.

Williams JGK, Kublelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531–6535.

Willocquet L, Noel M, Sackville HR, Savary S. 2012. Susceptibility of rice to sheath blight: an assessment of the diversity of rice germplasm according to genetic groups and morphological traits. Euphytica 183:227–241.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Wright S. 1969. Evolution and the Genetics of Populations. The Theory of Gene Frequencies, University of Chicago Press.

Wu A, Sun X, Pang Y, Tang K. 2002. Homozygous transgenic rice lines expressing GNA with enhanced resistance to the rice sap-sucking pest *Laodelphax striatellus*. Plant Breeding 121:93–95.

Xiang CC, Chen Y. 2000. cDNA microarray technology and its applications. Biotechnology Advances 18:35–46.

Xu DP, Xue QZ, McElroy D, Mawal Y, Hilder VA, Wu R. 1996. Constitutive expression of a cowpea trypsin inhibitor gene, *CpTI*, in transgenic rice plants confers resistance to two major rice insect pests. Molecular Breeding 2:167–173.

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W. 2012. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature Biotechnology 30:105–111.

Yamanaka S, Nakamura I, Nakai H, Sato Y. 2003. Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. Genetic Resources and Crop Evolution 50:529–538.

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92.

Zhang NS, Linscombe S, Oard J. 2003. Outcrossing frequency and genetic analysis of hybrids between transgenic glusfosinate herbicide-resistant rice and the weed, red rice. Euphytica 130:35–45.

Zhao Y, Vrieling K, Liao H, Xiao M, Zhu Y, Rong J, Zhang W, Wang Y, Yang J, Chen J, Song Z. 2013. Are habitat fragmentation, local adaptation and isolation-by-distance driving population divergence in wild rice *Oryza rufipogon*? Molecular Ecology 22:5531–5547.

Zheng X, Ge S. 2010. Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). Molecular Ecology 19:2439–2454.

Zhou HF, Zheng XM, Wei RX, Second G, Vaughan DA, Ge S. 2008. Contrasting population genetic structure and gene flow between *Oryza rufipogon* and *Oryza nivara*. Theoretical and Applied Genetics 117:1181–1189.

Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ. 1998. AFLP markers for the study of rice biodiversity. Theoretical and Applied Genetics 96:602–611.

Zhu K, Min C, Xia H, Yang Y, Wang B, Chen K. 2014. Characterisation of indica special protein (ISP), a marker protein for the differentiation of *Oryza sativa* subspecies *indica* and *japonica*. International Journal of Molecular Sciences 15:7332–7343.

Ziska LH, Gealy DR, Tomecek MB, Jackson AK, Black HL. 2012. Recent and projected increases in atmospheric CO_2 concentration can enhance gene flow between wild and genetically altered rice (*Oryza sativa*). PLoS ONE 7:1-6.

Chapter 3

Assessing diversity of rice varieties using agro-morphological characterisation

3.1 Abstract

Information on germplasm diversity and relationships is essential to help design experiments on gene flow. In this study 26 agro-morphological traits were used to assess the genetic diversity of 36 rice accessions comprising *O. sativa*, *O. glaberrima* and interspecific (*O. sativa* x *O. glaberrima*) hybrids. The Shannon-Weaver diversity index was low (H*'* = 0.23 ± 0.02). The highest level of diversity was detected in the selected Indica accessions. (H' = 0.32 ± 0.01) and the lowest diversity (H' = 0.18 ± 0.07) in the selected *O. glaberrima* accessions. Results of the phenotypic frequency showed that the selected Japonica accessions had brown or green apiculus and were awnless. Selected *O. glaberrima* and landrace accessions had purple apiculus and some were awned. Selected accessions of the Japonica group comprised mostly of improved varieties with a white pericarp whilst *O. glaberrima* accessions and landraces had a red pericarp. Selected interspecific hybrids combined traits of *O. sativa* and *O. glaberrima* to varying degrees. Principal component analysis revealed that flag leaf length, flag leaf width, 1000-grain weight, percentage grain filling and panicle length account for most of the variation. Hierarchical clustering revealed four clusters. Cluster 1 was composed of germplasm not adapted to Africa; cluster 2 contained *O. glaberrima* accessions; cluster 3 had improved varieties comprising the Japonica group and interspecific hybrid accessions and cluster 4 contained Indica landraces.

Keywords: Agro-morphological, diversity**,** interspecific hybrids, *O. glaberrima*, *O. sativa*

3.2 Introduction

Diversity of a population refers to the amount and combinations of different genes and phenotypic traits within the population. It reflects the totality of different alleles and genotypes within the gene pool of the population. This is expressed in appearance as different phenotypes in a given environment. Identification of homogenous groups of individuals is important in designing experiments to determine outcrossing in plants. Groups of similar individuals could be classified into one population and be treated alike (Evanno *et al.,* 2005). Plant breeding has contributed to yield increase and improvement in quality of food and industrial crops. However, the widespread adoption of modern cultivars and changing patterns in cropping practices and the climate may affect genetic diversity of cultivated crops and could threaten global agricultural production (Hammer and Teklu, 2008). Conservation of plant genetic diversity is important for species

survival and evolution. Conservation of germplasm, that habour alleles that make plants tolerant or resistant to stresses, is also necessary for crop production under various climatic, biotic and abiotic constraints. Assessing genetic diversity is a major focus in plant studies. Knowledge about relationships between genotypes is used in identifying suitable pollen donor and recipient plants with desired traits for assessing gene flow between plants. It is also useful in identifying parental material and planning new crosses. Such knowledge can be obtained from several sources which can be geographic, pedigree or from information about plant characteristics. Plant characteristics have been found to be reliable in assessing relationships between genotypes. Knowledge of agro-morphological diversity and the distribution pattern of variation are useful for conservation and management of germplasm, seed production and crop improvement (Gao, 2003; Sanni *et al.*, 2008). Diversity studies have revealed two forms of cultivated rice namely *O. sativa* and *O. glaberrima*. *Oryza sativa* was domesticated some 10000–14000 years ago in China and *O. glaberrima* some 3500 years ago in the Niger River delta in West Africa (Portères, 1970; Chang, 1976; 1984; Khush, 1997;). *Oryza glaberrima* has been found to be less diverse than *O. sativa* (Barry *et al*., 2007; Li *et al*., 2011). Several studies have reported outcrossing between *O. sativa* and *O. glaberrima*. (Song *et al*., 2003; 2004; Semon *et al*., 2005; Nuijten *et al*., 2009 Phan *et al*., 2012).

Agro-morphological characterisation has been useful in the evaluation of genetic diversity of plant genotypes. It largely involves scoring of characters that are related to the physical appearance of the plant. Suitable traits are those that can easily identify the genotype. Such traits are usually highly heritable, expressed in all environments and easily seen by the eye. Traits can be qualitative, quantitative or quasi-quantitative. A qualitative trait like anthocyanin colouration is either present or absent. Qualitative traits can be dominant or recessive. Quantitative traits like plant height or leaf width are measurable and continuous. Agro-morphological traits can also be quasi-quantitative like the intensity of green colour of a leaf. The expression of these traits in a genotype is influenced by the constituent genes. Therefore differences in genotypes with regards to inherent traits are a reflection of differences in the genes and indication of genetic relationships. Therefore agro-morphological traits can be used to effectively assess diversity of plant genotypes (Nuijten and Almekinders, 2008) and can also be used to identify genotypes with desired traits for evaluating outcrossing.

Agro-morphological characterisation has the advantage of not requiring sophisticated equipment or complex experiments. Traits are relatively simple and can be rapidly scored. There are also readily available statistical procedures for morphological trait analysis. Morphological

characterisation has proved to be an invaluable tool in the evaluation of relationships between genotypes. Rice scientists use agro-morphological traits to characterise and distinguish genotypes and assess their diversity. Farmers have also been reported to use agro-morphological traits to characterise and name rice varieties (Nuijten and Almekinders, 2008). However, agro-morphological traits that are well suited for evaluating germplasm are few, thus covering only a small proportion of the genome. They are also characterised by epistasis, pleiotropy and dominant-recessive relationships, further limiting their value as genetic markers (Smith and Smith, 1992).

Agro-morphological characterisation has been used in several studies to identify appropriate accessions and traits for the analysis of outcrossing in rice. Da Silva *et al*. (2005) used leaf pilosity to study outcrossing rates in upland rice. They reported outcrossing rates of 0.30–0.35% at 0.5 m from the donor. Endo *et al*. (2009) used purple grain and ligule colour to assess outcrossing in rice and observed hybrid plants at up to 20 m from the donor. Somaratne *et al*. (2012) also used short round grain and purple colour culm of donor to estimate outcrossing in a rice variety. They reported potential and average outcrossing rates of 3.41% and 1.29% respectively between the selected donor and recipient plants.

Evaluation of plant characteristics is a first step in the study of genetic diversity and relationships of germplasm. In their study of rice germplasm from Niger, Sow *et al*. (2014) assessed the variability of 270 rice accessions. They distinguished three groups based on agro-morphological traits. Another study of rice germplasm by Dong *et al*. (2013) indicated morphological traits that are suitable for the selection of double-purpose rice varieties. Willocquet *et al*. (2012) studied the susceptibility of 200 rice accessions to sheath blight. They reported a strong association between morphological traits and disease intensity. Babaei *et al*. (2011) used morphological markers to study the m_2 generation of mutant rice cultivars. Moukoumbi *et al*. (2011) successfully characterised a collection of 78 rice accessions into distinct groups adapted to specific ecologies. Sanni *et al*. (2008) used plant traits to assess diversity in rice landraces. Agro-morphological characterisation continues to be widely used in rice diversity studies (Ogunbayo *et al*., 2005; Yi *et al*., 2005; Fogliatto *et al*., 2011; Lee *et al*., 2013; Mazid *et al*., 2013; Shinada *et al*., 2013).

Agro-morphological characterisation is an effective tool to identify dominant traits suitable for outcrossing trials in rice and is also effective in assessing diversity in rice germplasm. It is one of the earliest methods used to assess diversity. It still remains an invaluable tool for plant diversity studies. In the present study, diversity of 36 rice accessions was evaluated using 26 agro-

morphological traits. Results were used to assign accessions into distinct clusters from which donor and recipient varieties were selected for subsequent outcrossing trials. Days to heading was used to determine planting dates of different varieties used in outcrossing trials so that flowering synchronisation was assured.

3.3 Materials and methods

Thirty six accessions comprising *O. glaberrima*, *O sativa* and *O. glaberrima* x *O. sativa* hybrids were utilised for the study (Table 3.1; Appendix 1).

Table 3.1 List of varieties used in the study

No.	Variety name	Taxon/Group	No.	Variety name	Taxon/Group
1	Afhikari	Japonica	19	WAB365-B-2-H1-HB	Japonica
2	PL87-3	*glaberrima*	20	WAB384-B-11-H2-H1-H1B	Japonica
3	ITA123	Indica	21	NERICA 12	Interspecific
4	ITA150	Indica	22	NERICA 13	Interspecific
5	WAB56-104	Japonica	23	WAB506-125-3	Japonica
6	Moroberekan	Japonica	24	WAB519-55-3	Japonica
7	NERICA 4	Interspecific	25	NERICA 2	Interspecific
8	WAB100	Japonica	26	WAB96-1-1	Japonica
9	WAB128	Japonica	27	IRAT104	Indica
10	WAB176	Japonica	28	B6144F	Indica
11	WAB217	Japonica	29	NERICA 14	Interspecific
12	WAB224	Japonica	30	NERICA 15	Interspecific
13	WAB272	Japonica	31	NERICA 16	Interspecific
14	WAB285	Japonica	32	NERICA 18	Interspecific
15	WAB306	Japonica	33	TOS15505	Indica
16	WAB307	Japonica	34	TOS15729	Indica
17	NERICA 7	Interspecific	35	TOS8076	Indica
18	WAB337	Japonica	36	CG14	*glaberrima*

The *O. sativa* material was composed of Indica and Japonica variety groups. Accessions were chosen on the basis of being adapted to the upland ecology and also widely used varieties in Africa. Materials were selected from a pool of AfricaRice mega varieties adapted to prevailing conditions in Africa. There was a large proportion of Japonica accessions in the collection. This is because they are the most suitable and widely used varieties in this ecology. All plant material was acquired from the gene bank of the Africa Rice Center, Cotonou, Benin. Field experiments were conducted at the Africa Rice Center, Cotonou, Benin. Cotonou is in the Littoral District in southern Benin, located at 6°25′15″N 2°19′35″E / 6.42083°N 2.32639°E at an altitude of 15.5 metres above sea level (MASL). Two experiments were conducted under upland irrigated conditions. The first was from June–October 2012 and the second from November 2012–March 2013.

An alpha lattice design with three replications was used. Each plot was 3 m x 3 m. The planting method was direct seeding. Three seeds were sown per hole which was thinned to one plant per stand at 14 days after sowing at a density of 20 cm x 20 cm. NPK (15:15:15) fertiliser was used as basal application at sowing at the rate of 200 kg ha^{-1}. Urea (46% N) was applied as top dressing at the rate of 46 kg N ha^{-1} in two equal splits; first at tillering and then at panicle initiation stage. Weeding and other cultural practices were done manually as and when required.

Data was collected from plants in middle rows (2.2 m x 2.2 m) in each plot leaving out two border rows on each side. Harvesting and threshing was done manually. Data was collected on agro-morphological traits according to the descriptors for rice developed by Bioversity International *et al*. (2007). Data was collected on 16 quantitative and 10 qualitative agro-morphological traits (Table 3.2).

3.4 Data analysis

3.4.1 Qualitative traits

The data was analysed using AGROBASE (2013) software. Variability between genotypes was estimated by subjecting the data to clustering and principal component analysis (PCA) using XLSTAT software (Addinsoft, 2008). Qualitative traits are particularly suitable to use in identifying plant varieties and desirable traits. They are highly controlled by the genetic background of the plant. As such, they are relatively more stable and less influenced by genotype by environment (GxE) interaction. In most cases they can easily be used to identify a genotype and expressed in all environments. In this study 10 qualitative traits were evaluated. The frequency of each trait

category was calculated. The Shannon-Weaver diversity index (Shannon and Weaver, 1949) was used to assess the phenotypic diversity for each trait. As the frequency of each trait indicates the contribution of that trait to the total diversity in the population, it was used to compute the diversity index.

The standardised Shannon-Weaver diversity index (H') was computed as:

$$\mathrm{H}' = -\sum_{i=1}^{n} P_i log_2 P_i,$$

where *n* is the number of phenotypic classes for a character and P_i is the relative abundance of individuals in the i^{th} class, which was calculated as the ratio of individuals in the phenotypic class *i* to the total number of individuals (*n*). Each standardised H' value was normalised by dividing it by minus its maximum value ($-log_2 n$) to give standardised H' values between 0 and 1 (Perry and McIntosh, 1991; Sanni *et al*., 2008). Various traits were pooled across different accessions and then additive properties of standardised H' were used to evaluate predominance of one trait state within the population.

3.4.2 Quantitative traits

Sixteen quantitative traits were evaluated. The data was analysed using AGROBASE (2013) software. Analysis of variance (ANOVA) was done separately for each season and for the two seasons combined. The data set was standardised as suggested by Ruiz *et al.* (1997). This was done by subtracting the mean value of the trait for each observation and then dividing it by its standard deviation. Pearson PCA was performed for traits and accessions assigned into groups by an agglomerative hierarchical clustering (AHC) method. Dissimilarities were computed based on the Euclidean distance and the Ward method (Ward, 1963) for the aggregation of accessions. These analyses were performed using XL-STAT 2008 application package.

Heritability, which was calculated using AGROBSE (2013) software, was broad sense. This refers to the proportion of phenotypic variance due to all genetic effects. Broad sense heritability can be calculated as the ratio of the genotypic variance to phenotypic variance according to Allard (1960) as:

$$\mathrm{H}^2 = \left(\frac{\sigma_g^2}{\sigma_{ph}^2}\right) \times 100$$

where H^2 = broad sense heritability (%), σ_g^2 = genotypic variance and $\sigma_{ph}^2 =$ phenotypic variance.

The principal axis method was used to extract the components, which was followed by a varimax (orthogonal) rotation. Only components and variables that met minimum standards for selection were retained and presented. The PCA was done using the eigenvalue-one criterion test (Kaiser, 1960). Only components with an eigenvalue of greater than or equal to 1.0 were retained for interpretation. The scree test (Cattell, 1966) was also done to decide on the number of components to retain. PCA was further subjected to a minimum proportion of variance test. For the latter only components that accounted for at least 18% of variability were retained. The first three components were selected and retained for rotation.

3.5 Results

3.5.1 Diversity as revealed by qualitative traits

Diversity level based on the H' index as revealed by the qualitative traits is shown in Table 3.3 and Figure 3.1. Some of the qualitative traits like ligule shape and ligule colour, were monomorphic and were not presented here. There were differences between as well as within variety groups. The average diversity across all accessions based on H' revealed by the qualitative traits was 0.23 ± 0.02. Diversity ranged from 0.14 to 0.31. Selected *Oryza sativa* accessions had an average diversity of 0.27 ± 0.02. Selected interspecific hybrid accessions had a diversity of 0.24 ± 0.04. Within *O. sativa*, accessions of the Indica group showed the highest diversity of 0.32 ± 0.01. Accessions of Japonica group and interspecific hybrids had a diversity of 0.24 ± 0.04.

Apiculus colour

Three colours were observed for apiculus colour namely: brown, green and purple; 44% of accessions had brown and 42% had green apiculus colour. Purple colour of apiculus was revealed in 14% of accessions (Figure 3.1B). Apiculus colour revealed the highest diversity level of all qualitative traits; it had a value of H' = 0.31 Accessions of Japonica group showed brown (47%) or green (53%) apiculus in almost equal proportions. On the other hand, 78% of all interspecific hybrid accessions had brown apiculus colour. For Indica group accessions 57% had green and another 33% and 10% had purple and brown apiculus colour respectively.

Table 3.2 Parameters measured for the agro-morphological characterisation of genotypes (Bioversity International *et al*., 2007)

S/N	Quantitative traits	
1	Flag leaf length (FLL) (cm)	Measured from the ligule to the tip of the blade on five representative plants
2	Flag leaf width (FLW) (cm)	Measured at the widest portion of the flag leaf on five representative plants
3	Leaf blade length (LBL) (cm)	Measured on leaf just below the flag leaf on the main culm, from the ligule to the tip of the blade, on five representative plants
4	Leaf blade width (LBW) (cm)	Measured at the widest portion on leaf just below the flag leaf on the main culm on five representative plants
5	Panicle exsertion (PnEx) (%)	Expressed as: length of panicle from neck to tip/total length (part exserted + portion within leaf sheath)
6	Panicle length (PnL) (cm)	Length of main axis of panicle measured from the panicle base to the tip (average of five representative plants)
7	Panicle number per plant (PnNum)	Recorded as the average of total number of panicles per plant on five representative plants
8	Plant height (PHt) (cm)	Measured from ground level to the base of the panicle on five representative plants

Table 3.2 continued. Parameters measured for the agro-morphological characterisation of genotypes (Biodiversity International *et al*., 2007)

S/N	Quantitative traits	
9	Spikelet opening angle (SOA) (^{0})	Measured as the angle between lemma and palea during flower opening on five representative plants
10	Spikelet opening duration (SOD) (min)	Measured as length of time between flower opening and closing (five top spikelets on 12 plants)
11	Tiller number (TNum)	Recorded as the average of total number of grain-bearing and non-bearing tillers on 12 representative plants
12	50% Heading date (HD) (days)	Number of days from effective seeding to 50% heading
13	Spikelets per panicle (SPKS)	Recorded as average of total number of spikelets per panicle on 12 representative plants
14	Grain filling (GRFLG) (%)	Number of well-developed spikelets as a percentage of the total number of spikelets on 12 representative plants
15	1000-grain weight (KGwt) (g)	Weight of 1000 grains
16	Grain yield (GYLD) (t ha^{-1})	Weight of grains ha^{-1} at 14% moisture content

Table 3.2 continued. Parameters measured for the agro-morphological characterisation of genotypes (Biodiversity International *et al*., 2007)

S/N	Qualitative traits	
17	Apiculus colour (Apicolor)	Visual observation: 1 = white; 2 = straw; 3 = brown (tawny); 4 = green; 5 = red; 6 = red apex; 7 = purple; 8 = purple apex; 9 = black
18	Awn distribution (Awndist)	The presence and distribution of awns along the panicle: 0 = none (awnless); 1 = tip only; 2 = upper quarter only; 3 = upper half only; 4 = upper three-quarters only; 5 = whole length
19	Basal leaf sheath colour (Blsc)	Visual observation of the colour of the outer surface of the basal leaf sheath: 1 = green; 2 = green with purple lines; 3 = light purple; 4 = purple
20	Caryopsis colour (Cryopcolor)	Visual observation of the colour of the pericarp: 1 = white; 2 = light brown; 3 = speckled brown; 4 = brown; 5 = red; 6 = variable purple; 7 = purple
21	Culm habit (Chbt)	Visual observation of angle of inclination of the base of the main culm from vertical: 1 = erect; 3 = semi-erect; 5 = open; 7 = spreading; 9 = procumbent
22	Culm lodging resistance Clrst)	Scored at maturity based on the observed degree of lodging: 1 = very weak; 3 = weak; 5 = intermediate; 7 = strong; 9 = very strong
23	Flag leaf attitude (Flgat)	Visual observation of angle of attachment between the flag leaf blade and the main panicle axis: 1 = erect; 3 = semi-erect; 5 = horizontal; 7 = descending
24	Leaf blade pubescence (Lbp)	Visual observation and by touch: 1 = glabrous; 2 = intermediate 3 = pubescent

Table 3.2 continued. Parameters measured for the agro-morphological characterisation of genotypes (Biodiversity International *et al*., 2007)

S/N	Qualitative traits	
25	Panicle shattering (Pansht)	Visual observation of the extent to which grains have shattered from the panicle: 1 = very low (<3%); 3 = low (~3%); 5 = moderate (~15%); 7 = high (~35%); 9 = very high (>50%)
26	Panicle type (Pntype)	Visual observation of compactness of panicle: 1 = erect (compact); 3 = semi-erect (semi-compact); 5 = spreading (open); 7 = horizontal; 9 = drooping

S/N = serial number

Table 3.3 Shannon-Weaver index (H') for qualitative traits

Trait	H'	H' *O. sativa*	H' Japonica	H' Indica	H' interspecific
Apicolor	0.31	0.32	0.29	0.32	0.23
Awndist	0.19	0.15	0.00	0.30	0.23
Blsc	0.14	0.22	0.20	0.30	0.17
Cryopcolor	0.16	0.22	0.11	0.34	0.36
Chbt	0.29	0.30	0.28	0.33	0.34
Clrst	0.26	0.34	0.33	0.34	0.23
Flgat	0.24	0.30	0.32	0.30	0.35
Lbp	0.20	0.27	0.26	0.30	0.00
Mean	0.23 ± 0.02	0.27 ± 0.02	0.22 ± 0.04	0.32 ± 0.01	0.24 ± 0.04

Apicolor = apiculus colour; Awndist = distribution of awns; Blsc = basal leaf sheath colour; Cryopcolor = caryopsis pericarp colour; Chbt = culm habit; Clrst = culm lodging resistance; Flgat = flag leaf attitude Lbp = leaf blade pubescence

Awn distribution

Awn distribution revealed diversity of H' = 0.19. The majority (81%) of the accessions were awnless (Figure 3.1C) including accessions of the Japonica group which were largely homogenous for the trait and had no awns. Only WAB384 had awns at the tip of the panicle only. Likewise, improved varieties of the Indica group had no awns. However, awns were on the tip of two landraces, TOS 15729 and TOS 8076.

Basal leaf sheath colour

Basal leaf sheath colour had the lowest diversity based on qualitative traits with a value of H' = 0.14. Only two colours (green and purple) were observed for basal leaf sheath colour; of which 81% of accessions were green and 19% were purple (Figure 3.1C). Accessions of the *O. sativa* group Japonica had green basal leaf sheath colour with only WAB384 and WAB519 having a purple basal leaf sheath. For *O. sativa* group Indica purple basal leaf sheath was observed on B6144F and TOS15505 whilst the rest had green basal leaf sheath. Interspecific hybrid accessions had green basal leaf sheath colour; only NERICA 2 had a purple basal sheath. Other accessions that had purple basal leaf sheath colour were PL87-3 and CG14 which belong to *O. glaberrima*.

Caryopsis colour

Caryopsis colour revealed a diversity value of H' = 0.16. Accessions were either of white or red caryopsis colour. The majority of accessions (72.2%) had white whilst only 27.8% had red caryopsis colour (Figure 3.1C). The latter were mostly accessions of interspecific hybrids and landraces of *O. sativa* group Indica. WAB96-1-1 was the only Japonica accession that had red caryopsis colour.

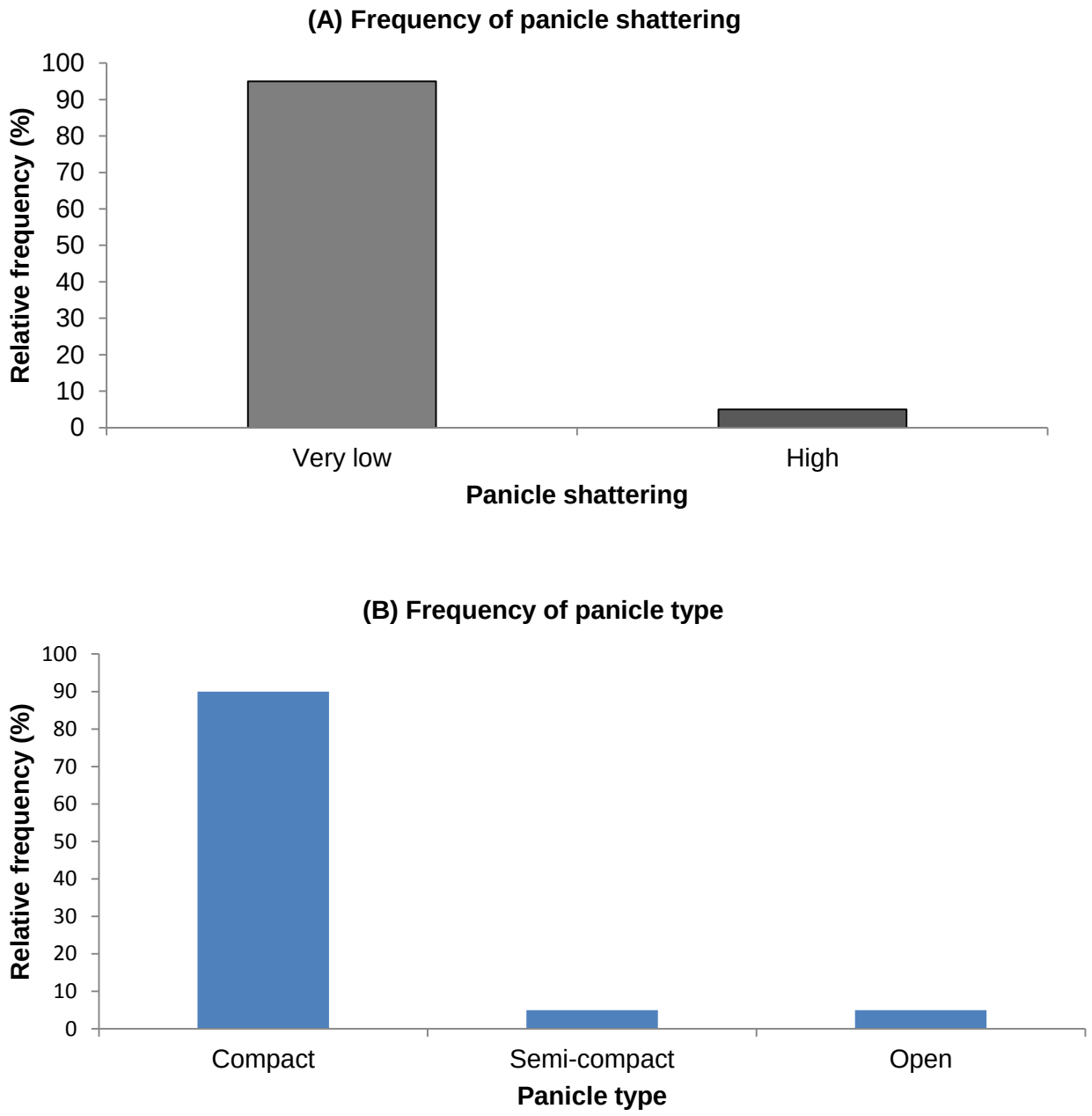


Figure 3.1A Frequency distribution of (A) Panicle shattering and (B) Panicle type

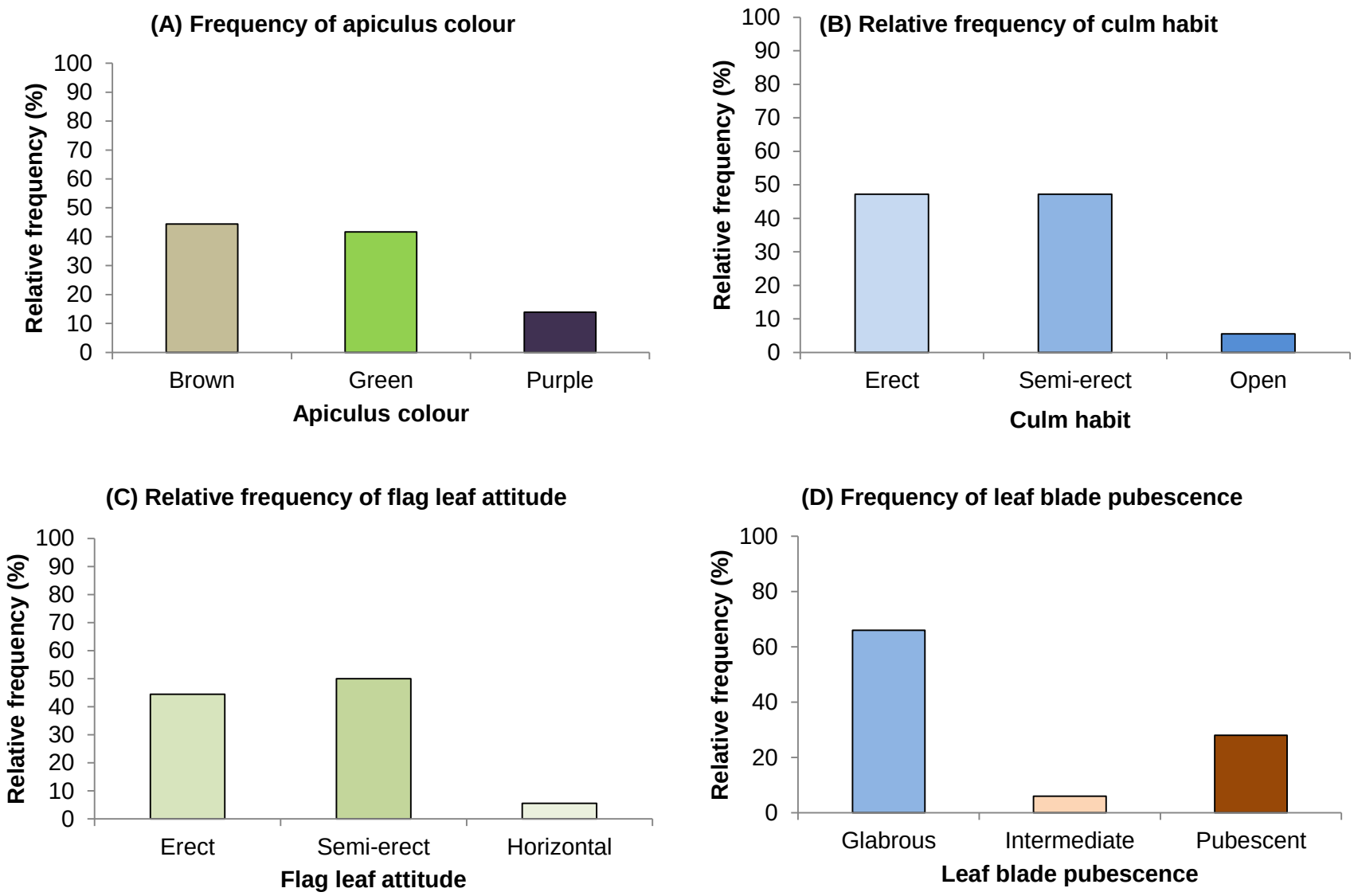


Figure 3.1B Frequency distribution of (A) Apiculus colour, (B) Culm habit, (C) Flag leaf attitude and (D) Leaf blade pubescence

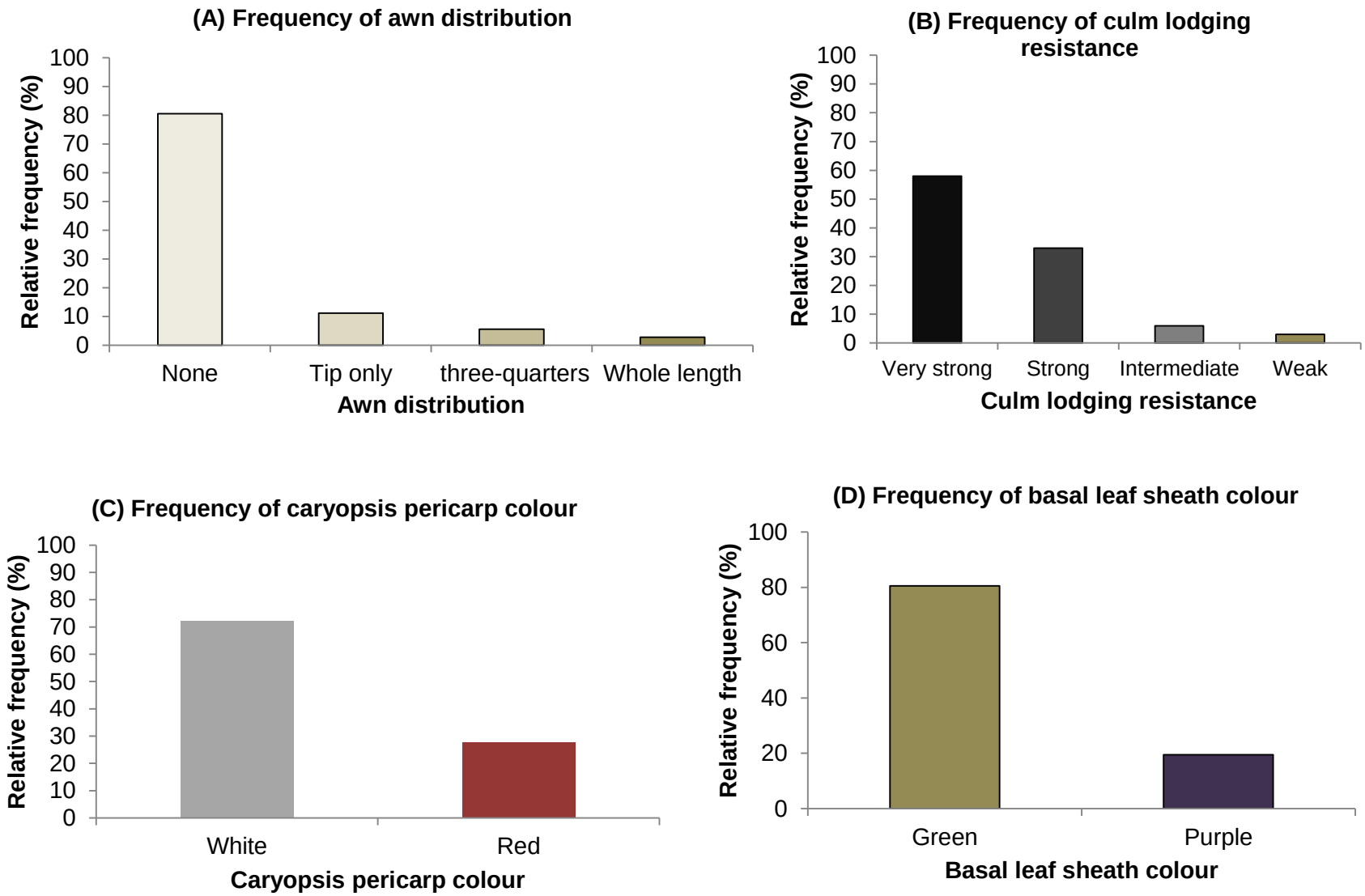


Figure 3.1C Frequency distribution of (A) Awn distribution, (B) Culm lodging resistance, (C) Caryopsis pericarp colour and (D) Basal leaf sheath colour

Culm habit

Culm habit revealed diversity of H' = 0.29. Three classes were observed for the trait; 47% of accessions were erect, another 47% were semi-erect whilst 6% were open (Figure 3.1B). Culm habit of *O. sativa* Japonica group accessions was either erect (56%) or semi-erect (44%). On the other hand, for interspecific hybrids, culm habit was 44% and 56% erect and semi-erect respectively. For *O. sativa* Indica group accessions the distribution was 43% erect and 57% semi-erect culm habit. CG14 and PL87-3 which are *O. glaberrima* accessions had an open culm habit. Erect and semi-erect culm habit are characteristics of improved cultivars whilst traditional cultivars and wild rice relatives often have an open culm habit.

Culm lodging resistance

Culm lodging resistance revealed diversity of H' = 0.26. Most of the accessions were strongly resistant to lodging; of these 58% were very strongly resistant and 33% were strongly resistant. In addition 6% were of intermediate resistance whilst 3% were weak (Figure 3.1C). All accessions of the *O. sativa* Japonica group and most of *O. sativa* Indica group were very strongly resistant to lodging. Interspecific hybrid accessions were strongly resistant however, NERICA 7 had intermediate resistance. CG14 and PL87-3 had weak lodging resistance and this is a characteristic of *O. glaberrima,* which includes both accessions.

Flag leaf attitude

Diversity level was H' = 0.24. Three classes were observed of which 50% of accessions were semi-erect, whilst 44% and 6% were erect and horizontal respectively (Figure 3.1B). Majority of *O. sativa* Japonica and Indica accessions were erect or semi-erect for flag leaf attitude; only WAB307 and WAB337 had open flag leaves. Accessions of interspecific hybrids also had erect or semi-erect flag leaves. CG14 and PL87-3 had semi-erect flag leaves.

Leaf blade pubescence

Leaf blade pubescence revealed a diversity level of H' = 0.20. Three classes were observed for leaf blade pubescence; of which 66% of accessions were glabrous, 28% were pubescent and 6% were intermediate (Figure 3.1B). Accessions of *O. sativa* Japonica group were mostly glabrous (67%) however, the following accessions: Afhikari, Moroberekan, WAB384, WAB519 and WAB96-1-1 have pubescent leaves. On the other hand, 71% of *O. sativa* Indica group accessions had pubescent leaves; only ITA150 and IRAT104 were observed to have glabrous leaves. Interspecific hybrids were monomorphic with 100% of accessions having glabrous leaves.

Panicle shattering

Panicle shattering revealed a diversity level of H' = 0.09. Most accessions (95%) were observed to have very low panicle shattering (Figure 3.1A). This conforms to expectations as most of the accessions are improved varieties and low panicle shattering is associated with improved varieties.

Panicle type

Panicle type revealed a diversity level of H' = 0.12. As a characteristic feature of improved varieties compact panicle type was observed in 90% of accessions (Figure 3.1A). Open panicle type was observed in the accessions CG14 and PL87-3. Semi-compact panicles were also observed in interspecific hybrid accessions.

3.5.2 Diversity as revealed by quantitative traits

Sixteen quantitative traits were evaluated in this study. The descriptive statistics indicating several parameters are shown in Table 3.4. The highest coefficient of variation was observed for tiller number per plant (34.7%). This trait values ranged from 4–13, but showed a high heritability of 0.86. The lowest coefficient of variation was for days to 50% heading (3.2%) which was followed by panicle exsertion (3.8%). The number of spikelets per panicle had the highest standard error followed by plant height. This reflects the broad range of differences of yield components and height of accessions.

Heritability scores ranged from 0.51 (leaf blade length) to 0.96 (1000-grain weight). The lowest heritability was followed by flag leaf length, spikelet opening duration and spikelet opening angle and grain yield, with heritability scores of 0.52, 0.55 0.58 and 0.58 respectively. The highest heritability value of 0.96 was followed in diminishing order, by panicle number per plant, panicle length and days to 50% heading with values of 0.93, 0.92 and 0.90 respectively.

3.5.2.1 Principal component analysis of quantitative traits

Results of the PCA also revealed diversity within the collection. Tables 3.5A and 3.5B give a summary for the first three components before and after Varimax rotation respectively. Cumulatively the first three principal components (PC1, PC2 and PC3) before Varimax rotation, accounted for 87.3% of the total variation. PC1 accounted for 44.6%, whilst PC2 and PC3 accounted for 24.6% and 18.1% of variation, respectively (Table 3.5A). After Varimax rotation the first three factor axes (D1, D2 and D3) similarly accounted for 87.3% of the total variation.

Table 3.4 Summary statistics of quantitative traits showing diversity of the collection

Trait	Min	Max	Mean	CV (%)	SE	Heritability
KGWt	23.2	41.1	31.2	5.1	0.9	0.96
FLL	24.4	42.6	31.4	12.9	2.3	0.52
FLW	1.5	2.06	1.7	10.2	0.1	0.66
GRFLG	10.6	82.9	67.0	10.1	3.9	0.77
GYLD	30.1	474.9	3.4	24.2	0.8	0.58
HD	58.3	106.8	77.7	3.2	1.4	0.90
LBL	32.9	54.3	43.9	10.3	2.6	0.51
LBW	1.1	1.8	1.4	10.7	0.1	0.82
PnEx	36.0	43.5	39.8	3.8	0.9	0.81
PnL	17.9	29.3	24.9	5.6	0.8	0.92
PnNum	3.6	12.1	5.8	19.3	0.7	0.93
PHt	72.4	136.7	111.3	7.7	4.9	0.85
SOA	25.7	30.8	28.8	6.9	1.1	0.58
SOD	43.1	70.7	55.0	11.8	3.7	0.55
SPKS	65.5	197.9	119.8	14.8	10.2	0.82
TNum	3.9	13.3	7.0	34.7	1.4	0.86

Min = minimum; Max = maximum; CV = coefficient of variation; SE = standard error; KGWt = 1000-grain weight; FLL = flag leaf length; FLW = flag leaf width; GRNFLG = % grain filling; GYLD = grain yield; HD = days to 50% heading; LBL = leaf blade length; LBW = leaf blade width; PnEx = panicle exsertion; PnL = panicle length; PnNum = panicle number; PHt = plant height; SOA = spikelet opening angle; SOD = spikelet opening duration; SPKS = spikelets per panicle; TNum = tiller number

Table 3.5A Summary results of principal components before Varimax rotation

	PC1	PC2	PC3
Eigenvalue	3.6	2.0	1.4
Variability (%)	44.6	24.6	18.1
Cumulative (%)	44.6	69.2	87.3

PC1 = first principal component; PC2 = second principal component; PC3 = third principal component

Table 3.5B Summary results of factor axes after Varimax rotation

	D1	D2	D3
Variability (%)	36.0	27.5	23.8
Cumulative (%)	36.0	63.5	87.3

D1 = first principal component after Varimax rotation; D2 = second principal component after Varimax rotation; D3 = third principal component after Varimax rotation

However, this time D1 accounted for 36.0%, whilst D2 and D3 accounted for 27.5% and 23.8% of the variation respectively (Table 3.5B).

Factor loadings of retained traits for rotation for the first three principal components (D1, D2 and D3) are given in Table 3.6. Traits that did not significantly contribute to variation were excluded from Table 3.6. Traits that contributed most to D1 were flag leaf length and flag leaf width with factor loadings of 0.963 each. These were followed by leaf blade length with a loading of 0.781. The highest loading trait for D2 was 1000-grain weight and leaf blade width both with a factor loading value of 0.971. Percentage filled grains per panicle and grain yield were the highest loading traits for the third component D3. They scored 0.938 and 0.915 respectively (Table 3.6).

Table 3.6 Factor loadings for the first three principal components after Varimax rotation

	D1	D2	D3
KGWt	0.083	0.971	0.165
FLL	0.963	0.060	0.055
FLW	0.963	0.060	0.055
GRNFLG	0.008	0.205	0.938
LBL	0.781	0.129	0.293
LBW	0.083	0.971	0.165
PnL	0.614	0.479	-0.197
GYLD	0.163	0.120	0.915

KGWt = 1000-grain weight; FLL = flag leaf length; FLW = flag leaf width; GRNFLG = % grain filling; LBL = leaf blade length; LBW = leaf blade width; PnL = panicle length; GYLD = grain yield; D1 = first principal component after Varimax rotation; D2 = second principal component after Varimax rotation; D3 = third principal component after Varimax rotation

The correlation circle map of the PCA is shown in Figure 3.2. The horizontal axis is linked to the three components and their related traits. Flag leaf length, flag leaf width and leaf blade length were linked to the first component. These three traits were strongly correlated. Traits of the second component were 1000-grain weight and leaf blade width. Grain filling percentage and grain yield linked to the third component were also distinctly projected and shown to be well correlated.

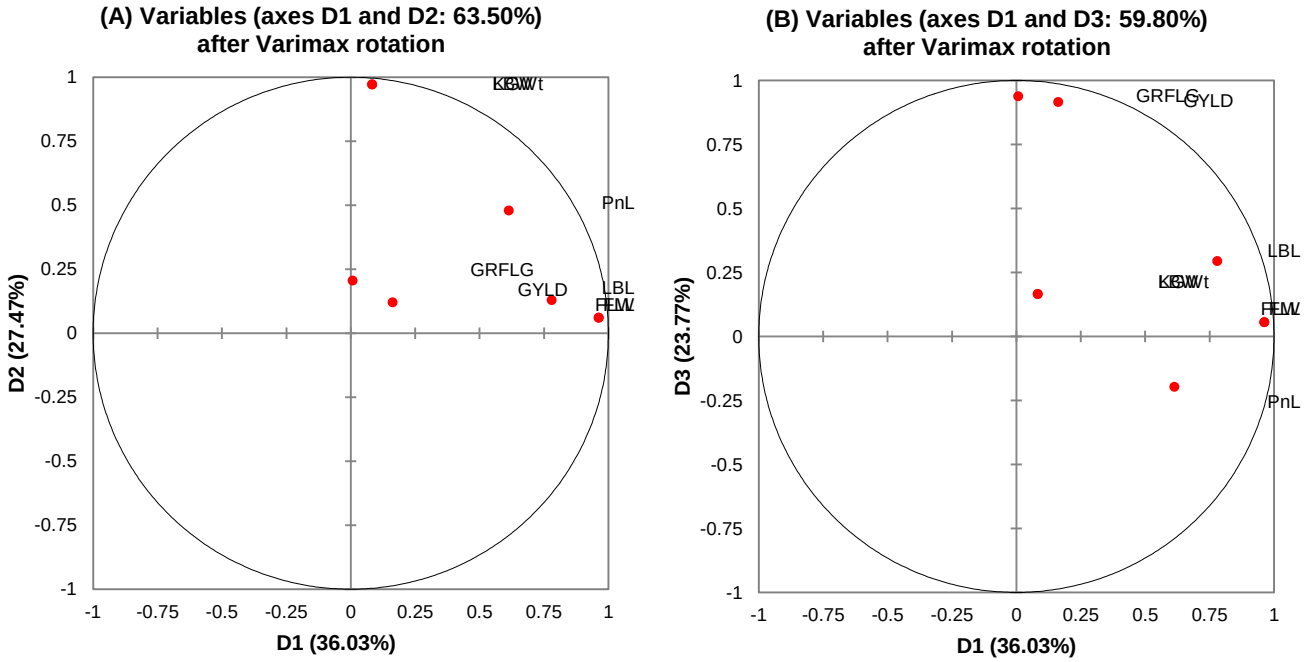


Figure 3.2 The correlation circle projecting selected variables (A) D1/D2 and (B) D1/D3

KGWt = 1000-grain weight; FLL = flag leaf length; FLW = flag leaf width; GRFLG = % grain filling; LBW = leaf blade length; LBW = leaf blade width; PnL = panicle length; GYLD = grain yield

The PC biplot which projected the different accessions is shown in Figure 3.3. It is apparent from Figure 3.3A and 3.3B that accessions Afhikari, ITA123, CG14 and PL87-3 are unique. Afhikari is an Asian genotype which belongs to the *O. sativa* Japonica group and it is not quite adapted to Africa. CG14 and PL 87-3 are *O. glaberrima* landraces and PL87-3 particularly, is adapted to the lowland ecology. All other accessions are adapted to the upland ecology.

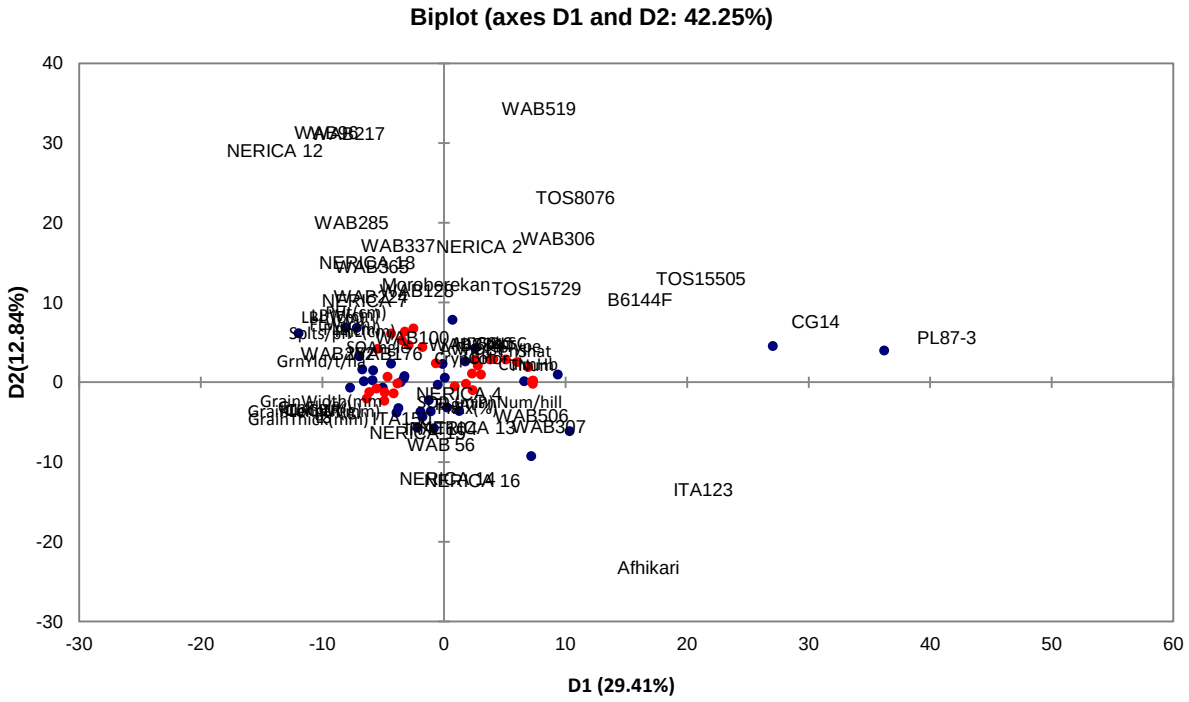


Figure 3.3A Biplot projecting the different accessions on axes D1/D2

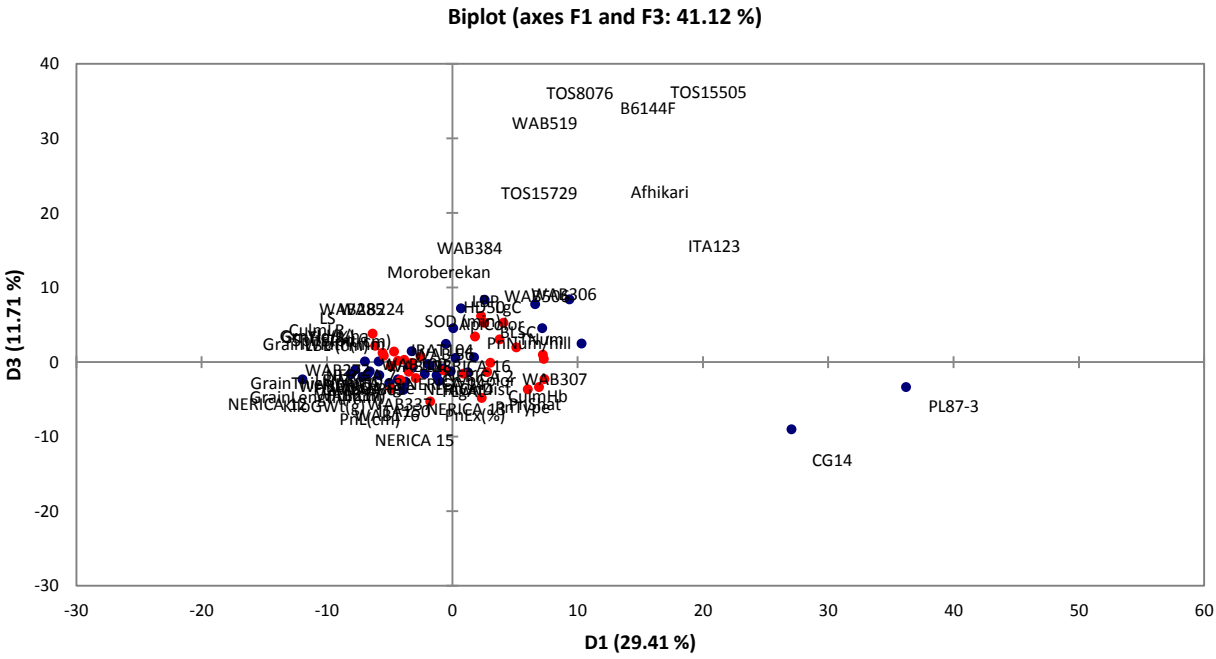


Figure 3.3B Biplot projecting the different accessions on axes D1/D3

3.5.2.2 Clustering of quantitative traits

The dendrogram projecting the clustering of accessions based on 16 quantitative agro-morphological traits is presented in Figure 3.4. There were two main groups A and B which were subdivided into several clusters. Group B had one cluster C2 which contained PL87-3 and CG14, both belonging to the species *O. glaberrima.* Group A subdivided into two subgroups A1 and A2. Subgroup A1 contained two clusters C1 and C4. Cluster C1 contained two accessions; Afhikari and ITA123. These belong to the species *O. sativa*, but different variety groups. Afhikari belongs to *O. sativa* Japonica group of Asian origin whilst ITA123 belongs to *O. sativa* Indica group. Cluster C4 contained mainly *O. sativa* Indica group accessions and was divided into two sub-clusters.

The first sub-cluster contained three accessions namely: TOS8076, TOS15729 and WAB519. TOS8076 and TOS15729 which clustered together, and belong to the *O sativa* Indica group. WAB519 which belongs to *O. sativa* Japonica group was an outlier within the sub-cluster. The second sub-cluster contained two accessions namely TOS15505 and B6144F, both of them belonging to *O. sativa* Indica group.

Subgroup A2 contained one cluster (Cluster C3). Cluster C3 was the largest cluster, containing 27 accessions. It contained the majority of *O. sativa* Japonica group accessions. The only Japonica group accessions outside of C3 were Afhikari and WAB519. Cluster C3 also contained all interspecific hybrid accessions and it was divided into several sub-clusters. The first sub-cluster contained WAB176 and NERICA 4 which clustered together and then IRAT105 and WAB56-104 which also clustered together. NERICA 16 was an outlier in this sub-cluster. The sub-cluster also contained WAB365 and WAB100, which clustered together. NERICA 15 and ITA150 clustered together with NERICA 14 as an outlier. NERICA 2 and WAB384 were also in this sub-cluster. The second sub-cluster contained WAB224, WAB128, WAB337, Moroberekan, NERICA 13, WAB307, WAB506 and WAB306. The third sub-cluster contained WAB96, WAB217 and NERICA 12 clustering together and then NERICA 7, WAB272, WAB285 and NERICA 18 also clustering together.

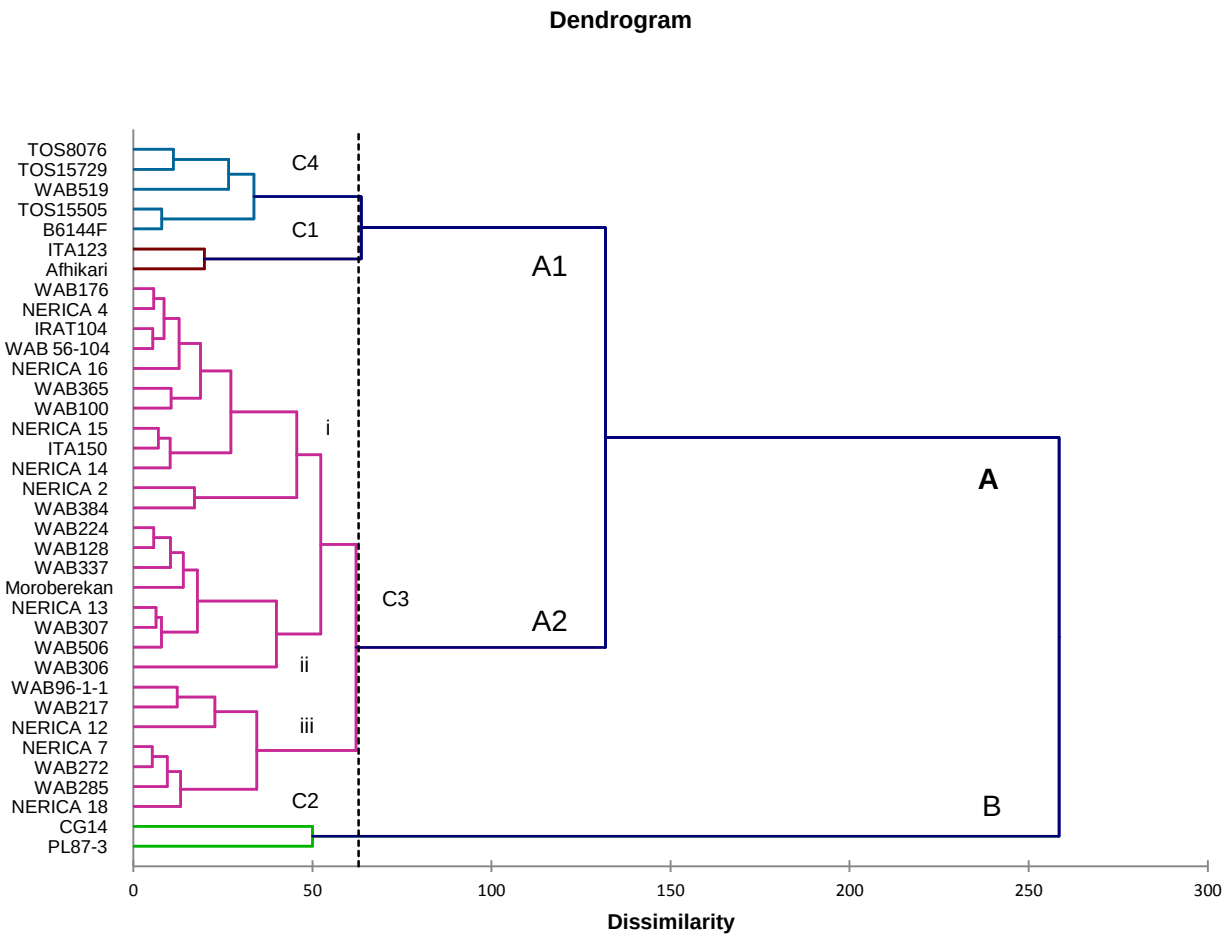


Figure 3.4 Dendrogram showing clusters among the 36 rice accessions based on 16 quantitative traits

Distances between clusters are shown in Table 3.7 and cluster variance is shown in Table 3.8. Cluster C1 was the most homogenous whilst C4 was the least homogenous (Figure 3.4). This is confirmed by the within-class variance shown in Table 3.8. C1 had a within-class variance of 605.10 whilst C4 had a value 1980.36.

Table 3.7 Distances between clusters

Cluster	1	2	3	4
1	0			
2	41.10	0		
3	60.53	70.49	0	
4	65.42	72.03	25.14	0

Table 3.8 Variance results by cluster

Cluster	C1	C2	C3	C4
Objects	2	2	27	5
Sum of weights	2	2	27	5
Within-cluster variance	605.10	1556.83	853.14	1980.36
Average distance to centroid	17.40	27.90	57.67	73.85

Cluster C2 contained accessions that were characterised by the presence of anthocyanin in the leaves and culms, therefore displaying purple colour of the basal leaf sheath. The colour of the apiculus was also purple. They exhibited profuse vegetative growth with a high number of tillers, but culms were susceptible to lodging. Accessions had open panicles that were well exserted but with a high shattering capacity. The spikelet opening duration was about 60 min at a moderately wide angle during anthesis. Accessions in the cluster had short grains that are awned with a red caryopsis colour. The presence of awns as well as lodging and panicle shattering are important traits in rice domestication (Vaughan *et al*., 2008; Ishii *et al*., 2013). Awn presence is associated with wild rice, where it is believed to deter birds and other animals and profuse vegetative growth is a favourable trait for weed competitiveness. Accessions in this cluster could be used as donors for increasing tiller number and profuse growth, which are key features for yield increase.

Cluster C1 contained accessions that are characterised by the lack of anthocyanin in the leaves and culms, therefore displaying green colour of leaves, culms and the basal leaf sheath. Culms were also strongly resistant to lodging. They had compact panicles that were moderately well exserted with low shattering capacity. Spikelet opening duration lasted 50–60 min at a moderately wide angle. Spikelet fertility was fairly low, but accessions gave medium yields. Accessions had awnless grains of medium size with a white caryopsis.

Cluster C3 contained accessions that largely consisted of improved varieties. The only *O. sativa* Japonica group accessions that were not in this cluster were Afhikari and WAB519. Similarly all interspecific hybrids were found in this cluster. Moroberekan, which is a landrace, is the only non-improved variety found in cluster C3. Culms were strongly resistant to lodging. Accessions had erect and compact panicles that were moderately well exserted with low shattering capacity. The spikelet opening duration was for about 60 min at a moderately wide angle during anthesis. Accessions in cluster C3 were high yielding with long and heavy grains that were awnless. Improved varieties tend to have no awns to ease harvest and post-harvest operations. The colour of the caryopsis was white for the majority of accessions in this cluster. They were early maturing with good yields. Early maturity is desirable for the upland ecology where drought is a constraint. It is an escape mechanism which is effective in conditions of terminal drought (Fischer and Fukai, 2003).

Cluster C4 contained the largest number of landraces. However, WAB519 which is an improved variety belonging to the *O. sativa* Japonica group, was also found in cluster C4. Accessions typically had anthocyanin in the leaves and culms, therefore displaying purple colour of the basal leaf sheath. The colour of the apiculus was also purple. Culms were strongly resistant to lodging. Accessions had compact panicles that were well exserted with low shattering capacity. Spikelet opening duration was from 50–60 min at a moderately wide angle. Spikelet fertility was fairly good and accessions gave high yields. Grains were long and had no awns and the pericarp colour was red. Only WAB519 had a white pericarp in the cluster.

3.6 Discussion

Agro-morphological characterisation has been widely used to identify suitable traits to be used as genetic markers in rice outcrossing trials (Da Silva *et al*., 2005; Endo *et al*., 2009; Somaratne *et al*., 2012). They have also been equally effective in evaluating diversity in germplasm collections of rice. Ogunbayo *et al.* (2005), Bajracharya *et al*. (2006), Sanni *et al.* (2008), Babaei *et al.* (2011), Wu *et al*. (2011), Willocquet *et al.* (2012), Dong *et al.* (2013), and Sow *et al.* (2014),) have all used agro-morphological traits to assess diversity in rice germplasm collections. Similarly the 26 agro-morphological traits evaluated in this study effectively revealed variation within the 36 rice accessions.

Frequency distributions revealed that most of the improved varieties of *O. sativa* had brown or green apiculus colour whilst landraces had a purple apiculus colour. Accessions PL 87-3 and CG14 belonging to *O. glaberrima* also had a purple apiculus colour. Though the sample size of

O. glaberrima in the study is too small in number to be representative, this is in conformity with Nuijten *et al.* (2009) who reported similar results in their study of *O. glaberrima* germplasm in West Africa. Accessions CG14 and PL87-3 also had a high tiller number, but they were susceptible to lodging and panicle shattering. This is in line with Jones *et al.* (1997) who reported *O. glaberrima* having profuse vegetative growth and susceptibility to lodging and grain shattering.

Results from standardised Shannon-Weaver diversity index (H') indicated a low diversity value (0.23 ± 0.02) for the collection based on qualitative traits. The qualitative trait with the highest diversity was apiculus colour (0.31) and the least diversified was the colour of the basal leaf sheath. The low level of diversity is expected as most accessions are improved genotypes targeting the same ecology, which is the upland ecology. Accessions must have similar adaptive traits for performance in the upland. Parents are also likely to have been selected from the same gene pool.

Selected accessions of *O. glaberrima* showed lower diversity (H' = 0.18 ± 0.07) than the *O. sativa* accessions which had a value of H' = 0.27 ± 0.02 and the interspecific hybrids with H' = 0.24 ± 0.04. The number of *O. glaberrima* accessions in this study was too small to make a valid comparison. However, results of the study conform to reports in other studies. Semon *et al.* (2005) reported low levels of diversity in their studies of *O. glaberrima.* In the present study, accessions of *O. sativa* Indica group showed higher diversity. This could be due to the presence of landraces in this group. Landraces have high levels of diversity due to the ecological heterogeneity and climatic variations and biotic stresses to which they are adapted. This is in conformity with the study of Sow *et al*. (2014), who investigated agro-morphological variability of rice and reported that accessions in the irrigated ecologies were less diverse than those in traditional systems. Improved varieties predominate in irrigated ecologies whilst landraces predominate in traditional systems.

PCA of the quantitative traits showed that several traits contributed to variation in the collection. The first three components accounted for 87.3% of the total variation. Traits related to photosynthesis like flag leaf length, flag leaf width and leaf blade length, were most correlated with PC1. Panicle length also correlated with PC1. Traits associated with grain quality, like 1000-grain weight, correlated with PC2 whilst traits of yield and yield components correlated with PC3. Other rice morphological studies agree partially with results in this study. Fogliatto *et al.* (2011) reported that seed weight contributed significantly to variation. Moukoumbi *et al.* (2011) in their study of interspecific hybrid rice reported that traits that contributed most to variation were tillering

ability, plant height, maturity and spikelet fertility. Similarly, Sanni *et al.* (2008) reported that panicle length, grain size (including weight), tillering and days to heading contributed most to variation. Ogunbayo *et al.* (2005) reported that tillering, spikelet fertility and heading significantly influenced grain yield. Rabara *et al.* (2014) reported that culm number and panicle number highly contributed to diversity.

The biplot of the PCA using quantitative trait data did not reveal a clear grouping pattern of accessions. However, Afhikari, ITA123, CG14 and PL 87-3 appeared to be unique genotypes. Afhikari is an Asian genotype that is not particularly adapted to Africa, whilst PL 87-3 and CG14 are landraces. PL 87-3 is an *O. glaberrima* genotype that is adapted to the lowland ecology. The correlation analysis indicated strong correlations between eight quantitative variables evaluated in the study. Flag leaf length and flag leaf width were correlated as were 1000-grain weight and leaf blade width. Grain yield was also correlated with the percentage grain filling. The flag leaf plays an important role in photosynthesis (Yue *et al*., 2006). Breeding programmes could aim at increasing flag leaf length or flag leaf width during selection. Increased yields could also be achieved by selecting for increased leaf blade width. Above ground vegetative tissue like leaves are essential for photosynthesis. Therefore breeding for increase in functional flag leaves and leaf blades would eventually lead to increased yields. It is also evident that increased grain filling would result in increased yield. The classification of accessions into distinct groups with specific traits would facilitate identification of genotypes with dominant traits that will be used in the outcrossing trial.

3.7 Conclusions

The study successfully used agro-morphological characterisation to assess diversity within the germplasm collection and reveal distinct groups of accessions with specific traits that will be used for the outcrossing trial. It revealed the pattern and level of diversity. Several important traits were found to be strongly correlated. Thus future studies could use one rather than two highly correlated traits. These findings have implications for the exploitation of these genotypes. For instance, accessions in cluster C2 displayed the capacity to be weed competitive. They could be used as parents for improving weed competitiveness of other genotypes. Likewise accessions in other clusters could be used to improve accessions in cluster C2. Findings of this study could help reduce the amount of time and resources allocated to breeding in screening these accessions for potential parents for a given ideotype. In addition, findings assigned accessions into distinct agro-

morphological groups that would be useful for selecting donor and recipient germplasm for outcrossing trials using agro-morphological traits as markers.

3.8 Recommendation

This study revealed differences in diversity between the 36 accessions that belonged to *O. glaberrima*, *O. sativa* and interspecific hybrids. It identified groups of accessions with specific traits that will be used as donor and recipient germplasm for trials on outcrossing under field conditions. However, *O. glaberrima* accessions were few; therefore it was difficult to make comparisons involving *O. glaberrima* and the other species. This underrepresentation could be due to the study focussing on released and adopted varieties of which there are few *O. glaberrima* accessions. A future study should be more representative in terms of number of accessions for each species.

Agro-morphological traits that are well suited for evaluating germplasm are few; therefore they cover only a small proportion of the genome. They are also characterised by epistasis, pleiotropy and dominant-recessive relationships, further limiting their value as genetic markers (Smith and Smith, 1992). Molecular markers are not limited by the aforementioned negative characteristics. A future study should also use molecular markers to assess the diversity of the collection.

3.9 References

Addinsoft. 2008. XLSTAT Version 3.02. http://www.xlstat.com/en/. (Accessed 18/07/2015)

AGROBASE. 2013. Version 36.8.1, Agronomix, Winnipeg, Canada.

Allard RW. 1960. Principles of Plant Breeding. John Wiley and Sons, Inc. New York, 485 pp.

Babaei A, Nematzadeh GA, Hashemi H. 2011. An evaluation of genetic differentiation in rice mutants using semi-random markers and morphological characteristics. Australian Journal of Crop Science 5:1715–1722.

Bajracharya J, Steele KA, Jarvis DI, Sthapit BR, Witcombe JR. 2006. Rice landrace diversity in Nepal: variability of agro-morphological traits and SSR markers in landraces from a high-altitude site. Field Crops Research 95:327–335.

Barry M, Pham J, Courtois B, Billot C, Ahmadi N. 2007. Rice genetic diversity at farm and village levels and genetic structure of local varieties reveal need for *in situ* conservation. Genetic Resources and Crop Evolution 54:1675–1690.

Bioversity International, IRRI WARDA. 2007. Descriptors for wild and cultivated rice (*Oryza spp*.). Bioversity International, Rome, Italy; International Rice Research Institute, Los Baños, Philippines; WARDA, Africa Rice Center, Cotonou, Benin. 63 pp.

Cattell RB. 1966. The scree test for the number of factors. Multivariate Behavioral Research 1:245–276.

Chang TT. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica 25:425–441.

Chang TT. 1984. Conservation of rice genetic resources: luxury or necessity? Science 224:251–256.

Da Silva EF, Da Silva LM, Montalvan R. 2005. Crossing rate and distance in upland rice. Bragantia, Campinas 64:197–201.

Dong CF, Ding CL, Xu NX, Cheng YH, Shen YX, Gu HR. 2013. Double-purpose rice (*Oryza sativa* L.) variety selection and their morphological traits. Field Crops Research 149:276–282.

Endo T, Sato H, Yamaguchi M, Kataoka T, Nakagomi K, Ito T, Mori K. 2009. Estimate of outcrossing rates in a rice plant (*Oryza sativa* L.) under field conditions using a purple grain rice cultivar, *Okunomurasaki*. Breeding Science 59:195–202.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Fischer KS, Fukai S. 2003. How rice responds to drought. In: Fischer KS, Lafitte R, Fukai S, Atlin G, Hardy B. (Eds.). Breeding Rice for Drought-Prone Environments. International Rice Research Institute (IRRI). pp 32–36.

Fogliatto S, Vidotto F, Ferrero A. 2011. Morphological characterization of Italian weedy rice (*Oryza sativa*) populations. Weed Research 52:60–69.

Gao L. 2003. The conservation of Chinese rice biodiversity: genetic erosion, ethnobotany and prospects. Genetic Resources and Crop Evolution 50:17–32.

Hammer K, Teklu Y. 2008. Plant genetic resources: selected issues from genetic erosion to genetic engineering. Journal of Agriculture and Rural Development in the Tropics and Subtropics 109:15–50.

Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, Htun TM, Yamasaki M, Komeda N, Matsumoto T, Terauchi R, Ishikawa R, Ashikari M. 2013. *OsLG1* regulates a closed panicle trait in domesticated rice. Nature Genetics 45:462–465.

Jones MP, Dingkuhn M, Aluko GK, Semon M. 1997. Interspecific *Oryza sativa* L *x O. glaberrima* Steud. progenies in upland rice improvement. Euphytica 92:237–246.

Kaiser HF. 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement 20:141–151.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25–34.

Lee J, Park JH, Koh HJ. 2013. Morphological and genetic characterization of off-type rice plants collected from farm fields in Korea. Journal of Plant Biology 56:160–167.

Li ZM, Zheng XM, Ge S. 2011. Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. Theoretical and Applied Genetics 123:21–31.

Mazid MS, Rafii MY, Hanafi MM, Rahim HA, Shabanimofrad M, Latif MA. 2013. Agro-morphological characterization and assessment of variability, heritability, genetic advance and divergence in bacterial blight resistant rice genotypes. South African Journal of Botany 86:15–22.

Moukoumbi YD, Sie M, Vodouhe R, N'dri B, Toulou B, Ogunbayo SA, Ahanchede A. 2011. Assessing phenotypic diversity of interspecific rice varieties using agro-morphological characterization. Journal of Plant Breeding and Crop Science 3:74–86.

Nuijten E, Almekinders DCJM. 2008. Mechanisms explaining variety naming by farmers and name consistency of rice varieties in The Gambia. Economic Botany 62:148–160.

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P. 2009. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers' fields. PLoS ONE 4:1–9.

Ogunbayo SA, Ojo DK, Guei RG, Oyelakin OO, Sanni KA. 2005. Phylogenetic diversity and relationships among 40 rice accessions using morphological and RAPDs techniques. African Journal of Biotechnology 4:1234–1244.

Perry MC, McIntosh MS. 1991. Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. Crop Science 31:1350–1355.

Phan PDT, Kageyama H, Ishikawa R, Ishii T. 2012. Estimation of the outcrossing rate for annual Asian wild rice under field conditions. Breeding Science 62:256–262.

Portères R. 1970. Primary cradles of agriculture in the African continent. In: Fage JD, Oliver RA. (Eds.). Papers in African Prehistory. Cambridge: Cambridge University Press. pp 43–58.

Rabara RC, Ferrer MC, Diaz CL, Newingham MCV, Romero GO. 2014. Phenotypic diversity of farmers' traditional rice varieties in the Philippines. Agronomy 4:217–241.

Ruiz M, Varela F, Carillo JM. 1997. Analysis of the discriminating power of agro/morphological and biochemical descriptors in a sample of the Spanish collection of barely (*Hordeum vulgare* L.). Genetic Resources and Crop Evolution 44:247–255.

Sanni KA, Fawole I, Guei RG, Ojo DK, Somado EA, Tia DD, Ogunbayo SA, Sanchez I. 2008. Geographical patterns of phenotypic diversity in *Oryza sativa* landraces of Cote d'Ivoire. Euphytica 160:389–400.

Semon M, Nielsen R, Jones MP, McCouch SR. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O*. *sativa* and ecological adaptation. Genetics 169:1639–1647.

Shannon CE, Weaver W. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, 125 pp.

Shinada H, Iwata N, Sato T, Fujino K. 2013. Genetical and morphological characterization of cold tolerance at fertilization stage in rice. Breeding Science 63:197–204.

Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. Advances in Agronomy 47:85–141.

Somaratne LHMYK, Abayawickrama ASMT, Wickramasinghe IP, Samarasinghe WLG. 2012. Estimating out-crossing rate of Bg 379-2 using morphological markers and confirmation by molecular markers. Rice Science 19:166–168.

Song ZP, Lu BR, Zhu YG, Chen JK. 2003. Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657–665.

Sork VL, Nason J, Campbell DR, Fernandez JF. 1999. Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology and Evolution 14:219–224.

Sow M, Sido A, Laing M, Ndjiondjop MN. 2014. Agro-morphological variability of rice species collected from Niger. Plant Genetic Resources 12:22–34.

Vaughan DA, Lu BR, Tomooka N. 2008. The evolving story of rice evolution. Plant Science 174:394–408.

Ward JH. 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58:236–244.

Willocquet L, Noel M, Sackville HR, Savary S. 2012. Susceptibility of rice to sheath blight: an assessment of the diversity of rice germplasm according to genetic groups and morphological traits. Euphytica 183:227–241.

Wu LL, Liu ZL, Wang JM, Zhou CY, Chen KM. 2011. Morphological, anatomical and physiological characteristics involved in development of the large culm trait in rice. Australian Journal of Crop Science 5:1356–1363.

Yi G, Choi JH, Jeong EG, Chon NS, Jena KK, Ku YC, Kim DH, Eun MY, Jeon JS, Nam MH. 2005. Morphological and molecular characterization of a new frizzy panicle mutant, ‘‘*fzp-9*(t)’’, in rice (*Oryza sativa* L.). Hereditas 142:92–97.

Yue B, Xue WY, Luo LJ, Xing YZ. 2006. QTL analysis for flag leaf characteristics and their relationships with yield and yield traits in rice. Acta Genetica Sinica 33:824–832.

Chapter 4

Assessing diversity of rice varieties using microsatellite markers

4.1 Abstract

Knowledge of diversity of plant germplasm is a prerequisite and vital for selection of suitable parental material for outcrossing trials. It is equally important for breeding programmes. Microsatellite markers are effective in assessing diversity of germplasm. In this study the diversity of a collection of 36 rice accessions comprising *O. sativa, O. glaberrima* and interspecific hybrids (*O. sativa* x *O. glaberrima*) was assessed using 27 microsatellite markers. Accessions were mainly improved varieties for the upland ecology. A total of 321 alleles were detected with an average of 11.9 alleles per locus and an average major allele frequency of 0.29 per locus. The average gene diversity value was 0.81 and polymorphism information content was 0.80 per locus. Diversity indices for interspecific hybrids were intermediate between *O. sativa* and *O. glaberrima*, but closer to *O. sativa* Japonica group. Two populations were revealed which corresponded to the *O. sativa* Indica group and *O. sativa* Japonica group. Interspecific hybrid accessions were dispersed between the two groups. There were subtle subdivisions within each major population group. The sub-clustering in the first population was based on colour of the caryopsis. In the second population it was based on days to maturity. Results showed that modern improved cultivars have high levels of diversity that can be exploited for breeding and other plant studies. In addition, interspecific hybrids could be suitable as bridge cultivars for exchange of suitable traits between *O. sativa* and *O. glaberrima*. Interspecific hybrids can also be used to assess gene flow in rice.

Keywords: Diversity, interspecific rice, *O. glaberrima, O. sativa,* microsatellites

4.2 Introduction

Diversity of a population refers to the genetic makeup and phenotypic expression of the population. Often there are relationships between individuals in populations. Analysis of genetic relationships of individuals within and between populations is important for crop improvement. This is important for assessing genetic variability in cultivars, identifying parental combinations and introgressing desirable alleles into germplasm. In addition, such relationship can be used for studying outcrossing in rice. The identification of homogenous groups of individuals is important in plant science. Analysis of genetic diversity of accessions in a population facilitates

classification. Groups of similar individuals could be classified into one population and be treated alike (Mohammadi and Prasanna, 2003; Evanno *et al.,* 2005).

Knowledge of genetic diversity and the distribution pattern of variation within and amongst related crop varieties are useful for assessing outcrossing in rice, the conservation and management of germplasm, seed production and crop improvement. Molecular markers are useful tools to portray the structure and assess the genetic variability within and amongst different species (Garris *et al*., 2005; Semon *et al*., 2005; Barry *et al*., 2007; Kumar *et al*., 2012).

Molecular markers reflect DNA sequences that are readily detected and whose inheritance can be easily monitored. Their use is based on naturally occurring DNA polymorphisms. As molecular markers reflect DNA sequences, they are particularly suited for evaluating relationships between genotypes. They represent genetic variation, permitting the estimation of relationships between different genotypes. There are several molecular marker types that have been developed and used to characterise rice populations. Classification of molecular markers is based on the strategy of the technique. Markers that involve the use of the PCR are referred to as PCR-based markers as opposed to those that do not involve the use of PCR which are referred to as non-PCR based. The latter involve pioneering methods in molecular marker techniques. RFLP is a prominent non-PCR-based technique (Nathans and Smith, 1975; Semagn *et al*., 2006; Edwards *et al*., 2008; Chen *et al*., 2014).

Molecular markers are particularly suitable for studies on fingerprinting, genetic diversity, qualitative gene tagging and quantitative trait loci (QTL) mapping. They have also been effectively used in studies of plant genetics and evolution (Rick *et al*., 1994; Kumagai *et al*., 2010; Wang *et al*., 2014). They have been widely used in rice outcrossing trials (Nuijten *et al*., 2009; Somaratne *et al*., 2012). PCR-based approaches are techniques that were developed with the advent of PCR, which was first reported by Kleppe *et al*. (1971) but generally credited to Kary Mullis (Rabinow, 1996). PCR is a robust and reliable technique for amplifying DNA. It involves three steps namely, denaturing (94–98ºC), annealing (45–65ºC) and elongation (72ºC). These steps are repeated 25–50 times and the amplified DNA is separated by gel electrophoresis (Dale *et al*., 2012).

Microsatellites, also referred to as SSRs, are a prominent PCR-based approach. Microsatellites are tandem repeats of sequence units generally less than five bp in length. They are produced as a result of errors in DNA replication. Errors arise whilst DNA polymerase, in copying repeats, change the number of repeats in the affected region. Polymorphisms are therefore reflected based on the number of repeat units in a defined region of the genome being investigated.

Microsatellite markers are particularly attractive to work with. They are codominant markers that are highly variable and also reproducible. They are locus specific, displaying simple Mendelian inheritance and are randomly dispersed throughout the genome. Microsatellites are easy to work with and can be analysed by PCR and are easily detected on polyacrylamide or agarose gels. In addition, only small amounts of DNA samples (50~100 ng per individual) are required for assays (Akkaya *et al.,* 1995; Brown *et al.,* 1996; Agarwal *et al.,* 2008).

Microsatellites are a powerful tool in the assessment of outcrossing in rice and in the elucidation of genetic relationships within and amongst rice collections (Semon *et al*., 2005; Barry *et al*., 2007; Semagn *et al.,* 2007; Ndjiondjop *et al.,* 2010; Kumar *et al.,* 2012). In this study with the aim of identifying suitable donor and recipient plants, the genetic diversity of 36 rice accessions was evaluated using 27 microsatellite markers. Molecular markers are not affected by neutral variation and are capable of detecting minor differences between genotypes. Microsatellites were used in this study as a tool to further clarify diversity that might have not been detected using morphological characterisation.

4.3 Materials and methods

Thirty six accessions comprising *O. glaberrima*, *O. sativa* and interspecific (*O. sativa* x *O. glaberrima*) hybrids were utilised in the study (Table 3.1). The *O. sativa* material consisted of Indica and Japonica variety groups. All plant material was collected from the gene bank of the Africa Rice Center, Cotonou, Benin. Laboratory analysis was conducted at the Africa Rice Center Biotechnology Laboratory Cotonou, Benin.

4.3.1 DNA isolation

Approximately 250 mg of leaf tissue from fresh young leaves of two-week old seedlings were used for DNA extraction, using the Cetyl Trimethyl Ammonium Bromide (CTAB) protocol (Risterrucci *et al*., 2000).

A volume of 750 µl CTAB buffer [100 mM tris hydroxymethyl aminomethane, pH 8.0; 20 mM EDTA (ethylene-diaminetetra acetate), pH 8.0; 1.4 M NaCl; 2% (w/v) CTAB; 0.2% (v/v) β-Mercapthoethanol] was added to approximately 250 mg leaf tissue in a 1.5 ml microfuge tube and incubated in a water bath at 65°C for 1 h. The suspension was extracted with 500 µl chloroform: isoamylalcohol [24:1 (v/v)] and the phases separated by centrifugation at 12000 rpm for 5 min. DNA was precipitated from the aqueous phase with 0.66 volumes isopropanol at room temperature for 20 min and centrifuged at 12000 rpm for 10 min. The precipitate was washed at

room temperature with 500 ml 70% (v/v) ethanol for 20 min followed by centrifugation at 12000 rpm for 5 min. The pellet was air-dried for 1 h and resuspended in TE buffer (10 mM tris hydroxymethyl aminomethane, pH 8.0; 1 mM EDTA, pH 8.0). Resuspended DNA was precipitated with 0.75 M ammonium acetate and an equal volume chloroform: isoamylalcohol [24:1 (v/v)]. DNA was precipitated from the aqueous layer with two volumes of ice-cold absolute ethanol. After an overnight incubation at -20°C, DNA was recovered by centrifugation at 12000 *g* for 15 min and washed twice with ice-cold 70% (v/v) ethanol for 5 min. The pellet was air-dried and resuspended in TE buffer and treated with 0.1 mg ml^{-1} DNase-free RNase for 2 h at 37°C. The concentration of the DNA was determined using the NanoDrop 3300 spectrophotometer (Applied Biosystems), by loading 1 µl of each sample and measuring absorbance at A_{260} and A_{280}.

To check the quality of DNA, 3 µl of each sample was loaded on 0.3 µg ml^{-1} ethidium bromide stained 1% (w/v) agarose gels and electrophoresed at 100 V for 1 h and visualised under UV light. All samples were normalised to a final concentration of 15 ng $µl^{-1}$ by adding variable volumes of double distilled water to a final volume of 100 µl.

4.3.2 PCR amplification and genotyping

Twenty seven SSR markers (Appendix 3) selected from Gramene were utilised for the study (http://archive.gramene.org/markers/). Amplifications were carried out in a Biometra Tprofessional® Basic Gradient Thermal Cycler for DNA amplification. The PCR programme was as described by Folkertsma *et al*. (2005). Initial denaturation at 94°C for 15 min followed by 10 cycles of 94°C for 15 s, annealing for 20 s at touchdown temperatures declining from 61 to 51°C and extension at 72°C for 30 s; this was followed by 35 cycles of 94°C for 10 s, 54°C for 20 s and 72°C for 30 s; and a final extension step of 20 min at 72°C. Reliability of amplification for each primer set was confirmed by resolving 3 µl of each of the PCR products on a 2% (w/v) agarose gel at 100 V for 55 min. PCR products were subjected to electrophoresis on a 2% (w/v) agarose gel at 100 V for 55 min. DNA bands were observed under UV light. Allele sizes were determined using AlphaImager HP for Windows software package (2007), which encodes genes in base pairs.

4.4 Data analysis

Allele number per locus (Na), major allele frequency (MAF), gene diversity (D) and polymorphic information content (PIC) were computed using the software PowerMarker V3.0 (Liu and Muse, 2005). Variation in populations can be estimated by either heterozygosity or gene diversity. The

latter has been reported to be appropriate for inbred populations (Weir, 1996). Gene diversity was used to estimate variation in the collection of accessions. Gene diversity, often referred to as expected heterozygosity, is defined as the probability that two randomly chosen alleles from the population are different. An unbiased estimator of gene diversity at the *l*th allele is

$$\widehat{D}_l = (1 - \sum_{u=1}^{k} \tilde{P}_{lu}^2) \Big/ (1 - \frac{1+f}{n})$$

where n is the number of individuals; P_{lu} the frequency of the *l*th allele and *f* the inbreeding coefficient (Weir, 1996).

The PIC was calculated as

$$\widehat{PIC} = 1(-\sum_{i=1}^{n} P_i^2 - 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P_i^2 P_j^2$$

where P_i and P_j are the frequencies of the *i* and *j* alleles respectively (Botstein *et al*., 1980).

The data was run through a Bayesian clustering analysis to determine the underlying population structure using STRUCTURE software, version 2.3 (Pritchard *et al*., 2000; Hubisz *et al*., 2009). Most parameters were set to default values as recommended in the manual of STRUCTURE 2.0 (Pritchard and Wen, 2003) and those used by Evanno *et al*. (2005). An admixture model with assumption of correlated allele frequencies and unlinked loci was used (Falush *et al*., 2003). Posterior probabilities were estimated using a burn-in period of 50000 iterations and 500000 iterations of the Markov Chain Monte Carlo (MCMC). The number of populations tested (K) varied from one to seven with 10 repetitions each. Five independent simulations were performed. The best K value was estimated using the *ad hoc* statistic ΔK as suggested by Evanno *et al*. (2005). Posterior probabilities were again estimated for best K using a burn-in period of 90000 iterations and 900000 iterations of the MCMC.

Estimation of ΔK is based on the second-order rate of change of likelihood of data between consecutive K values as:

$$\Delta K = \frac{m|L''(K)|}{Stdev\,[L(K)]}$$

$|L'' (K)| = |L'(K + 1) - L'(K)|$; which corresponds to the second order rate of change of L(K) with respect to K. L´ (K) = L(K) – L(K - 1); which corresponds to the rate of change of likelihood function with respect to K. L(K) = average value of Ln P(D). Where Ln P(D) = posterior probability of the data for a given K.

To investigate genetic relationships between accessions, a genetic dissimilarity matrix was computed in DARwin software 5.0.155 (Perrier and Jacquemoud-Collet, 2006). The dissimilarity between samples was calculated by using the simple matching coefficient based on the Sokal and Michener (1958) index. The dissimilarity coefficient formula is:

$$d_{ij} = 1 - \frac{1}{L}\sum_{l=1}^{L}\frac{m_l}{\pi}$$

where d*ij* represents dissimilarity between units *i* and *j*, L the number of loci, π ploidy and m_l number of matching alleles for locus *l*.

Dissimilarity indices were used to perform factorial coordinate analysis for graphical representations on Euclidean planes that conserve, at best, distances between units. Dissimilarity values were also employed for the construction of an unweighted neighbour-joining tree (Perrier *et al*., 2003). Levels for support for the nodes were evaluated using a bootstrap analysis with 500 replicates.

4.5 Results

4.5.1 Genetic diversity

A summary of genetic diversity results for the total number of accessions is presented in Table 4.1. The 27 microsatellite markers revealed a total of 334 alleles with an average of 12.37 alleles per locus. This ranged from four alleles for RM10018 to 22 alleles for RM6842.

Table 4.1 Genetic diversity of 36 rice accessions genotyped using 27 microsatellite markers

Marker	Chr	Total (N = 36)			
		Na	MAF	D	PIC
RM11	7	19.00	0.14	0.93	0.92
RM19	12	14.00	0.22	0.89	0.88
RM60	3	12.00	0.22	0.87	0.86
RM85	3	14.00	0.19	0.89	0.88
RM167	11	12.00	0.22	0.86	0.85
RM222	10	17.00	0.25	0.89	0.88
RM248	7	5.00	0.44	0.66	0.60
RM264	8	19.00	0.17	0.91	0.91
RM286	11	9.00	0.57	0.61	0.56
RM542	7	11.00	0.44	0.75	0.73
RM1227	12	11.00	0.19	0.87	0.86
RM3341	1	8.00	0.36	0.72	0.68
RM3529	5	19.00	0.13	0.93	0.92
RM3907	9	11.00	0.33	0.81	0.79
RM5463	6	13.00	0.39	0.81	0.80
RM5590	11	19.00	0.18	0.91	0.91
RM5812	2	9.00	0.31	0.82	0.80
RM6314	4	11.00	0.18	0.87	0.86
RM6673	10	16.00	0.19	0.90	0.89
RM6840	1	13.00	0.22	0.89	0.88
RM6842	2	22.00	0.25	0.88	0.87
RM10018	1	4.00	0.81	0.34	0.32
RM15281	3	14.00	0.17	0.90	0.89
RM18452	5	13.00	0.14	0.91	0.91
RM23662	9	9.00	0.19	0.86	0.85
RM24035	9	5.00	0.44	0.71	0.66
RM27973	12	5.00	0.31	0.77	0.73
Total		334.00			
Mean		12.37	0.28	0.82	0.80

N = number of accessions; Chr = chromosome; Na = mean number of alleles per locus; MAF = major allele frequency; D = mean gene diversity per locus; PIC = polymorphic information content

The frequency of major alleles ranged from 0.13 (RM3529) to 0.81 (RM10018) with an average of 0.28 per locus. Gene diversity, on the other hand, was highest for RM3529 and RM11 with a value of 0.93 and lowest for RM10018 with a value of 0.34. The mean gene diversity value per locus was 0.82. The PIC had a mean value of 0.80 per locus; ranging from 0.32 for RM10018 to 0.92 for RM3529 and RM11. Markers were highly informative showing PIC values greater than 0.5 with only RM10018 showing a PIC value of less than 0.5. Results for heterozygosity were not

taken into account as gene diversity is more suitable in characterising variation in inbred populations (Weir, 1996).

A summary of genetic diversity results for the three species *O. glaberrima*, *O. sativa* and interspecific hybrids is shown in Table 4.2, Figure 4.1 and Figure 4.2. Samples of *O. sativa* revealed 272 alleles with an average of 10.07 alleles per locus. The allele number ranged from four for RM10018 and RM248 to 17 for RM5590. Major allele frequency values ranged from 0.12 for RM11 to 0.76 for RM10018 with a mean of 0.31. Gene diversity revealed an average of 0.80 and ranged from 0.40 for RM10018 to 0.92 for RM11. Similarly PIC values revealed an average of 0.78 and ranged from 0.37 for RM10018 to 0.92 for RM11. Markers were highly informative with all but RM10018 revealing PIC values of more than 0.5.

A total of 168 alleles and an average of 6.2 alleles per locus were detected for interspecific hybrid accessions. The allele number ranged from two for RM10018 to 11 for RM6842. Major allele frequency values ranged from 0.11 for RM19 and RM18452 to 0.89 for RM10018 with a mean of 0.33. Gene diversity revealed an average of 0.76 and ranged from 0.20 for RM10018 to 0.89 for RM18452 and RM19. Similarly PIC values for interspecific hybrids revealed an average of 0.73 and ranged from 0.18 for RM10018 to 0.88 for RM19 and RM18452. Markers were highly informative with all but RM10018 revealing PIC values of more than 0.5.

A total of 53 alleles were detected for the two accessions of *O. glaberrima*. The average number of alleles per locus was 1.96 and that for major allele frequency was 0.58. Average values for diversity and PIC were 0.43 and 0.33 respectively. Selected accessions of *O. sativa* had a higher level of diversity (Figure 4.2). Values for D and PIC were higher for *O. sativa* than for interspecific hybrids.

Table 4.2A Genetic diversity of *O. sativa* and interspecific rice

Marker	Chr	*O. sativa* (N = 25)				Interspecific (N = 9)			
		Na	MAF	D	PIC	Na	MAF	D	PIC
RM11	7	15.00	0.12	0.92	0.92	7.00	0.22	0.84	0.82
RM19	12	11.00	0.28	0.86	0.84	9.00	0.11	0.89	0.88
RM60	3	11.00	0.20	0.87	0.85	5.00	0.33	0.77	0.73
RM85	3	12.00	0.28	0.85	0.83	6.00	0.22	0.81	0.79
RM167	11	11.00	0.20	0.86	0.85	6.00	0.33	0.79	0.76
RM222	10	14.00	0.20	0.90	0.89	4.00	0.44	0.67	0.61
RM248	7	4.00	0.48	0.60	0.52	4.00	0.33	0.72	0.66
RM264	8	15.00	0.20	0.90	0.89	8.00	0.22	0.86	0.85
RM286	11	7.00	0.54	0.62	0.57	4.00	0.56	0.62	0.57
RM542	7	8.00	0.44	0.73	0.70	5.00	0.44	0.72	0.68
RM1227	12	11.00	0.20	0.88	0.87	5.00	0.22	0.79	0.76
RM3341	1	6.00	0.44	0.67	0.61	5.00	0.44	0.72	0.68
RM3529	5	12.00	0.18	0.89	0.88	8.00	0.22	0.85	0.83
RM3907	9	9.00	0.40	0.78	0.76	6.00	0.44	0.74	0.71
RM5463	6	9.00	0.44	0.76	0.74	7.00	0.33	0.81	0.79
RM5590	11	17.00	0.16	0.91	0.91	9.00	0.28	0.85	0.83
RM5812	2	7.00	0.28	0.80	0.77	5.00	0.44	0.72	0.68
RM6314	4	9.00	0.22	0.84	0.83	8.00	0.17	0.86	0.85
RM6673	10	13.00	0.20	0.89	0.88	7.00	0.22	0.84	0.82
RM6840	1	10.00	0.28	0.85	0.83	7.00	0.22	0.84	0.82
RM6842	2	15.00	0.26	0.87	0.85	11.00	0.22	0.88	0.86
RM10018	1	4.00	0.76	0.40	0.37	2.00	0.89	0.20	0.18
RM15281	3	10.00	0.20	0.87	0.86	8.00	0.22	0.86	0.85
RM18452	5	13.00	0.16	0.90	0.89	9.00	0.11	0.89	0.88
RM23662	9	9.00	0.24	0.86	0.84	6.00	0.33	0.79	0.76
RM24035	9	5.00	0.60	0.59	0.54	4.00	0.44	0.67	0.61
RM27973	12	5.00	0.40	0.73	0.68	3.00	0.44	0.64	0.57
Total		272.00				168.00			
Mean		10.07	0.31	0.80	0.78	6.22	0.33	0.76	0.73

N = number of accessions; Chr = chromosome; Na = mean number of alleles per locus; MAF = major allele frequency; D = mean gene diversity per locus; PIC = polymorphic information content

Table 4.2B Genetic diversity of *O. glaberrima*

Marker	**Chr**	***O. glaberrima* (N = 2)**			
		Na	MAF	D	PIC
RM11	7	2.00	0.50	0.50	0.38
RM19	12	2.00	0.50	0.50	0.38
RM60	3	2.00	0.50	0.50	0.38
RM85	3	2.00	0.50	0.50	0.38
RM167	11	2.00	0.50	0.50	0.38
RM222	10	2.00	0.50	0.50	0.38
RM248	7	2.00	0.50	0.50	0.38
RM264	8	2.00	0.50	0.50	0.38
RM286	11	1.00	1.00	0.00	0.00
RM542	7	2.00	0.50	0.50	0.38
RM1227	12	2.00	0.50	0.50	0.38
RM3341	1	2.00	0.50	0.50	0.38
RM3529	5	2.00	0.50	0.50	0.38
RM3907	9	2.00	0.50	0.50	0.38
RM5463	6	3.00	0.50	0.63	0.55
RM5590	11	1.00	1.00	0.00	0.00
RM5812	2	1.00	1.00	0.00	0.00
RM6314	4	3.00	0.50	0.63	0.55
RM6673	10	2.00	0.50	0.50	0.38
RM6840	1	2.00	0.50	0.50	0.38
RM6842	2	4.00	0.25	0.75	0.70
RM10018	1	1.00	1.00	0.00	0.00
RM15281	3	2.00	0.50	0.50	0.38
RM18452	5	1.00	1.00	0.00	0.00
RM23662	9	2.00	0.50	0.50	0.38
RM24035	9	2.00	0.50	0.50	0.38
RM27973	12	2.00	0.50	0.50	0.38
Total		53.00			
Mean		1.96	0.58	0.43	0.33

N = number of accessions; Chr = chromosome; Na = mean number of alleles per locus; MAF = major allele frequency; D = mean gene diversity rate per locus; PIC = polymorphic information content

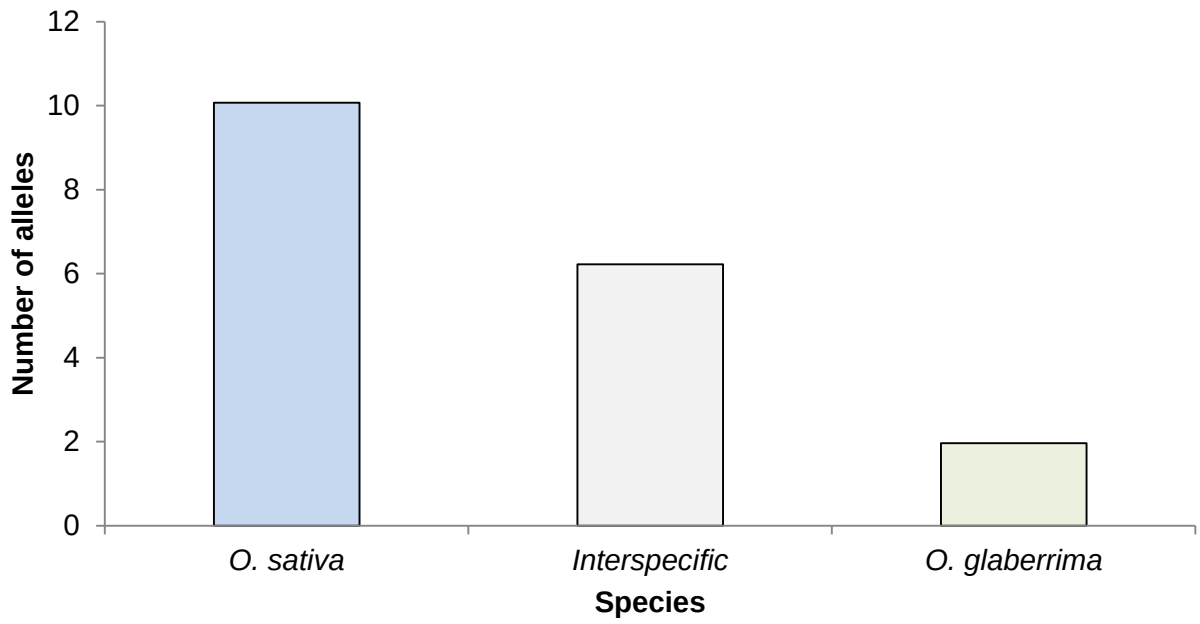


Figure 4.1 The average number of alleles per locus for *O. glaberrima*, *O. sativa* and interspecific hybrids

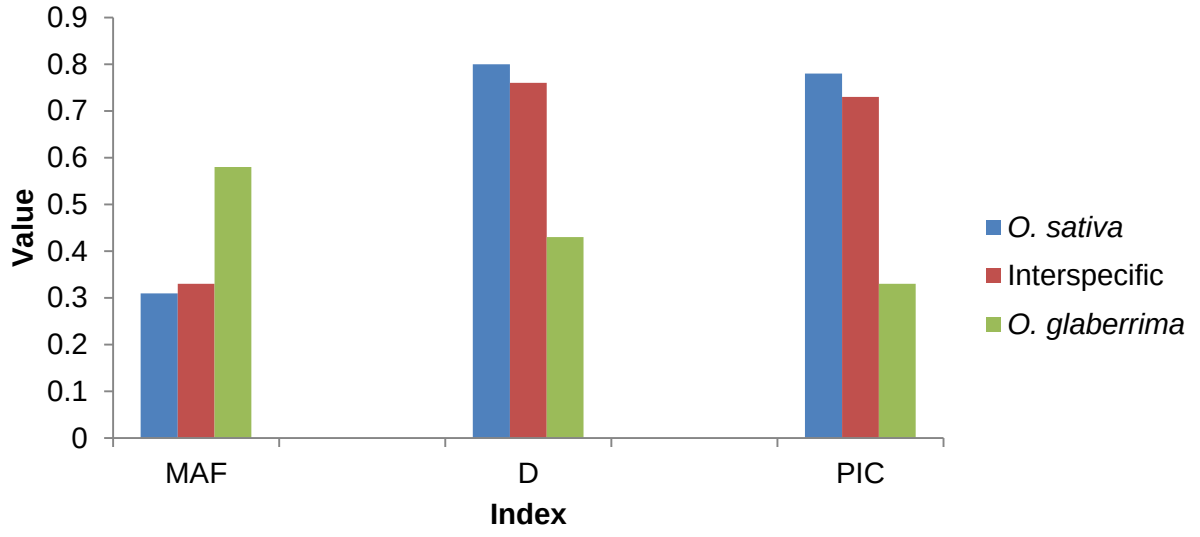


Figure 4.2 Average values per locus for different indices of diversity. MAF = major allele frequency; D = gene diversity; PIC = polymorphic information content

4.5.2 Structure of the diversity

The *ad hoc* statistic ΔK showed a value of best K = 2, indicating the presence of two clusters or populations (Figure 4.3 and Figure 4.4). Table 4.3 shows the inferred ancestry of accessions in the populations. All accessions derived high percentage (> 90%) of ancestry from one population or the other.

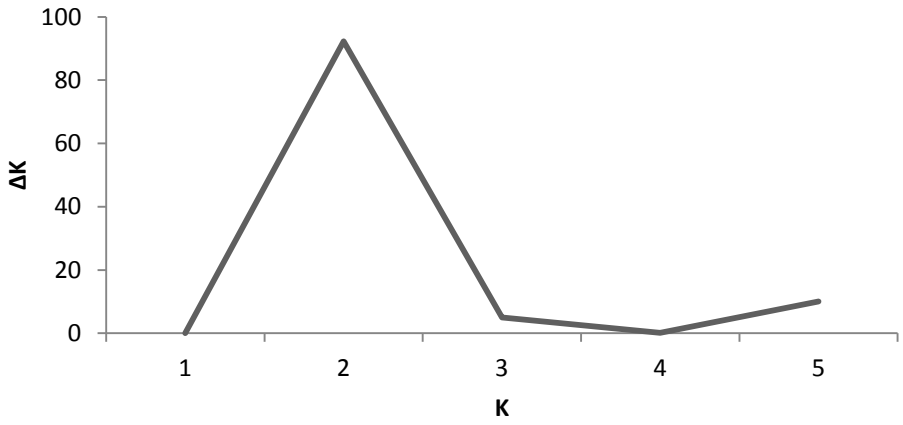


Figure 4.3 A plot of Evanno's *ad hoc* ΔK statistic against different possible values for K (the modal value indicates the most probable value of K = 2)

Figure 4.4 Model based ancestries of 36 rice accessions estimated from 27 nuclear SSR loci using STRUCTURE software (Pritchard *et al*., 2000)

Cluster C1 contained 64.4% of accessions while cluster C2 had 35.6% of accessions. Accessions in cluster C1 were typically *O. sativa* Japonica group accessions (Figure 4.5). However, four interspecific hybrid accessions (NERICA 4, NERICA 7, NERICA 12, NERICA 13) were assigned to cluster C1. In addition, ITA150 which belongs to *O. sativa* Indica group, and PL87-3 which belong to *O. glaberrima* were in cluster C1. Accessions in cluster C1 were of intermediate plant

height with early to medium maturity with glabrous leaves. They had wide and heavy grains of medium length with white colour of the caryopsis. Both ITA150 and PL87-3 had glabrous leaves.

Table 4.3 Inferred ancestry of accessions in the two populations (cluster C1 and cluster C2) using STRUCTURE software (Pritchard *et al*., 2000)

Accession	Cluster C1	Cluster C2
Afhikari	0.992	0.008
WAB56-104	0.999	0.001
Moroberekan	0.981	0.019
WAB100	0.998	0.002
WAB128	0.998	0.002
WAB176	0.999	0.001
WAB217	0.998	0.002
WAB224	0.999	0.001
WAB272	0.999	0.001
WAB285	0.997	0.003
WAB306	0.991	0.009
WAB307	0.999	0.001
WAB337	0.996	0.004
WAB365	0.998	0.002
WAB384	0.998	0.002
WAB506	0.998	0.002
WAB519	0.997	0.003
WAB96-1-1	0.002	0.998
ITA123	0.239	0.761
ITA150	0.997	0.003
IRAT104	0.022	0.978
B6144F	0.001	0.999
TOS15505	0.001	0.999
TOS15729	0.001	0.999
TOS8076	0.001	0.999
NERICA 4	0.997	0.003
NERICA 7	0.999	0.001
NERICA 12	0.998	0.002
NERICA 13	0.999	0.001
NERICA 2	0.003	0.997
NERICA 14	0.002	0.998
NERICA 15	0.002	0.998
NERICA 16	0.004	0.996
NERICA 18	0.001	0.999
PL87	0.992	0.008
CG14	0.001	0.999

Accessions in C2 were typically of *O. sativa* Indica group. However, five interspecific hybrid accessions (NERICA 2, NERICA 14, NERICA 15, NERICA 16 and NERICA 18) were clustered in cluster C2. In addition WAB96-1-1 which is an *O. sativa* Japonica group accession and CG14 which is an *O. glaberrima* accession were assigned to cluster C2. Accessions in cluster C2 were of intermediate plant height with medium to late maturity with pubescent leaves. They had wide and heavy grains of medium length with red colour of the caryopsis. It is interesting to note that CG14, WAB96-1-1, NERICA 14, NERICA 15, NERICA 16 and NERICA 18, though not belonging to *O. sativa* Indica group, showed a red colour of the caryopsis. WAB96-1-1 also had pubescent leaves. All landraces, namely TOS15505, TOS15729 and TOS8076 which are of the Indica group, were clustered in C2.

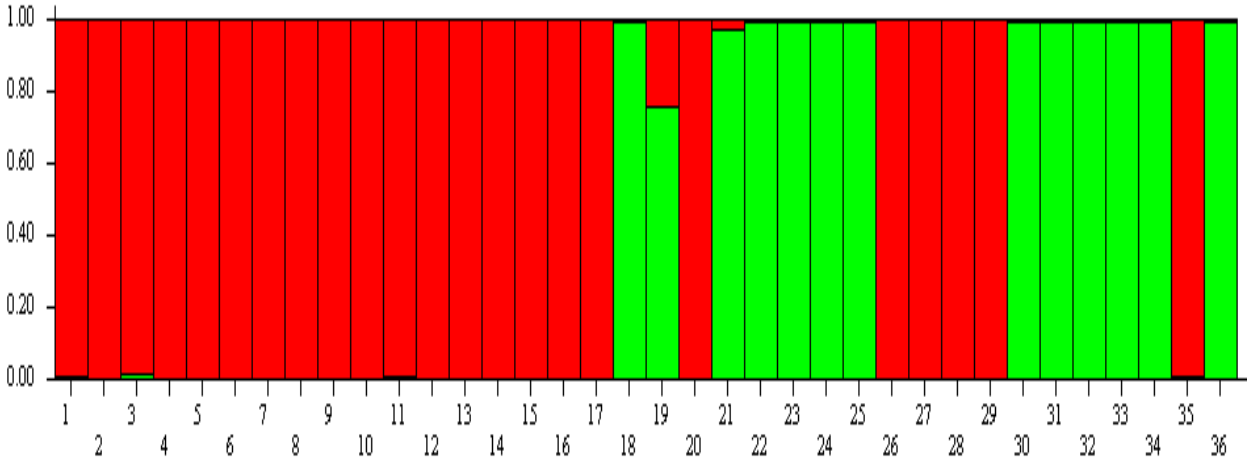


Figure 4.5 Estimated population structure at K = 2. Each accession is represented by a vertical coloured line, which is partitioned into coloured segments that represent the individual's membership fraction in K clusters using STRUCTURE software (Pritchard *et al*., 2000)

1 = Afhikari; 2 = WAB56-104; 3 = Moroberekan; 4 = WAB100; 5 = WAB128; 6 = WAB176; 7 = WAB217; 8 = WAB224; 9 = WAB272; 10 = WAB285; 11 = WAB306; 12 = WAB307; 13 = WAB337; 14 = WAB365; 15 = WAB384; 16 = WAB506; 17 = WAB519; 18 = WAB96-1-1; 19 = ITA123; 20 = ITA150; 21 = IRAT104; 22 = B6144F; 23 = TOS15505; 24 = TOS15729; 25 = TOS8076; 26 = NERICA 4; 27 = NERICA 7; 28 = NERICA 12; 29 = NERICA 13; 30 = NERICA 2; 31 = NERICA 14; 32 = NERICA 15; 33 = NERICA 16; 34 = NERICA 18; 35 = PL87-3; 36 = CG14

Figure 4.6 shows the result of factorial analysis. Accessions were divided into two major groups (Grp1 and Grp2). This largely agrees with the underlying population structure inferred by STRUCTURE.

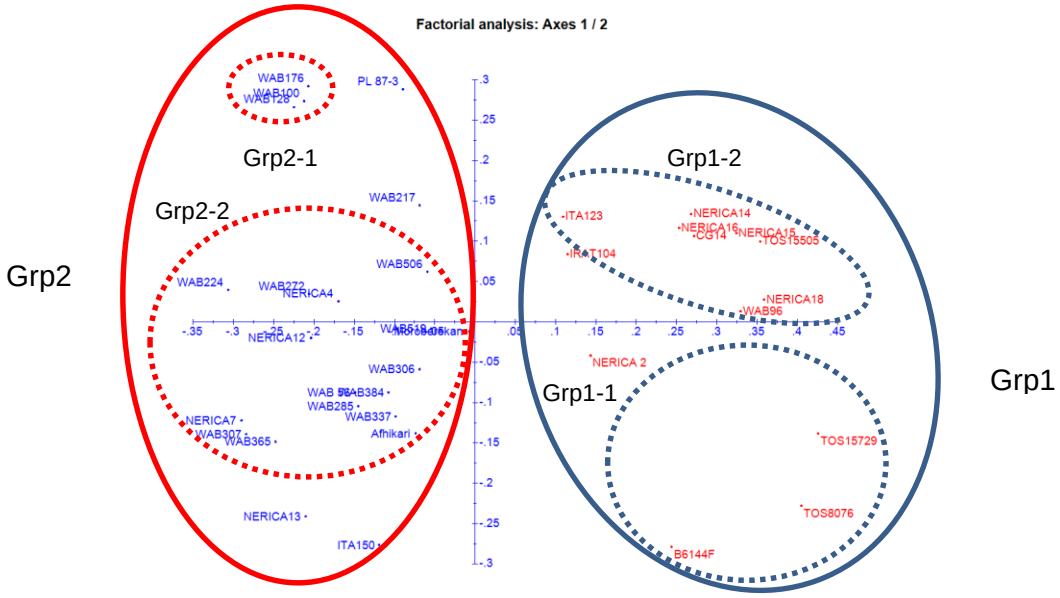


Figure 4.6 Factorial analysis of 36 rice accessions using 27 microsatellite markers

Grp = group

There were sub-groups within the two groups. The first group (Grp1), had two sub-groups: Grp1-1 and Grp1-2. Sub-group Grp1-1 comprised accessions of *O. sativa* Indica group as well as landraces. Sub-group Grp1-2 contained mainly accessions of *O. sativa* Indica group namely ITA123, IRAT104 and TOS15505. Accessions CG14 and WAB96-1-1 which were found in Grp1-2, belong to *O. glaberrima* and *O. sativa* Japonica group respectively. NERICA 14, NERICA 15, NERICA 16 and NERICA 18, also found in Grp1-2, are interspecific hybrids. The second group, Grp2, also had two sub-groups namely Grp2-1 and Grp2-2. Group Grp2 comprised mainly *O. sativa* Japonica group accessions. Early maturing accessions of intermediate plant height were clustered in Grp2-1, whilst accessions of medium maturity were assigned to Grp2-2. The interspecific hybrid accessions NERICA 4, NERICA 7, NERICA 12 and NERICA 13 were also assigned to Grp2-2. Figure 4.7 is a presentation of the tree that was generated based on dissimilarities between samples, calculated using the simple matching coefficient, using the unweighted neighbour joining method (UNJM).

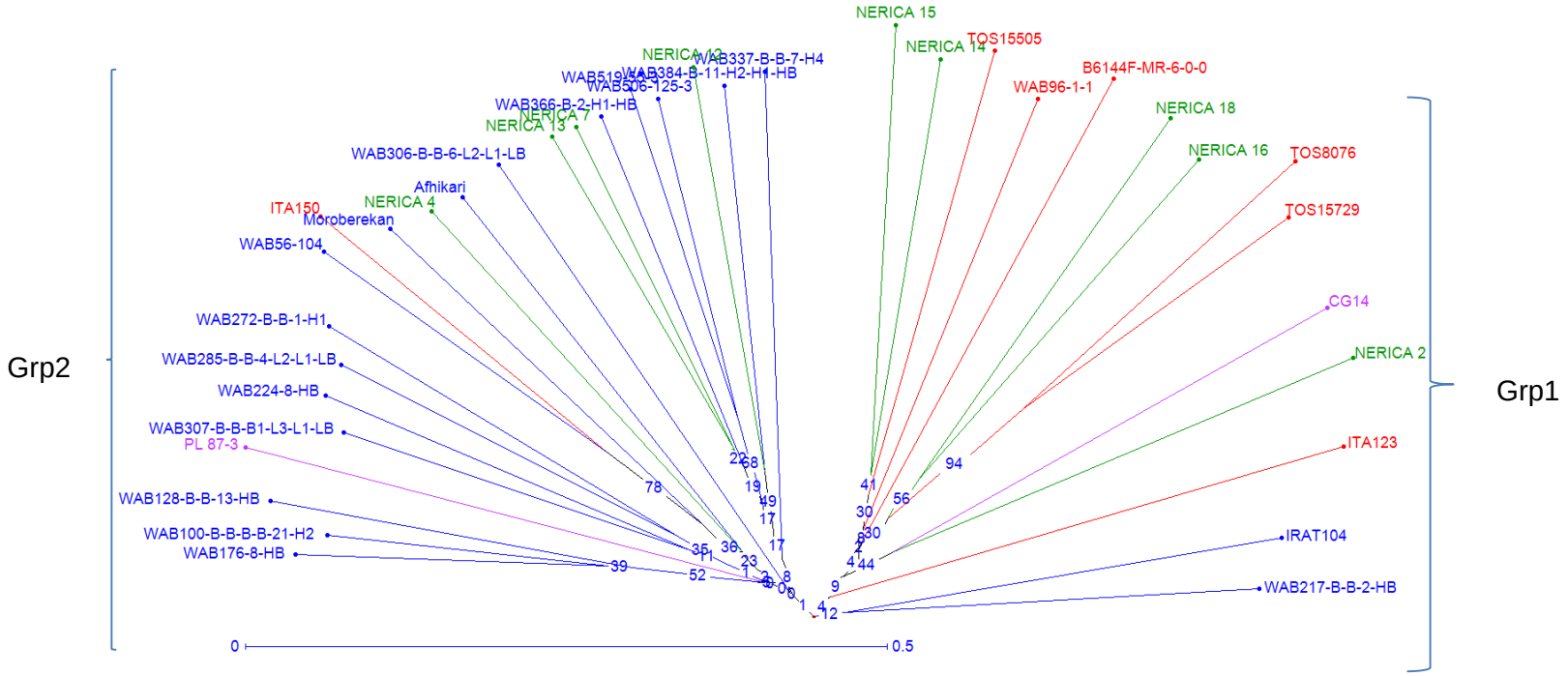


Figure 4.7 Unrooted tree of simple matching distances, based on Unweighted Neighbour Joining Method of 36 rice accessions at 27 microsatellite loci; numbers at the nodes indicate bootstrap values

Grp = group, red colour = *O. sativa* ssp. *indica*, blue colour = *O. sativa* ssp. *japonica*, purple = *O. glaberrima,* green colour = interspecific hybrid

The clustering using UNJM showed two groups, Grp1 and Grp2. This largely agreed with the inference of STRUCTURE with ΔK = 2 and factorial analyses (Figures 4.3, 4.4, 4.5 and 4.6). Grp1 contained 36% of the accessions which were mostly *O. sativa* Indica group and accessions of other species that showed red colour of the caryopsis. Within Grp1 accessions ITA123 and IRAT104, with white colour of the caryopsis, grouped together but were further apart from the other accessions that showed red colour of the caryopsis. Grp2 contained 64% of the accessions and they were mainly *O. sativa* Japonica group. Accessions of other species that clustered in this group were ITA150 and PL87-3 belonging to *O. sativa* Indica group and *O. glaberrima* respectively. In addition NERICA 4, NERICA 7, NERICA 12 and NERICA 13, belonging to interspecific rice, also clustered in Grp2 (Figure 4.7).

All Japonica accessions clustered together in group2, except for IRAT104, WAB217 and WAB96-1-1 which were found in Grp1. All Indica accessions clustered together in Grp1, except for ITA150 which was found in Grp2. Interspecific hybrid accessions were almost equally distributed between Grp 1 and Grp2. Grp1 had five interspecific hybrids (NERICA 2, NERICA 14, NERICA 15, NERICA 16 and NERICA 18). Grp2 had four interspecific hybrids (NERICA 4, NERICA 7, NERICA 12 and NERICA 13). Accessions CG14 and PL87-3 belonging to *O. glaberrima* were distributed in both groups and did not cluster together.

Grp1 and Grp2 contained accessions from all four classes (Indica, Japonica, *O glaberrima* and interspecific hybrids). All landraces but one clustered in Grp1. Only Moroberekan (landrace) clustered in Grp2. Grp 1 had a higher level of variation whereas Grp2 was more homogenous with the majority of its accessions belonging to the Japonica variety group.

Accessions TOS8076 and TOS15729 grouped very close, showing the highest bootstrap value in the dendrogram. The two are *O. sativa* landraces belonging to variety group Indica. Bootstrap values indicated in Figure 4.5 are generally low.

Classifications by PCA, STRUCTURE and dendrogram were similar in dividing all accessions into two major groups. In addition for all of them the two *O. glaberrima* accessions CG14 and PL87-3 always clustered in different groups.

4.6 Discussion

Assessing diversity of plant genotypes is important for crop improvement and increased agricultural production. It is equally important for identifying dominant traits that can be used in outcrossing studies. The relevance of conservation and maintenance of crop germplasm for such

a purpose has been known for a long time (Chang, 1984). There have been studies on assessment of diversity of rice germplasm using microsatellite markers in different regions worldwide (Garris *et al*., 2005; Qi *et al*., 2009; Choudhury *et al.,* 2013). In Africa, similar studies have been carried out (Semon *et al.,* 2005; Barry *et al.,* 2007; Girma *et al.,* 2010; Ndjiondjop *et al.,* 2010). However, these studies evaluated mainly landraces and *O. glaberrima*.

It is also equally important to evaluate the diversity of improved varieties adapted to African conditions. This is important for selecting parental material for conducting outcrossing trials and breeding programmes. The 27 microsatellite markers revealed a large number of alleles. Na values for the total collection had an average of 12.37 which ranged from 4–22 alleles per locus. These values compare favourably with the study of Qi *et al*. (2009) of modern varieties of *O. sativa* Japonica and Indica groups in China. In that study they reported a mean Na of 11.6 for the total collection with values ranging from 2–22. In another study by Barry *et al.* (2007), evaluating the diversity of 170 rice accessions from Guinea using 11 microsatellite markers, reported the average Na to be 12, ranging from 9–15. Their finding is similar to results from this study, where the average Na for total accessions was 12.4. Garris *et al*. (2005) reported a Na value of 11.8 for the total collection. The similar levels of diversity detected by this study in interspecific hybrids and *O. sativa* accessions is expected, as the interspecific accessions in this study were a result of a cross between *O. sativa* and *O. glaberrima*. The progeny was then backcrossed to a recurrent *O. sativa* parent (Jones *et al.,* 1997).

Results revealed by the 27 microsatellite markers for this collection of 36 accessions are in general similar to what has been reported for rice by other studies (Garris *et al*., 2005; Semon *et al.* 2005; Barry *et al.,* 2007; Qi *et al*., 2009). However, the current study had a higher proportion of Japonica than Indica accessions in the collection compared to previous studies. Differences could equally be due to differences in the number and type of accessions and the choice of microsatellite markers used in the different studies. For instance Barry *et al.* (2007) evaluated the diversity of landraces of *O. sativa* and *O. glaberrima* using 11 markers. Semon *et al*. (2005) evaluated only *O. glaberrima* accessions using 198 accessions and 93 microsatellite markers. Qi *et al*. (2009) evaluated 512 modern varieties of Indica and Japonica groups of *O. sativa* using 36 microsatellite markers. Garris *et al*. (2005) evaluated 234 accessions representing the geographic range of *O. sativa* using 169 microsatellite markers. However, this study focused on improved varieties adapted to the upland ecology in West Africa. Suitable genotypes for this ecology are mainly *O. sativa* Japonica group accessions. Comparing results of different studies could be difficult due to differences in plant material composition of different accessions. Differences in

population size and species composition could influence results. Studies have also shown that microsatellite diversity was influenced by genetic, ecological and edaphic factors (Li *et al.,* 2000; Baek *et al.,* 2003).

Accessions assigned by STRUCTURE analysis to population C1 were mainly *O. sativa* Indica group and other varieties that showed red colour of the caryopsis. Some of the interspecific rice hybrids were also clustered in C1. They were generally medium to late maturing accessions. Genotypes in C1 could be good donor material for outcrossing trials using red caryopsis colour trait as a morphological marker. They could also serve as parental material for breeding varieties with red kernels. This is a trait that is preferred by specific rice growing communities in Africa. Accessions in C2 were mainly *O. sativa* Japonica. These were improved varieties with early to medium maturity and were white seeded. Subtle subdivisions of the two populations were revealed by factorial and clustering analyses. Low bootstrap numbers observed on the dendrogram implies elevated levels of outcrossing between accessions (Felsenstein, 1982). Most accessions were developed from the same gene pool. Sub-groups were formed amongst genotypes that shared common parents like NERICA 16 and NERICA 18 in Grp1 and NERICA 12 with NERICA 13 in Grp2. Sub-groups were also formed based on common traits like colour of the caryopsis like NERICA 14 and NERICA 15. Accessions CG14 and PL87-3, both *O. glaberrima*, always clustered in different populations. This difference could have been due to differences in adapted ecology. CG14 is adapted to upland conditions whilst PL87-3 is adapted to lowland conditions. The genetic makeup of the two accessions is different. The accession WAB96-1-1 which is classified as a *O. sativa* Japonica group accession consistently clustered with *O. sativa* Indica group accessions in classifications by PCA, UNJM dendrogram and STRUCTURE. This could be a misclassification of the accession.

This study focused on genotypes adapted to the upland ecology in West Africa. The large population size of *O. sativa* Japonica accessions was due to this variety group being more adapted to the upland ecology in West Africa.

4.7 Conclusions

This study assessed the diversity of mainly improved mega varieties for the upland ecology for mainland West Africa. Improved cultivars are generally created from small gene pools. Notwithstanding, there are high levels of diversity within improved genotypes for the upland ecology. Therefore improved cultivars could have desirable alleles that could be exploited by rice and plant breeding programmes to serve as parental material for the creation of even more

superior varieties. On the other hand interspecific hybrids can be used as bridge cultivars for transfer of desirable traits from normally incompatible species.

Though the population structure of improved cultivars created from related parents might not be clear cut, there are still subtle subdivisions that define it. Molecular techniques using microsatellite markers are effective in assessing diversity of rice genotypes and can be used to genetically evaluate diversity and population structure of even closely related genotypes.

Accessions were divided into groups from which suitable material could be selected for studies on rice outcrossing.

4.8 Recommendation

Diversity is exploited in breeding progammes to create superior varieties. In most cases diversity levels in improved varieties are underexploited. This study evaluated the diversity level in a collection of improved genotypes in order to identify suitable material for outcrossing trials. Whilst the population size for the different variety groups may be adequate for identifying accessions with specific traits that could be used for outcrossing trials, to fully reflect the diversity in the different variety groups, larger population sizes of each group is recommended. Further studies with better representation of different species could reveal more details.

Simple sequence repeats are robust and effective in evaluating diversity levels in rice. However, there are other markers namely morphological and molecular or a combination of both which can be used to complement microsatellite techniques.

4.9 References

Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports 27:617–631.

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Science 35:1439–1445.

Baek HJ, Beharav A, Nevo E. 2003. Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan. Theoretical and Applied Genetics 106:397–410.

Barry MB, Pham JL, Noyer JL, Billot C, Courtois B, Ahmadi N. 2007. Genetic diversity of the two cultivated rice species (*O. sativa* and *O. glaberrima*) in Maritime Guinea. Evidence for interspecific recombination. Euphytica 154:127–137.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314–331.

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93:190–198.

Chang TT. 1984. Conservation of rice genetic resources: luxury or necessity? Science 24:251–256.

Chen JS, Lin SC, Chen CY, Hsieh YT, Pai PH, Chen LK, Lee S. 2014. Development of a microarray for two rice subspecies: characterization and validation of gene expression in rice tissues. BMC Research Notes 7:1–18.

Choudhury B, Khan ML, Dayanandan S. 2013. Genetic structure and diversity of indigenous rice (*Oryza sativa*) varieties in the Eastern Himalayan region of Northeast India. SpringerPlus 2:228–237.

Dale JW, von Schantz M, Plant N. 2012. From Genes to Genomes: Concepts and Applications of DNA Technology, Jonh Wiley, West Sussex, UK. 360 pp.

Edwards JD, Janda J, Sweeney MT, Gaikwad AB, Liu B, Leung H, Galbraith DW. 2008. Development and evaluation of a high-throughput, low-cost genotyping platform based on oligonucleotide microarrays in rice. Plant Methods 4:13.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Felsenstein J. 1982. How can we infer geography and history from gene frequencies? Journal of Theoretical Biology 96:9–20.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638.

Girma G, Tesfaye K, Bekele E. 2010. Inter simple sequence repeat (ISSR) analysis of wild and cultivated rice species from Ethiopia. African Journal of Biotechnology 9:5048–5059.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.

Jones MP, Dingkuhn M, Aluko GK, Semon M. 1997. Interspecific *Oryza sativa* L x *O. glaberrima* Steud. progenies in upland rice improvement. Euphytica 92:237–246.

Kleppe K, Ohtsuka E, Kleppe R, Molineux I, Khorana HG. 1971. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. Journal of Molecular Biology 56:341–336.

Kumagai M, Wang L, Ueda S. 2010. Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. Gene 462:44–51.

Kumar R, Singh AK, Kumar A, Radha. 2012. Evaluation of genetic diversity in rice using simple sequence repeats (SSR) markers. African Journal of Biotechnology 11:14956–14995.

Li YC, Fahima T, Korol AB, Peng J, Roder MS, Kirzhner V, Beiles A, Nevo E. 2000. Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in north Israel. Molecular Biology and Evolution 17:851–862.

Liu K, Muse SV. 2005. PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129.

Mohammadi SA, Prasanna BM. 2003. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Science 43:1235–1248.

Nathans D, Smith HO. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. Annual Review of Biochemistry 44:273–293.

Ndjiondjop MN, Cisse F, Girma G, Sow M, Bocco R, Djedatin G, Fatondji B. 2010. Morpho-agronomic and molecular characterisation of *Oryza glaberrima* germplasm from Mali. African Journal of Biotechnology 9:7409–7417.

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P. 2009. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers‘ fields. PLoS ONE 4:1-9.

Perrier X, Flori A, Bonnot F. 2003. Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC. (Eds.). Genetic Diversity of Cultivated Tropical Plants. Enfield, Science Publishers. Montpellier. pp 43–76.

Perrier X, Jacquemoud-Collet JP. 2006. DARwin software http://darwin.cirad.fr/darwin

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pritchard JK, Wen W. 2003. Documentation for STRUCTURE software: Version 2. Available from http://pritch.bsd.uchicago.edu. (Accessed on 20/10/2014)

Qi Y, Zhang H, Zhang D, Wang M, Sun J, Ding L, Wang F, Li Z. 2009. Assessing indica-japonica differentiation of improved rice varieties using microsatellite markers. Journal of Genetics and Genomics 36:305–312.

Rabinow P. 1996. Making PCR: A Story of Biotechnology. Chicago: University of Chicago Press. 190pp.

Rick VK, Ilan P, Richard WM. 1994. Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435–1446.

Risterucci AM, Grivet L, N'Goran JAK, Pietetti I, Flament MH, Lanaud C. 2000. A high–density linkage map of *Theobroma cacao* L. Theoretical and Applied Genetics 101:948-955.

Semagn K, Bjørnstad Å, Ndjiondjop MN. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540–2568.

Semagn K, Ndjiondjop MN, Lorieux M, Cissoko M, Jones M, McCouch S. 2007. Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). African Journal of Biotechnology 6:2014–2022.

Sokal RR, Michener CD. 1958. A statistical method for evaluating systematic relationships. University Kansas Science Bulletin 38:1409–1438.

Somaratne LHMYK, Abayawickrama ASMT, Wickramasinghe IP, Samarasinghe WLG. 2012. Estimating out-crossing rate of Bg 379-2 using morphological markers and confirmation by molecular markers. Rice Science 19:166–168.

Weir BS. 1996. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sunderland, Massachusetts: Sinaur Associates. 371 pp.

Wang M, Yu Y, Haberer G, Marri PR, Fan C. 2014. The genome sequence of African rice *(Oryza glaberrima*) and evidence for independent domestication. Nature Genetics 46:982–988.

Chapter 5

Assessing diversity of rice varieties using agro-morphological and microsatellite markers

5.1 Abstract

Agro-morphological and microsatellite markers are effective in assessing diversity of germplasm. In this study the diversity of a collection of 36 rice accessions of *O. sativa, O. glaberrima* and interspecific hybrids (*O. sativa* x *O. glaberrima*) was assessed using 10 qualitative agro-morphological and 27 microsatellite markers. The aim of the diversity assessment was to identify plant material that could be used in outcrossing trials. Accessions used in the study are adapted to the upland ecology and were mainly improved cultivars and some landraces. Each marker type was independently used to assess diversity and then a combination of the agro-morphological and microsatellite markers was also used for diversity assessment. The similarity distances ranged from 0.00–0.39 for the agro-morphological matrix, 0.81–0.91 for the microsatellite matrix and 0.73–0.91 for the combined agro-morphological and microsatellite matrix. Microsatellite data detected higher diversity between accessions in comparison to agro-morphological data. There was correlation between the different matrices, indicating that they all reflected similar patterns in the structure of diversity of the collection. However, there were differences between the datasets that reflected the type of marker used for the assessment. Microsatellite characterisation was found to be most effective in assessing diversity. However, the two marker types could be used complementary as variation revealed by microsatellite markers could be explained by agro-morphological variation.

Keywords: Agro-morphological, diversity, interspecific rice, microsatellites, *O. glaberrima, O. sativa*

5.2 Introduction

Assessing genetic diversity of germplasm in order to reveal the constituent genotypes is important for germplasm conservation, management and utilisation. The information could be useful for selecting suitable donor and recipient plants for outcrossing trials, parents for crosses, protecting intellectual property rights and for the production of good quality seed. Knowledge about existent diversity in the germplasm is important for the improvement and efficient use of available material (Saffdar *et al.,* 2009). There are different methods for estimating genetic diversity. The

comparison of such different methods could indicate how effective they are for evaluating genetic diversity and hence their usefulness in plant studies, breeding and conservation programmes.

Agro-morphological traits have been routinely used for assessing diversity of germplasm. It is the earliest method used for assessing germplasm diversity in plant studies. Agro-morphological characterisation largely involves phenotypic characters. In general, traits are either qualitative, such as the colouration of a specific appendage or quantitative in which case they are measurable and continuous like plant height. This division is not always clear, as traits can also be quasi-quantitative like the intensity of green colour of a leaf. Qualitative traits are controlled by a few major genes which could be dominant or recessive. These genes have a major effect on the phenotype which is relatively insensitive to environmental influences. The distribution of qualitative traits in a typical segregating population like an F_2 shows multiple peaks, but individuals in a category show continuous variation (Clewer and Scarisbrick, 2001). Therefore each individual can be classified into distinct categories corresponding to different genotypes. They are simple to score and exhibit Mendelian inheritance. Quantitative traits, on the other hand, are controlled by many minor genes each expressing a minor effect on the phenotype. They are largely influenced by the environment (Buckler *et al.,* 2009). Quantitative traits exhibit a normal distribution in a typical segregating F_2 population. As a result, it is more difficult to classify each individual into distinct categories. However, groups of individuals can be assigned to a class range which corresponds to a given genotype. Quantitative traits are more complex and do not exhibit Mendelian inheritance (Clewer and Scarisbrick, 2001).

Agro-morphological traits have been effectively used to assess germplasm diversity. Since the pioneering studies of Mendel in the mid-19th century these traits are still widely used in diversity assessment studies. They are simple and scoring of data is relatively rapid. There are also available statistical procedures for morphological trait analysis. There is an abundance of morphological markers available for genetic studies in rice (Khush, 1987; Biodiversity *et al*., 2007). Morphological traits are often used by farmers to characterise and name varieties (Nuijten and Almekinders, 2008).

Agro-morphological traits are widely used for rice diversity studies (Sanni *et al*., 2008; Babaei *et al*., 2011; Willocquet *et al*., 2012; Dong *et al*., 2013; Sow *et al.*, 2014). Several other studies have demonstrated the effectiveness of using agro-morphological traits in assessing diversity of rice populations. Despite their immense potential and use, agro-morphological markers have some limitations. They detect low levels of polymorphism, have low heritability and late expression.

They are also characterised by epistasis, pleiotropy and dominant-recessive relationships. They could also have a deleterious effect on the phenotype and are largely influenced by growth phases and the environment (Smith and Smith, 1992; Fufa *et al*., 2005). In addition, genomes do have sites of neutral variation at DNA level. Since these neutral sites have no effect on the phenotype, they cannot be detected using agro-morphological markers (Xu, 2010).

The paper of Botstein *et al.* (1980) that used RFLP for the construction of the linkage map of the human genome pioneered the use of DNA polymorphisms (molecular markers) as genetic markers for assessing diversity of germplasm. The advent of the PCR technique saw the development and use of many more molecular markers. A molecular marker reflects a particular segment of DNA that is representative of the differences at genome level (Agarwal *et al*., 2008). They are particularly attractive as genetic markers with numerous desirable features like being stable and detectable in all tissues, regardless of growth phase. They detect high levels of polymorphism and are generally independent of environmental influences. They are also capable of detecting neutral variation sites that have no impact on the phenotype (Maheswaran *et al*., 1997; Ridout and Donini, 1999; Semagn *et al*., 2006; Arif *et al*., 2010).

Microsatellites are a prominent PCR-based approach that is widely used in assessing diversity of germplasm. They are suitable as genetic markers as they are characterised as having high allelic variation and are highly reproducible. They are codominant, locus specific and are randomly dispersed in the genome. Microsatellites have the advantages of being easily amplified by PCR and easily detected on polyacrylamide or agarose gels. Microsatellite markers are excellent for genetic diversity analyses and genotype identification in self-pollinated species such as rice (Garris *et al*., 2005; Semon *et al*., 2005; Semagn *et al*., 2006; 2007 Barry *et al*., 2007; Qi *et al*., 2009; Ndjiondjop *et al.,* 2010; Kumar *et al.,* 2012; Tabkhkar *et al*., 2012).

Some of the drawbacks associated with microsatellites are that they survey only one locus at a time, they are complex in nature, initial screenings are expensive and they may yield only limited potential loci. There is also the problem of incompatibility of the nature of microsatellite mutations with assumptions of classical population genetics theories. Microsatellite mutations are a result of strand slippage and are not independent of the previous allele (Bhargava and Fuentes, 2010; Putman and Carbone, 2014). No particular technique is best for the study of genetic diversity in germplasm collections. Different methods test genetic variation at different levels. However, there is the need for an unambiguous, reliable, fast and cost-effective assessment of genetic diversity as this is important for determining the uniqueness and distinctiveness of the phenotypic and

genetic constitution of genotypes in order to protect breeder's intellectual property rights (Franco *et al*., 2001).

The objectives of this study were (i) to investigate genetic diversity and relationships among 36 rice genotypes adapted to the upland ecology using morphological and microsatellite markers, (ii) to assess the correlation between distance estimates based on morphological traits and molecular markers and (iii) to classify the accessions into groups based on a combination of molecular profiles and morphological traits. Results will be used to select donor and recipient genotypes and plan planting dates for subsequent outcrossing trials.

5.3 Materials and methods

5.3.1 Agro-morphological characterisation

Agro-morphological characterisation was done as described in Chapter 3, section 3.3.

5.3.2 Molecular characterisation using microsatellite markers

This was done as described in Chapter 4, subsections 4.3.1 and 4.3.2.

5.4 Data analysis

Only qualitative agro-morphological traits that are controlled by few major genes were used in this study and since they are dominant or recessive the traits can be scored as present or absent. Quantitative agro-morphological traits were not used, as they vary continuously and cannot be scored as present or absent. The 10 qualitative agro-morphological traits (Table 3.2) were divided into classes and scored as present (1) or absent (0). Similarly bands for microsatellite data were scored as present (1) or absent (0). This makes the two data sets compatible for combined analyses. Agro-morphological and microsatellite data were analysed using Numerical Taxonomy and Multivariate Analysis (NTSYSpc) version 2.21q software (Rohlf, 2012). A similarity matrix was constructed based on the simple matching coefficient using the SIMQUAL programme (Rohlf, 2012). Similarity coefficients were used to construct a dendrogram by the unweighted pair group method with arithmetic mean (UPGMA) using the sequential agglomerative hierarchical nested (SAHN) programme in NTSYS-pc 2.21q (Zhong and Steffenson, 2001). Cophenetic values were calculated using COPH and MXCOMP programmes to ascertain the goodness of fit between the cluster in the dendrogram and the matrix.

The relationship between the data sets of similarity matrices for agro-morphological and microsatellite markers was analysed using XLSTAT version 3.02 software (Addinsoft, 2008). This was done by running a two-tailed test as developed by Mantel (1967) using a Pearson correlation coefficient for 1000 permutations at P = 0.5.

The objective of the Mantel test was to compare matrices and accept the null hypothesis (H0: the matrices are not correlated) or reject the null hypothesis (Ha: the matrices are correlated). The Mantel test was not done on the matrix for the combined data set of agro-morphological and microsatellite markers as this would have violated the fundamental assumption of correlation analysis which is the independence of data sets (Clewer and Scarisbrick, 2001). In this case the combined data contained the agro-morphological and microsatellite data sets and could therefore not be tested against either of them for correlation effects.

5.5 Results

5.5.1 Genetic diversity based on agro-morphological data

The level of diversity based on the H' index as revealed by qualitative agro-morphological traits (Table 3.3) showed differences between accessions. The dendrogram generated from the clustering analysis of qualitative agro-morphological data revealed two major clusters I and II (Figure 5.1).

Cluster I consisted of *O. sativa* accessions whilst cluster II consisted of *O. glaberrima* accessions. Cluster I was subdivided into six sub-clusters namely; IA, IB, IC, ID, IE and 1F. Sub-cluster IA contained eight accessions namely; Afhikari, Moroberekan, ITA150, WAB224, WAB306, WAB100, WAB217 and WAB506. Accessions in this sub-cluster were improved *O. sativa* genotypes with the exception of Moroberekan which is *O. sativa* Japonica group landrace. Sub-cluster IB consisted exclusively of interspecific hybrid accessions namely; NERICA14, NERICA16, NERICA15 and NERICA18. Accessions in sub-cluster IB had red caryopsis colour. Sub-cluster IC contained six accessions namely; NERICA14, WAB176, WAB272, NERICA13, NERICA7 and NERICA12. This sub-cluster consisted of improved varieties of *O. sativa* Japonica group and interspecific hybrid accessions. Sub-cluster ID contained the following accessions: ITA123, WAB365, WAB56, WAB285, IRAT104, WAB128, WAB307 and WAB337. Accessions in sub-cluster ID are all improved varieties belonging to *O. sativa* Japonica group with the exception of ITA123 and IRAT104 which belong to *O. sativa* Indica Group. Sub-cluster 1E contained three accessions WAB96, TOS15729 and TOS8076.

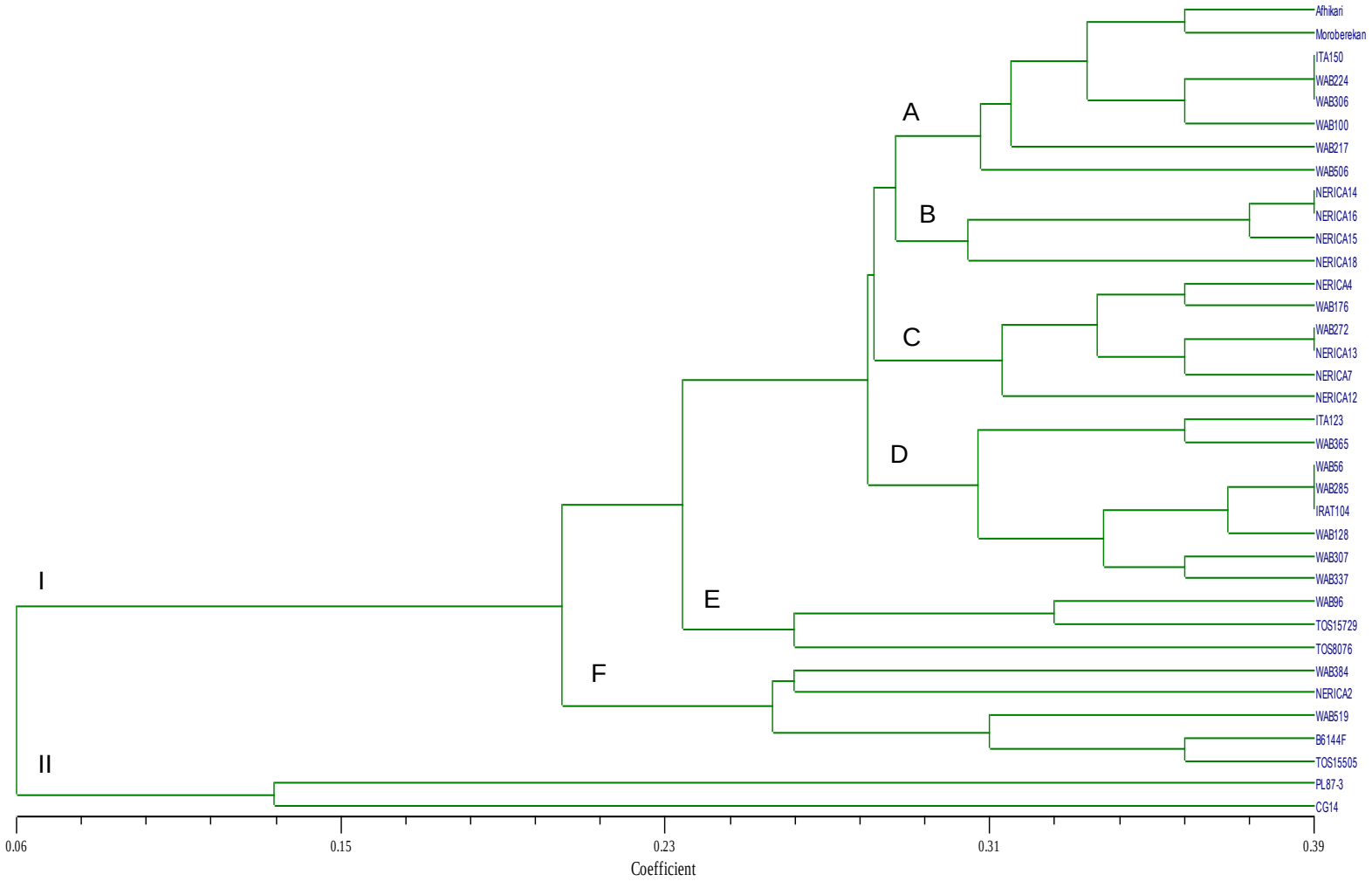


Figure 5.1 Clustering of 36 rice accessions based on 10 qualitative agro-morphological traits using simple similarity matching coefficient

Accessions in this sub-cluster had pubescent leaves and red caryopsis colour. TOS15729 and TOS8076 are landraces that belong to *O. sativa* Indica group whilst WAB96 is an improved variety of *O. sativa* Japonica group. Sub-cluster 1F contained five accessions namely; WAB384, NERICA2, WAB519, B6144F and TOS15505. Cluster II contained *O. glaberrima* accessions CG14 and PL87-3. Both accessions had open panicles and red colour of the caryopsis. Both accessions were susceptible to lodging and panicle shattering.

Table 5.1 shows the distance matrix for qualitative agro-morphological data. There was a cophenetic coefficient of r = 0.9; indicating a very good fit between generated clusters and genetic distances between accessions as determined by the similarity coefficient. The genetic similarity ranged from 0.00–0.39 with an average of 0.22 (Table 5.1). Accessions PL87-3 and NERICA15; PL87-3 and WAB176; PL87-3 and NERICA4; PL87-3 and WAB56; PL87-3 and WAB285; PL87-3 and IRAT104; PL87-3 and WAB307 and PL87-3 and WAB337 shared nothing in common agro-morphologically with a value of 0.00. Accession PL87-3 was the most distant and unique, showing an average pairwise distance of 0.05 with all other accessions in the collection. PL87-3 is a tropical *O. glaberrima* genotype that is adapted to the lowland ecology. This could explain the uniqueness of this accession as all other accessions are *O. sativa* and interspecific hybrid genotypes that are adapted to the upland ecology. Accession CG14, the only other *O. glaberrima*, had a distance of 0.13 with PL87-3.

Accessions that were most similar with the highest similarity value of 0.39 were WAB224 and WAB306; ITA150 and WAB224; ITA150 and WAB306; NERICA16 and NERICA14; NERICA15 and NERICA16; NERICA13 and WAB272; WAB285 and WAB56-104; IRAT104 and WAB56; IRAT104 and WAB285 and WAB128 and WAB285. Accessions that were similar to ITA150 and IRAT104 were generally of medium maturity with exserted panicles and a white caryopsis. Similarity of the interspecific rice accessions, which are the NERICA accessions, was based on red colour of the caryopsis.

5.5.2 Genetic diversity based on microsatellite data

Summary results for evaluation of genetic diversity using microsatellite markers are presented in Table 4.1. The dendrogram generated from the clustering analysis of microsatellite data revealed two major clusters; I and II (Figure 5.2). Cluster I was subdivided into five sub-clusters. Sub-cluster IA contained 11 accessions namely; Afhikari, NERICA4, WAB100, WAB176, WAB128, WAB224, WAB307, WAB272, WAB285, WAB217 and IRAT104.

Table 5.1 Genetic distance matrix for 10 qualitative agro-morphological traits using simple matching similarity coefficient

	Afrikari	Morobere	WAB224	WAB306	ITA150	WAB100	WAB217	WAB506	NERICA14	NERICA16	NERICA15	NERICA18	WAB176	WAB272	NERICA13	NERICA7	NERICA4	NERICA12	WAB56	WAB285	IRAT104	WAB128	WAB307	WAB337	WAB365	ITA123	WAB96	TOS15729	TOS8076	WAB384	WAB519	B6144F	TOS15505	NERICA2	PL87-3	CG14
Afrikari	1																																			
Morobereka	0.35	1																																		
WAB224	0.32	0.35	1																																	
WAB306	0.32	0.35	0.39	1																																
ITA150	0.32	0.35	0.39	0.39	1																															
WAB100	0.29	0.32	0.35	0.35	0.35	1																														
WAB217	0.26	0.29	0.32	0.32	0.32	0.35	1																													
WAB506	0.29	0.29	0.35	0.32	0.32	0.29	0.26	1																												
NERICA14	0.26	0.26	0.32	0.29	0.29	0.29	0.26	0.23	1																											
NERICA16	0.26	0.29	0.32	0.32	0.32	0.29	0.26	0.26	0.39	1																										
NERICA15	0.23	0.26	0.29	0.29	0.29	0.32	0.29	0.23	0.35	0.39	1																									
NERICA18	0.26	0.29	0.32	0.32	0.32	0.29	0.29	0.26	0.29	0.32	0.29	1																								
WAB176	0.26	0.26	0.29	0.29	0.29	0.32	0.29	0.32	0.29	0.29	0.32	0.23	1																							
WAB272	0.29	0.29	0.32	0.32	0.32	0.29	0.26	0.32	0.29	0.35	0.29	0.26	0.35	1																						
NERICA13	0.29	0.29	0.32	0.32	0.35	0.29	0.26	0.32	0.29	0.32	0.29	0.26	0.35	0.39	1																					
NERICA7	0.26	0.26	0.29	0.29	0.29	0.26	0.23	0.29	0.26	0.29	0.26	0.26	0.32	0.35	0.35	1																				
NERICA4	0.23	0.23	0.26	0.26	0.26	0.29	0.32	0.26	0.26	0.26	0.29	0.19	0.35	0.32	0.32	0.32	1																			
NERICA12	0.23	0.23	0.29	0.26	0.26	0.26	0.23	0.26	0.29	0.26	0.26	0.19	0.32	0.32	0.32	0.29	0.29	1																		
WAB56	0.23	0.26	0.29	0.29	0.32	0.32	0.29	0.23	0.29	0.29	0.32	0.29	0.32	0.29	0.29	0.29	0.29	0.26	1																	
WAB285	0.23	0.26	0.32	0.29	0.29	0.32	0.32	0.23	0.29	0.29	0.32	0.29	0.32	0.32	0.29	0.29	0.29	0.26	0.39	1																
IRAT104	0.23	0.26	0.29	0.29	0.29	0.32	0.29	0.23	0.29	0.29	0.35	0.29	0.32	0.29	0.32	0.29	0.29	0.26	0.39	0.39	1															
WAB128	0.26	0.29	0.32	0.32	0.32	0.35	0.35	0.26	0.29	0.26	0.29	0.32	0.29	0.26	0.26	0.26	0.26	0.23	0.35	0.39	0.35	1														
WAB307	0.23	0.26	0.29	0.29	0.29	0.29	0.26	0.23	0.29	0.29	0.29	0.29	0.32	0.29	0.29	0.29	0.26	0.26	0.35	0.35	0.35	0.32	1													
WAB337	0.23	0.23	0.26	0.26	0.26	0.26	0.23	0.26	0.26	0.26	0.26	0.26	0.32	0.32	0.32	0.32	0.29	0.29	0.32	0.35	0.32	0.29	0.35	1												
WAB365	0.26	0.29	0.29	0.29	0.32	0.32	0.29	0.23	0.23	0.23	0.26	0.29	0.26	0.23	0.26	0.23	0.26	0.19	0.35	0.32	0.32	0.35	0.29	0.26	1											
ITA123	0.32	0.29	0.26	0.26	0.26	0.29	0.26	0.23	0.19	0.19	0.23	0.26	0.26	0.23	0.26	0.23	0.23	0.19	0.29	0.29	0.32	0.32	0.26	0.26	0.35	1										
WAB96	0.29	0.29	0.26	0.26	0.26	0.23	0.19	0.26	0.29	0.32	0.29	0.26	0.29	0.32	0.32	0.29	0.26	0.26	0.23	0.23	0.26	0.19	0.23	0.26	0.19	0.23	1									
TOS15729	0.29	0.29	0.26	0.26	0.26	0.23	0.19	0.26	0.23	0.26	0.26	0.26	0.23	0.29	0.26	0.23	0.19	0.26	0.16	0.16	0.16	0.19	0.16	0.19	0.19	0.23	0.32	1								
TOS8076	0.23	0.26	0.23	0.23	0.23	0.26	0.23	0.16	0.23	0.23	0.26	0.23	0.19	0.16	0.16	0.13	0.16	0.19	0.19	0.19	0.19	0.23	0.16	0.13	0.23	0.23	0.23	0.29	1							
WAB384	0.19	0.19	0.16	0.16	0.16	0.19	0.16	0.16	0.16	0.16	0.19	0.16	0.26	0.23	0.23	0.23	0.23	0.23	0.26	0.26	0.26	0.26	0.23	0.26	0.23	0.26	0.23	0.19	0.16	1						
WAB519	0.26	0.26	0.23	0.23	0.23	0.19	0.16	0.23	0.16	0.19	0.16	0.16	0.23	0.26	0.26	0.23	0.19	0.19	0.16	0.16	0.16	0.16	0.16	0.19	0.16	0.19	0.26	0.23	0.19	0.23	1					
B6144F	0.19	0.19	0.16	0.16	0.16	0.19	0.19	0.16	0.13	0.13	0.16	0.16	0.23	0.19	0.23	0.19	0.19	0.16	0.23	0.23	0.23	0.23	0.19	0.23	0.26	0.29	0.19	0.16	0.19	0.29	0.32	1				
TOS15505	0.16	0.16	0.13	0.13	0.16	0.16	0.13	0.13	0.16	0.16	0.19	0.19	0.19	0.16	0.16	0.16	0.16	0.16	0.19	0.19	0.19	0.19	0.16	0.19	0.19	0.23	0.23	0.19	0.23	0.26	0.29	0.35	1			
NERICA2	0.19	0.19	0.23	0.23	0.23	0.26	0.23	0.26	0.19	0.23	0.23	0.16	0.32	0.26	0.26	0.26	0.26	0.29	0.23	0.23	0.23	0.23	0.19	0.23	0.19	0.19	0.19	0.19	0.16	0.26	0.26	0.26	0.23	1		
PL87-3	0.10	0.06	0.06	0.06	0.06	0.03	0.03	0.06	0.03	0.03	0.00	0.06	0.00	0.03	0.03	0.03	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.06	0.10	0.03	0.10	0.10	0.03	0.10	0.06	0.06	0.03	1	
CG14	0.06	0.06	0.10	0.10	0.10	0.06	0.06	0.13	0.10	0.13	0.10	0.13	0.13	0.10	0.13	0.13	0.10	0.06	0.06	0.06	0.06	0.06	0.10	0.06	0.03	0.03	0.10	0.10	0.03	0.10	0.10	0.06	0.10	0.13	0.13	1

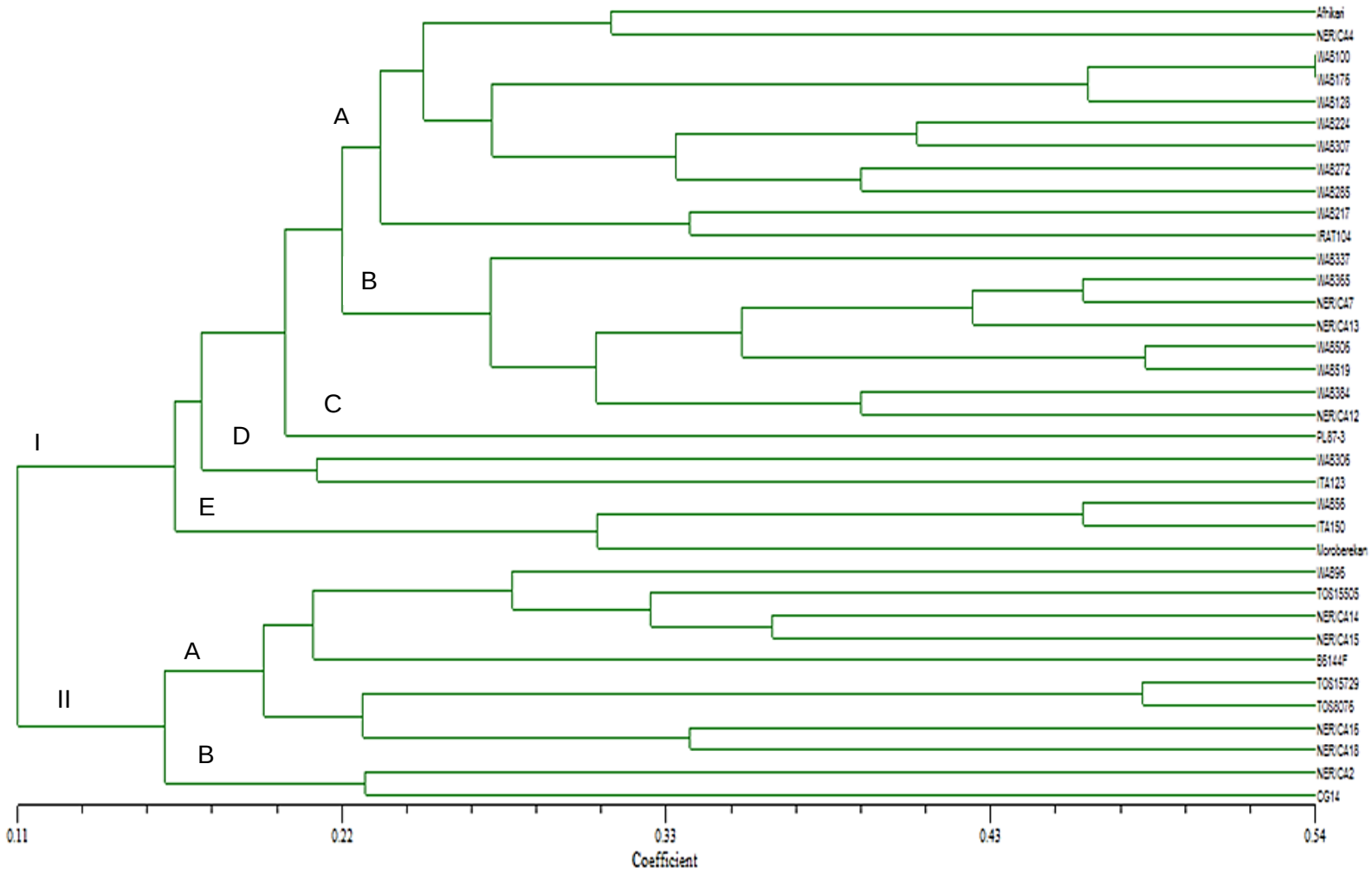


Figure 5.2 Clustering of 36 rice accessions based on 27 microsatellite marker data using the simple matching similarity coefficient

Accessions in the sub-cluster are mainly of *O. sativa* Japonica group with the exception of NERICA4 which is an interspecific hybrid. Sub-cluster IB consisted of the following accessions; WAB337, WAB365, NERICA7, NERICA13, WAB506, WAB519, WAB384 and NERICA12. Accessions in the sub-cluster mainly belong to *O. sativa* Japonica group. However, NERICA7, NERICA12 and NERICA13, which are interspecific hybrid accessions, also clustered in IB. Accession PL87-3 was an outlier and is the only one in sub-cluster IC.

Sub-cluster ID comprised two accessions namely; WAB306 and ITA123 belonging to *O. sativa* Japonica group and *O. sativa* Indica group. Sub-cluster IE contained WAB56, ITA150 and Moroberekan. Whilst WAB56 and ITA150 are improved varieties Moroberekan is a landrace. WAB56 and Moroberekan belong to *O. sativa* Japonica group and ITA150 belong to *O. sativa* Indica group.

Cluster II was subdivided into two sub-clusters (IIA and IIB). Sub-cluster IIA contained nine accessions namely; WAB96, TOS15505, NERICA14, NERICA15, B6144F, TOS15729, TOS8076, NERICA16 and NERICA18. All accessions that had red colour of the caryopsis were found in this cluster. The only exception was PL87-3 which was found in sub-cluster IC. Sub-cluster IIB consisted of NERICA2 (an interspecific hybrid accession) and CG14 which belong to *O. glaberrima*.

Table 5.2 shows the similarity matrix for the microsatellite data. There was a cophenetic coefficient of r = 0.8. The similarity matrix indicated a high level of similarity between accessions (Table 5.2). Average similarity was 0.85 which ranged from 0.73 (least related) to 0.91 (most related). The least related accessions were TOS8076 and WAB176; B6144F and WAB224; NERICA18 and WAB224; NERICA16 and ITA150; TOS8076 and WAB337 and TOS8076 and NERICA7. The most related accessions were WAB128 and WAB100 and also WAB176 and WAB100.

5.5.3 Genetic diversity based on the combined agro-morphological and microsatellite data

Table 5.3 shows the genetic similarity matrix for the combined data sets. The dendrogram generated from clustering analysis is presented in Figure 5.3. There was a cophenetic coefficient of r = 0.9. The matrix revealed high similarity of accessions (Table 5.3). Similarity ranged from 0.73 (NERICA2 and WAB56) to 0.91 (NERICA7 and NERICA13). The average genetic distance between accessions for all pairwise comparisons was 0.81.

Table 5.2 Genetic similarity distance matrix for microsatellite data

	Afrikari	NERICA4	WAB100	WAB128	WAB176	WAB217	WAB224	WAB56	ITA150	Moroberekan	WAB272	WAB285	WAB307	WAB337	WAB365	NERICA7	NERICA13	WAB506	WAB519	WAB384	NERICA12	WAB306	ITA123	PL87-3	IRAT104	B6144F	WAB96	TOS15505	NERICA14	NERICA15	TOS15729	TOS8076	NERICA16	NERICA18	NERICA2	CG14
Afrikari	1																																			
NERICA4	0.87	1																																		
WAB100	0.85	0.88	1																																	
WAB128	0.84	0.84	0.91	1																																
WAB176	0.87	0.87	0.91	0.89	1																															
WAB217	0.86	0.86	0.86	0.85	0.86	1																														
WAB224	0.85	0.87	0.87	0.88	0.86	0.88	1																													
WAB56	0.86	0.86	0.87	0.86	0.86	0.87	0.84	1																												
ITA150	0.85	0.84	0.83	0.83	0.85	0.85	0.86	0.90	1																											
Moroberekan	0.86	0.85	0.87	0.85	0.86	0.86	0.84	0.88	0.86	1																										
WAB272	0.84	0.84	0.86	0.86	0.85	0.86	0.87	0.84	0.84	0.84	1																									
WAB285	0.87	0.86	0.85	0.84	0.86	0.86	0.86	0.84	0.85	0.85	0.88	1																								
WAB307	0.87	0.87	0.87	0.86	0.87	0.84	0.89	0.85	0.86	0.86	0.87	0.88	1																							
WAB337	0.85	0.84	0.84	0.84	0.85	0.83	0.85	0.84	0.84	0.84	0.84	0.87	0.89	1																						
WAB365	0.85	0.86	0.88	0.87	0.85	0.86	0.87	0.85	0.83	0.83	0.86	0.86	0.88	0.85	1																					
NERICA7	0.86	0.85	0.86	0.86	0.85	0.84	0.86	0.84	0.84	0.84	0.85	0.87	0.88	0.89	0.89	1																				
NERICA13	0.85	0.84	0.86	0.86	0.84	0.83	0.85	0.84	0.83	0.83	0.83	0.86	0.87	0.86	0.88	0.90	1																			
WAB506	0.85	0.83	0.86	0.86	0.87	0.85	0.83	0.83	0.82	0.83	0.85	0.84	0.85	0.86	0.87	0.87	0.89	1																		
WAB519	0.83	0.86	0.85	0.84	0.84	0.86	0.84	0.83	0.82	0.83	0.84	0.84	0.85	0.84	0.86	0.87	0.89	0.89	1																	
WAB384	0.85	0.83	0.86	0.87	0.86	0.86	0.86	0.83	0.83	0.84	0.85	0.85	0.84	0.87	0.88	0.86	0.86	0.88	0.85	1																
NERICA12	0.83	0.84	0.85	0.85	0.84	0.85	0.86	0.82	0.83	0.83	0.86	0.86	0.86	0.87	0.88	0.88	0.89	0.86	0.86	0.89	1															
WAB306	0.84	0.85	0.85	0.85	0.86	0.85	0.85	0.85	0.85	0.84	0.85	0.86	0.84	0.85	0.85	0.84	0.85	0.86	0.85	0.85	0.84	1														
ITA123	0.84	0.85	0.85	0.84	0.85	0.86	0.84	0.85	0.83	0.84	0.85	0.83	0.84	0.84	0.84	0.83	0.83	0.84	0.84	0.83	0.83	0.86	1													
PL87-3	0.83	0.86	0.87	0.85	0.85	0.84	0.85	0.84	0.83	0.85	0.85	0.85	0.84	0.84	0.83	0.85	0.83	0.86	0.85	0.84	0.84	0.85	0.86	1												
IRAT104	0.86	0.83	0.85	0.84	0.84	0.86	0.83	0.85	0.85	0.84	0.84	0.85	0.84	0.84	0.85	0.86	0.86	0.84	0.85	0.83	0.85	0.85	0.85	0.83	1											
B6144F	0.83	0.83	0.83	0.82	0.82	0.83	0.81	0.84	0.86	0.83	0.82	0.82	0.84	0.82	0.84	0.83	0.86	0.83	0.84	0.83	0.83	0.83	0.85	0.82	0.86	1										
WAB96	0.83	0.82	0.83	0.83	0.83	0.86	0.82	0.82	0.82	0.83	0.83	0.86	0.83	0.84	0.84	0.82	0.84	0.84	0.86	0.83	0.84	0.85	0.85	0.83	0.86	0.85	1									
TOS15505	0.84	0.84	0.84	0.83	0.84	0.84	0.82	0.83	0.83	0.82	0.82	0.84	0.82	0.83	0.85	0.82	0.83	0.84	0.85	0.85	0.84	0.84	0.87	0.83	0.85	0.86	0.87	1								
NERICA14	0.87	0.84	0.85	0.84	0.85	0.85	0.83	0.83	0.82	0.84	0.82	0.83	0.83	0.84	0.84	0.83	0.83	0.85	0.84	0.85	0.84	0.83	0.84	0.83	0.86	0.84	0.87	0.87	1							
NERICA15	0.83	0.83	0.84	0.82	0.82	0.84	0.82	0.83	0.82	0.83	0.83	0.84	0.82	0.83	0.84	0.82	0.82	0.84	0.85	0.84	0.85	0.83	0.84	0.84	0.84	0.85	0.86	0.88	0.89	1						
TOS15729	0.84	0.82	0.82	0.82	0.82	0.84	0.83	0.82	0.83	0.86	0.82	0.83	0.82	0.83	0.82	0.82	0.82	0.83	0.83	0.83	0.82	0.85	0.84	0.82	0.85	0.84	0.86	0.85	0.85	0.86	1					
TOS8076	0.82	0.84	0.82	0.82	0.81	0.82	0.83	0.82	0.82	0.84	0.82	0.82	0.82	0.81	0.82	0.81	0.84	0.83	0.84	0.83	0.82	0.84	0.82	0.82	0.82	0.85	0.85	0.84	0.84	0.85	0.90	1				
NERICA16	0.83	0.84	0.85	0.84	0.84	0.84	0.83	0.84	0.81	0.83	0.83	0.84	0.82	0.84	0.84	0.82	0.83	0.84	0.82	0.85	0.83	0.85	0.83	0.85	0.84	0.84	0.86	0.85	0.86	0.84	0.86	0.87	1			
NERICA18	0.83	0.83	0.83	0.82	0.83	0.83	0.81	0.82	0.82	0.83	0.82	0.86	0.82	0.83	0.82	0.82	0.83	0.84	0.84	0.85	0.82	0.84	0.83	0.83	0.86	0.83	0.86	0.86	0.85	0.86	0.87	0.85	0.88	1		
NERICA2	0.83	0.83	0.84	0.83	0.83	0.83	0.83	0.83	0.82	0.82	0.83	0.82	0.84	0.86	0.85	0.85	0.85	0.86	0.86	0.84	0.83	0.83	0.84	0.83	0.85	0.83	0.86	0.84	0.84	0.83	0.83	0.83	0.83	0.84	1	
CG14	0.83	0.83	0.83	0.83	0.85	0.83	0.82	0.82	0.82	0.83	0.84	0.84	0.83	0.84	0.85	0.84	0.83	0.85	0.85	0.86	0.84	0.83	0.86	0.84	0.85	0.84	0.85	0.87	0.85	0.84	0.85	0.83	0.85	0.86	0.86	1

The dendrogram generated from the clustering analysis of the combined qualitative agro-morphological and microsatellite data revealed two major clusters A and B and nine sub-clusters. Cluster B contained two accessions WAB56 and CG14. These two belong to the species *O. sativa* Japonica group and *O. glaberrima* respectively. Cluster A which contained 34 accessions is subdivided into two sub-clusters a and b. Sub-cluster a contains two accessions Afrikari and Moroberekan. Whilst both accessions belong to *O. sativa* Japonica group, the first is of Asian origin and the second is an African landrace. Sub-cluster b contained 32 accessions and is further subdivided into b1 and b2. Sub-cluster b2 contained five accessions namely PL87-3, NERICA16, NERICA18, TOS8076 and TOS15729. Accessions in sub-cluster b2 all have red colour of the caryopsis. Sub-cluster b1 contains 27 accessions that are sub-clustered into six groups.

The first sub-group b1(i) contained eight accessions namely; WAB100, WAB176, WAB128, WAB224, WAB272, WAB307, WAB217, and WAB285. Accessions in the sub-group are improved varieties belonging to *O. sativa* Japonica group. WAB100, WAB176 and WAB128 clustered together with WAB100 and WAB176 being closer whilst WAB128 is an outlier within the sub-group. Similarly WAB224 is an outlier in a clustering of WAB224, WAB272 and WAB307. The latter two accessions clustered more closely.

The second sub-group b1(ii) contained six accessions namely; WAB337, WAB365, NERICA7, NERICA13, WAB506 and WAB519. Accessions in this sub-group belong to *O. sativa* Japonica group and interspecific hybrids. The two interspecific hybrid accessions NERICA7 and NERICA13 were tied together. Sub-group b1(iii) contained eight accessions namely; ITA123, IRAT104, B6144F, NERICA 2, WAB96, TOS15505, NERICA14 and NERICA15. Accessions in the sub-group belong to *O. sativa* Indica group (ITA123, IRAT104, B6144F and TOS15505), *O. sativa* Japonica group (WAB96) and interspecific hybrids (NERICA2, NERICA14 and NERICA15). Sub-group b1(iv) contained two accessions; WAB306 (*O. sativa* Japonica group genotype) and NERICA4 (interspecific hybrid). Sub-group b1(v) contained a single accession which is WAB384 and belonging to *O. sativa* Japonica group. Sub-group b1(vi) contained two accessions; ITA150 which belongs to *O. sativa* Indica group and NERICA12 which is an interspecific hybrid.

Table 5.3 Genetic similarity distance matrix for qualitative agro-morphological and microsatellite data

	Afrikari	Moroberekan	WAB100	WAB176	WAB128	WAB224	WAB272	WAB307	WAB217	WAB285	WAB337	WAB365	NERICA7	NERICA13	WAB506	WAB519	WAB96	TOS15505	NERICA14	NERICA15	ITA123	IRAT104	B6144F	NERICA2	WAB306	NERICA4	WAB384	ITA150	NERICA12	TOS15729	TOS8076	NERICA16	NERICA18	PL87-3	WAB56	CG14
Afrikari	1																																			
Moroberekan	0.84	1																																		
WAB100	0.82	0.83	1																																	
WAB176	0.85	0.84	0.90	1																																
WAB128	0.80	0.82	0.90	0.87	1																															
WAB224	0.81	0.81	0.87	0.85	0.87	1																														
WAB272	0.80	0.82	0.85	0.85	0.86	0.86	1																													
WAB307	0.83	0.82	0.87	0.86	0.85	0.87	0.87	1																												
WAB217	0.81	0.82	0.85	0.85	0.86	0.86	0.84	0.83	1																											
WAB285	0.81	0.81	0.85	0.84	0.84	0.86	0.86	0.87	0.87	1																										
WAB337	0.80	0.81	0.86	0.84	0.85	0.83	0.84	0.87	0.84	0.85	1																									
WAB365	0.80	0.80	0.86	0.83	0.87	0.85	0.87	0.87	0.85	0.84	0.86	1																								
NERICA7	0.82	0.81	0.86	0.84	0.88	0.86	0.84	0.87	0.86	0.87	0.89	0.89	1																							
NERICA13	0.79	0.80	0.85	0.82	0.86	0.85	0.82	0.85	0.85	0.86	0.86	0.87	0.91	1																						
WAB506	0.81	0.81	0.85	0.85	0.85	0.84	0.84	0.86	0.83	0.84	0.85	0.86	0.86	0.87	1																					
WAB519	0.80	0.79	0.85	0.84	0.83	0.82	0.83	0.85	0.85	0.83	0.85	0.86	0.87	0.87	0.88	1																				
WAB96	0.79	0.79	0.83	0.82	0.82	0.82	0.83	0.83	0.84	0.85	0.83	0.84	0.82	0.82	0.86	0.85	1																			
TOS15505	0.82	0.80	0.84	0.83	0.84	0.82	0.82	0.83	0.84	0.82	0.83	0.84	0.83	0.83	0.85	0.85	0.86	1																		
NERICA14	0.83	0.81	0.85	0.83	0.85	0.84	0.81	0.83	0.85	0.83	0.85	0.84	0.85	0.85	0.85	0.84	0.86	0.88	1																	
NERICA15	0.78	0.79	0.86	0.82	0.84	0.82	0.83	0.82	0.83	0.82	0.85	0.83	0.82	0.83	0.83	0.85	0.84	0.87	0.88	1																
ITA123	0.81	0.81	0.86	0.85	0.85	0.86	0.84	0.83	0.86	0.83	0.84	0.83	0.84	0.84	0.83	0.84	0.83	0.86	0.84	0.85	1															
IRAT104	0.80	0.79	0.85	0.82	0.84	0.82	0.83	0.83	0.85	0.84	0.84	0.85	0.85	0.84	0.83	0.84	0.83	0.82	0.84	0.83	0.85	1														
B6144F	0.80	0.80	0.85	0.84	0.83	0.82	0.82	0.85	0.85	0.85	0.84	0.83	0.85	0.86	0.84	0.85	0.85	0.86	0.86	0.85	0.86	0.86	1													
NERICA2	0.79	0.79	0.85	0.83	0.84	0.82	0.82	0.84	0.85	0.85	0.85	0.84	0.86	0.87	0.86	0.85	0.85	0.83	0.86	0.84	0.85	0.85	0.87	1												
WAB306	0.79	0.82	0.83	0.83	0.83	0.84	0.86	0.85	0.84	0.83	0.83	0.86	0.83	0.83	0.84	0.83	0.82	0.83	0.81	0.80	0.85	0.84	0.80	0.81	1											
NERICA4	0.83	0.81	0.86	0.86	0.82	0.83	0.82	0.86	0.83	0.83	0.82	0.84	0.83	0.82	0.82	0.85	0.81	0.82	0.82	0.81	0.83	0.83	0.82	0.83	0.83	1										
WAB384	0.81	0.79	0.85	0.83	0.84	0.82	0.83	0.82	0.82	0.81	0.85	0.85	0.83	0.82	0.85	0.82	0.82	0.84	0.83	0.81	0.80	0.80	0.81	0.80	0.82	0.81	1									
ITA150	0.81	0.83	0.82	0.84	0.83	0.84	0.84	0.83	0.83	0.82	0.84	0.83	0.83	0.82	0.80	0.81	0.81	0.82	0.81	0.81	0.82	0.83	0.83	0.80	0.82	0.82	0.79	1								
NERICA12	0.79	0.83	0.81	0.81	0.82	0.83	0.84	0.81	0.80	0.80	0.83	0.84	0.83	0.83	0.83	0.81	0.80	0.81	0.81	0.81	0.80	0.80	0.78	0.78	0.80	0.79	0.83	0.83	1							
TOS15729	0.80	0.82	0.80	0.80	0.79	0.78	0.79	0.81	0.82	0.81	0.81	0.79	0.80	0.79	0.80	0.81	0.82	0.82	0.81	0.82	0.80	0.80	0.82	0.80	0.82	0.81	0.79	0.81	0.81	1						
TOS8076	0.78	0.81	0.81	0.81	0.83	0.81	0.81	0.82	0.83	0.83	0.80	0.81	0.82	0.83	0.82	0.83	0.84	0.84	0.82	0.82	0.82	0.81	0.86	0.84	0.81	0.82	0.79	0.81	0.79	0.88	1					
NERICA16	0.79	0.81	0.80	0.81	0.81	0.80	0.80	0.78	0.80	0.80	0.79	0.81	0.81	0.81	0.81	0.79	0.81	0.82	0.84	0.80	0.80	0.79	0.79	0.80	0.81	0.80	0.80	0.79	0.84	0.84	0.83	1				
NERICA18	0.80	0.81	0.81	0.82	0.80	0.80	0.80	0.81	0.80	0.82	0.81	0.80	0.81	0.81	0.83	0.82	0.82	0.85	0.84	0.85	0.81	0.82	0.81	0.81	0.81	0.81	0.83	0.80	0.80	0.85	0.81	0.87	1			
PL87-3	0.80	0.79	0.83	0.82	0.81	0.81	0.81	0.79	0.79	0.79	0.80	0.80	0.82	0.80	0.82	0.81	0.80	0.81	0.81	0.82	0.83	0.80	0.78	0.80	0.80	0.82	0.81	0.80	0.80	0.79	0.77	0.84	0.85	1		
WAB56	0.77	0.78	0.77	0.76	0.77	0.74	0.74	0.76	0.77	0.75	0.74	0.75	0.75	0.75	0.74	0.74	0.74	0.75	0.75	0.74	0.75	0.74	0.75	0.73	0.74	0.77	0.77	0.78	0.74	0.75	0.74	0.76	0.75	0.75	1	
CG14	0.74	0.75	0.78	0.77	0.74	0.76	0.77	0.78	0.75	0.76	0.77	0.76	0.75	0.75	0.80	0.79	0.78	0.80	0.78	0.77	0.77	0.76	0.77	0.78	0.75	0.76	0.77	0.76	0.78	0.77	0.77	0.78	0.78	0.74	0.78	1

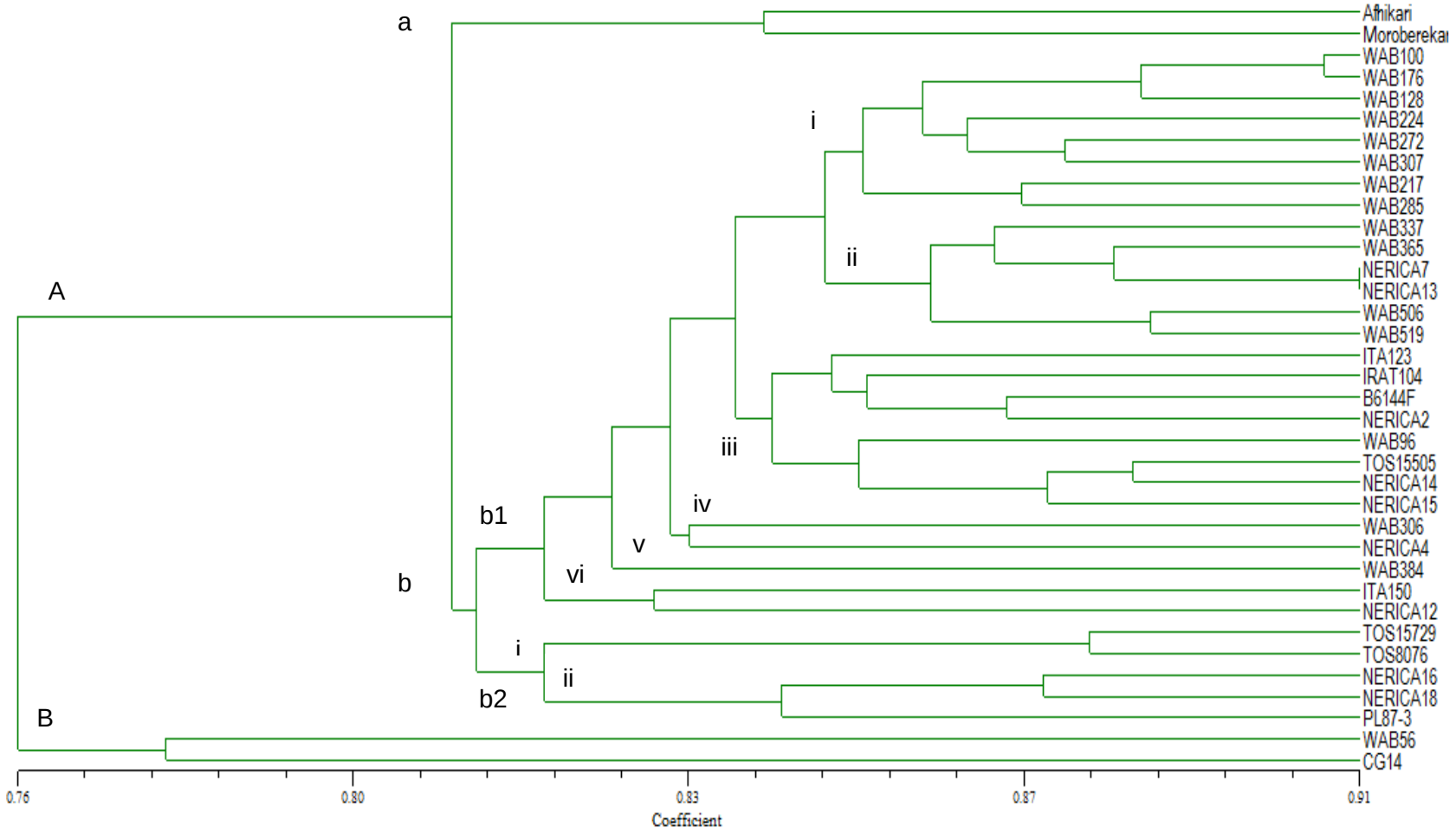


Figure 5.3 Clustering of 36 rice accessions based on combined data of 10 qualitative agro-morphological and 27 microsatellite marker data using the simple matching similarity coefficient

Accessions in sub-cluster b1 were improved cultivars of early to medium maturity and of intermediate plant height. Accessions are resistant to lodging and had well exserted panicles with long grains and white caryopsis colour. This sub-cluster contained accessions belonging to *O. sativa* Japonica and Indica groups and interspecific hybrids. However, landrace accessions TOS15505 (*O. sativa* Indica group) and Moroberekan (*O. sativa* Japonica group) also clustered in sub-cluster b1.

Sub-cluster b2 was subdivided into two sub-groups; b2(i) and b2(ii) and contained accessions that had red colour of the caryopsis. Sub-group b2(i) contained two accessions namely TOS15729 and TOS8076. Both accessions are landraces that belong to *O. sativa* Indica group and have pubescent leaves and red caryopsis colour. Sub-group b2(ii) contained three accessions namely; NERICA16; NERICA18 and PL87-3. NERICA16 and NERICA18 are interspecific hybrids and they clustered close together whilst PL87-3 which is *O. glaberrima,* was an outlier in the group. All accessions have red colour of the caryopsis.

5.5.4 Correlations between similarity matrices

Table 5.4 shows the correlation between the distance matrices of the qualitative agro-morphological and microsatellite data sets. There was a positive correlation between the matrices of the two data sets. This indicated that the two data sets reflected similar diversity within the collection of rice accessions used in this study.

Table 5.4 Mantel test for qualitative agro-morphological and microsatellite distance matrices

Item	Value
r|AB|	0.092
p-value	0.023
alpha	0.05

Table 5.5 gives the summary statistics of similarity coefficients for the three data sets. The agro-morphological distance matrix showed lower similarity but with a larger range. The microsatellite data showed higher similarity between accessions. The combined data showed similarity that was closer to that shown by microsatellite data.

Table 5.5 Minimum, maximum and average similarity coefficient for qualitative agro-morphological, microsatellite and combined marker data

Marker type	Minimum	Maximum	Average
Agro-morphological	0.00	0.39	0.22
Microsatellite	0.81	0.91	0.85
Combined marker data	0.73	0.91	0.81

5.6 Discussion

Agro-morphological and microsatellite characterisation effectively revealed diversity in the collection of rice accessions used in the study. Accessions were clustered into two major clusters irrespective of data type. However, subdivisions of clusters differed between data type. The number of accessions allocated to each cluster was similar for clustering based on qualitative agro-morphological and combined agro-morphological and microsatellite data. Each of the two clustering methods allocated 34 accessions to one cluster and only two accessions to the other. However, for clustering based on microsatellite data 25 accessions were allocated to one cluster and 11 accessions to the second.

Each of the clustering methods detected sets of accessions that had no difference and were tied together. Clustering based on microsatellite data detected no difference between WAB100 and WAB176. Both accessions were developed by AfricaRice and share the same female parent. Clustering based on qualitative agro-morphological data detected no difference between ITA150 and WAB224. The two accessions belong to variety group Indica and Japonica respectively. Both accessions are improved varieties which are high yielding with white caryopsis and resistant culms. Similarly agro-morphological clustering did not detect any difference between NERICA14 and NERICA16. Both accessions share a common gene pool and they are high yielding with long grains and red caryopsis colour. WAB56, WAB285 and IRAT104 were also tied together. Clustering based on combined microsatellite and agro-morphological data detected no difference between NERICA7 and NERICA13. Both accessions share similar donor parents. They are tall, early maturity with long and heavy grains (Appendix 5).

WAB100 and WAB224 always clustered together irrespective of data type. Accessions Afhikari, NERICA4, WAB100, WAB176, WAB128, WAB224, WAB307, WAB272, WAB285, WAB217 and IRAT104 that clustered together in sub-cluster IA of the microsatellite dendrogram generally

clustered together in sub-cluster Ab1(i) in the combined dendrogram. Only Afhikari, NERICA4 and IRAT104 clustered in sub-clusters Aa, Ab1(iv) and Ab1(iii). However, accessions in sub-cluster IA of the microsatellite dendrogram clustered in sub-clusters IA, IC, and ID in the agro-morphological dendrogram. Accessions WAB337, WAB365, NERICA7, NERICA13, WAB506, WAB519, WAB384, and NERICA12 in sub-cluster IB in microsatellite dendrogram clustered together in the combined dendrogram with the exception of WAB384 and NERICA12. In the agro-morphological dendrogram, the same accessions clustered in several sub-clusters: IA (WAB506), IC (NERICA7, NERICA and NERICA13), ID (WAB337 and WAB365), IF (WAB519 and WAB384). Accessions WAB96, TOS15505, NERICA14, NERICA15, B6144F, TOS15729, TOS8076, NERICA16 and NERICA18 which clustered together in sub-cluster IIA in the microsatellite dendrogram clustered in sub-cluster Ab1(iii) (WAB96, TOS15505, NERICA14, NERICA15 and B6144F) and sub-cluster Ab2(i) (TOS15729 and TOS8076) and in sub-cluster Ab2(ii) (NERICA16 and NERICA 18) in the combined dendrogram. The same accessions were found in sub-cluster IB (NERICA14, NERICA15, NERICA18); in sub-cluster IE (WAB96, TOS15729, TOS8076); and in sub-cluster IF (TOS15505, B6144F) in the agro-morphological dendrogram.

The agro-morphological dendrogram revealed a similarity from 0.06 to 0.39. The most similar accessions were thus only 39% similar (e.g. NERICA14 and NERICA16) and there was a total of 6% similarity between groups I and II. The microsatellite dendrogram detected lower levels of variation, ranging from 0.11 to 0.48. The most similar accessions (e.g. TOS15729 and TOS8076) were 48% similar. It thus detected less variation than agro-morphological data.

The microsatellite dendrogram clustered accessions based on genetic similarity whereas, agro-morphological dendrogram was based on phenotypic similarity. For instance CG14 is one of the parents of NERICA2 and the two accessions clustered together in sub-cluster IIB in the microsatellite dendrogram. However, the two accessions were separated in the agro-morphological dendrogram; CG14 and PL87-3 clustered together in cluster II. The two are phenotypically similar with high levels of panicle shattering and culm lodging, and awned grains with red caryopsis colour.

The significant correlation between the distance matrices of agro-morphological and microsatellite matrices indicated that both approaches displayed similar results. Similar trends between agro-morphological and microsatellite markers were reported in other studies (Geleta and Labuschagne, 2005; Geleta *et al*., 2006; Singh *et al*., 2011). Both agro-morphological characterisation and molecular characterisation by microsatellite data revealed two major

clusters. The latter largely reflected the underlying population structure of the collection as revealed by STRUCTURE analysis for population structure, which showed a value of K = 2. The dendrogram revealed by the combined data set of agro-morphological and microsatellite markers revealed two major clusters; however, the assignment of accessions within the clusters was closer to that of microsatellite characterisation. Notwithstanding, dendrograms generated from agro-morphological and microsatellite data showed some differences.

The difference in distance matrices and clustering by dendrograms between microsatellite and agro-morphological characterisation is a reflection of the inherent characteristics of the markers. Agro-morphological markers are widely used in rice studies (Ndjiondjop *et al*., 2010; Moukoumbi *et al*., 2011; Sow *et al*., 2014). However, agro-morphological markers with distinct polymorphisms are few and can only detect traits that are being expressed visually. They are further characterised by GxE interactions, epistasis, pleiotropy and dominant-recessive relationships (Smith and Smith, 1992). They also cannot detect sites of neutral variation at DNA level (Xu, 2010). Microsatellite markers have high allelic variation, are codominant and are dispersed throughout the genome. Therefore they are more effective in detecting similarities and differences between genotypes and are more effective than agro-morphological markers in assessing genetic diversity. Molecular markers, including microsatellites, are routinely used in rice diversity studies (Semon *et al*., 2005; Barry *et al*., 2007; Kumar *et al*., 2012). Both marker types effectively divided the collection into two clusters and several sub-clusters that contained groups of accessions that could be used for outcrossing studies.

5.7 Conclusions

The assessment of genetic diversity can be effectively done using agro-morphological or microsatellite markers. However, there are differences in the level and pattern of diversity revealed by either of the two approaches. The agro-morphological and microsatellite markers used in studying the collection of 36 rice accessions effectively revealed levels of diversity in the collection. Accessions were grouped into two major clusters by agro-morphology and microsatellites respectively. Notwithstanding the differences between agro-morphological and molecular characterisation in evaluating diversity, the two are equally useful. An approach that combined the agro-morphological and microsatellite data sets reflected a similar pattern as for the two other approaches.

5.8 Recommendation

Molecular markers aid plant breeders to identify desirable alleles that could be incorporated into crops. They ensure that traits are incorporated at early stages of breeding, therefore saving time and resources. However, sometimes the application of molecular techniques in breeding is impractical, as is the case for complex traits like yield. Agro-morphological markers are also effective in assessing diversity of germplasm, especially for traits that are distinct and are reflected on the phenotype or for complex traits, like yield. However, there are several limitations associated with agro-morphological markers. Microsatellite and agro-morphological characterisation are effective in assessing variation in rice germplasm and are important tools for crop improvement and improved productivity. The two approaches are complementary; agro-morphological characterisation could be used for preliminary assessment of diversity among morphologically distinguishable accessions, followed by molecular characterisation to further detect similarities and differences and clarify the variation.

Knowledge generated from the structure of diversity could be effectively used to select parental material, and to manage the conservation and utilisation of *in situ* and *ex situ* germplasm. It could also be used to elucidate outcrossing levels under field conditions. Such knowledge could be essential to limit unwanted outcrossing in order to assure the production of good quality seed.

5.9 References

Addinsoft. 2008. XLSTAT Version 3.02. http://www.xlstat.com/en/. (Accessed 18/07/2015)

Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports 27:617–631.

Arif IA, Bakir MA, Khan HA, Farhan AHA, Homaidan AAA, Bahkali AH, Sadoon MA, Shobrak M. 2010. A brief review of molecular techniques to assess plant diversity. International Journal of Molecular Sciences 11:2079–2096.

Babaei A, Nematzadeh GA, Hashemi H. 2011. An evaluation of genetic differentiation in rice mutants using semi-random markers and morphological characteristics. Australian Journal of Crop Science 5:1715–1722.

Barry MB, Pham JL, Noyer JL, Billot C, Courtois B, Ahmadi N. 2007. Genetic diversity of the two cultivated rice species (*O. sativa* and *O. glaberrima*) in Maritime Guinea. Evidence for interspecific recombination. Euphytica 154:127–137.

Bhargava A, Fuentes FF. 2010. Mutational dynamics of microsatellites. Molecular Biotechnology 44:250–266.

Bioversity International, IRRI, WARDA. 2007. Descriptors for Wild and Cultivated Rice (*Oryza* spp.), Bioversity International, Rome, Italy; International Rice Research Institute, Los Baños, Philippines; WARDA, Africa Rice Center, Cotonou, Benin. 63 pp.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314–333.

Buckler ES, Holland JB, Brandbury PJ, Acharya CB, Brown PJ, Ersoz E, McMullen MD. 2009. The genetic architecture of maize flowering time. Science 325:714–718.

Clewer AG, Scarisbrick DH. 2001. Practical statistics and experimental design for plant and crop science. John Wiley and Sons Ltd. Chichester, 346 pp.

Dong CF, Ding CL, Xu NX, Cheng YH, Shen YX, Gu HR. 2013. Double-purpose rice (*Oryza sativa* L.) variety selection and their morphological traits. Field Crops Research 149:276–282.

Franco J, Crossa J, Ribaut JM, Betran J, Warburton ML, Khairallah M. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. Theoretical and Applied Genetics 103:944–952.

Fufa H, Baenziger PS, Beecher BS, Dweikat I, Graybosch RA, Eskridge KM. 2005. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133–146.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638.

Geleta N, Labuschagne MT. 2005. Qualitative traits variation in sorghum [*Sorghum bicolour* (L.) Moench] germplasm from eastern highland of Ethiopia. Biodiversity and Conservation 14:3055–3064.

Geleta N, Labuschagne MT, Viljoen CD. 2006. Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodiversity and Conservation 15:3251–3265.

Khush GS. 1987. List of gene markers maintained in the Rice Genetic Stock Center, IRRI. Rice Genetics Newsletter 4:56–62.

Kumar R, Singh AK, Kumar A, Radha. 2012. Evaluation of genetic diversity in rice using simple sequence repeats (SSR) markers. African Journal of Biotechnology 11:14956–14995.

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N. 1997. Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. Theoretical and Applied Genetics 94:39–45.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220.

Moukoumbi YD, Sie M, Vodouhe R, N'dri B, Toulou B, Ogunbayo SA, Ahanchede A. 2011. Assessing phenotypic diversity of interspecific rice varieties using agro-morphological characterization. Journal of Plant Breeding and Crop Science 3:74–86.

Ndjiondjop MN, Cisse F, Girma G, Sow M, Bocco R, Djedatin G, Fatondji B. 2010. Morpho-agronomic and molecular characterisation of *Oryza glaberrima* germplasm from Mali. African Journal of Biotechnology 9:7409–7417.

Nuijten E, Almekinders DCJM. 2008. Mechanisms explaining variety naming by farmers and name consistency of rice varieties in the Gambia. Economic Botany 62:148–160.

Putman AI, Carbone I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecology and Evolution 4: 4399–4428.

Qi Y, Zhang H, Zhang D, Wang M, Sun J, Ding L, Wang F, Li Z. 2009. Assessing *indica-japonica* differentiation of improved rice varieties using microsatellite markers. Journal of Genetics and Genomics 36:305–312.

Ridot CJ, Donini P. 1999. Use of AFLP in cereals research. Trends in Plant Science 4:76–79.

Rohlf FJ. 2012. NTSYSpc: numerical taxonomy system version 2.21q. Exeter Software: Setauket: New York.

Saffdar H, Ashfaq M, Hameed S, Ul Haque I, Mujeeb-Kazi A. 2009. Molecular analysis of genetic diversity in elite II synthetic hexaploid wheat screened against Barley yellow dwarf virus. African Journal of Biotechnology 8:3244–3250.

Sanni KA, Fawole I, Guei RG, Ojo DK, Somado EA, Tia DD, Ogunbayo SA, Sanchez I. 2008. Geographical patterns of phenotypic diversity in *Oryza sativa* landraces of Cote d'Ivoire. Euphytica 160:389–400.

Semagn K, Bjørnstad Å, Ndjiondjop MN. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540–2568.

Semagn K, Ndjiondjop MN, Lorieux M, Cissoko M, Jones M, McCouch S. 2007. Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). African Journal of Biotechnology 6:2014–2022.

Semagn K, Ndjiondjop MN, Cissoko M. 2006. Microsatellites and agronomic traits for assessing genetic relationships among 18 New Rice for Africa (NERICA) varieties. African Journal of Biotechnology 5:800–810.

Semon M, Nielsen R, Jones MP, McCouch SR. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O*. *sativa* and ecological adaptation. Genetics 169:1639–1647.

Singh VK, Upadhyay P, Sinha P, Mall AK, Jaiswal SK, Singh A, Ellur RK, Biradar S, Sundaram RM, Singh S, Ahmed I, MIshra B, Singh AK, Kole C. 2011. Determination of genetic relationships among elite thermosensitive genic male sterile lines (TGMS) of rice (*Oryza sativa* L.) employing morphological and simple sequence repeat (SSR) markers. Journal of Genetics 90:11–19.

Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. Advances in Agronomy 47:85–141.

Sow M, Sido A, Laing M, Ndjiondjop MN. 2014. Agro-morphological variability of rice species collected from Niger. Plant Genetic Resources 12:22–34.

Tabkhkar N, Rabiei B, Sabouri A. 2012. Genetic diversity of rice cultivars by microsatellite markers tightly linked to cooking and eating quality. Australian Journal of Crop Science 6:980–985.

Willocquet L, Noel M, Sackville HR, Savary S. 2012. Susceptibility of rice to sheath blight: an assessment of the diversity of rice germplasm according to genetic groups and morphological traits. Euphytica 183:227–241.

Xu Y. 2010. Molecular plant breeding. CAB International, UK, 743 pp.

Zhong S, Steffenson BJ. 2001. Virulence and molecular diversity in *Cochliobolus sativus*. Phytopathology 91:469–476.

Chapter 6

Estimation of outcrossing rates in intraspecific (*Oryza sativa*) and interspecific (*Oryza sativa* x *Oryza glaberrima*) rice under field conditions using agro-morphological markers

6.1 Abstract

Outcrossing rates in rice under field conditions were investigated using cultivar WAB96-1-1 as a pollen donor and WAB56-104, NERICA 2, NERICA 4 and NERICA 7 as pollen recipients. Levels of outcrossing were investigated up to 30 m from the pollen donor. Dominant morphological markers of red kernel colour and pubescent leaves of the donor were used to identify hybrids. A total of 721 134 plants were investigated. There was an average outcrossing rate of 0.7% $\pm$ 0.51, with a potential outcrossing rate of 2.45% $\pm$ 0.86. Outcrossing rates decreased with increase in distance. It ranged from 2.45% at 0.2 m from the donor to 0.05% at 25 m from the donor. Differences were observed between genotypes and seasons. In season 1 the highest average outcrossing rate of 1.2% $\pm$ 0.63 was with WAB56-104 and in season 2 it was 1.1% $\pm$ 0.69 with NERICA 4. Outcrossing occurred up to 30 m from the donor. This has implications for germplasm management and conservation and the production of high quality seed. Spatial isolation remains the most practical method to prevent undesirable gene flow. The study indicated that red kernel colour and leaf pubescence can be used to effectively assess outcrossing under field conditions in rice.

Keywords: Agro-morphological markers, gene flow, outcrossing, rice, seed quality

6.2 Introduction

Rice is an important food crop worldwide and especially for people in the developing world. It is also important for rice research and plant studies serving as a model crop for which the genome sequence is known (IRGSP, 2005). Rice is mainly a self-pollinating plant, with fertilisation of the ovary by pollen occurring within the same flower. The viability of rice pollen is about 10–20 min, and it is influenced by temperature and humidity (Matsui *et al*., 2000; Jagadish *et al.,* 2007). The ovary has been reported to be viable and remaining receptive for several days (Gealy *et al*., 2003).

Outcrossing, which is fertilisation of the ovary by pollen from a different flower or plant, has been reported in several studies on rice (Langevin, 1990; Ellstrand *et al.*, 1999; Ellstrand*,* 2003; Chen *et al*., 2004; Da Silva *et al*., 2005; Endo *et al*., 2009; Nuijten *et al*., 2009). Cross pollination is a prerequisite for outcrossing and certain conditions must be met for cross pollination to be

effective. The conditions are: there must be a pollen vector; plants must be close enough; flowering periods must overlap and there should be a receptive ovary for viable pollen.

Rice is an anemophilous plant in which pollination and by inference outcrossing, is mainly by wind. However, insects have also been found to influence levels of outcrossing (Gealy *et al*., 2003). Outcrossing with an undesirable donor would lead to the creation of segregants. In seed production such plants, referred to as off-types, would adversely affect the acceptable genetic purity and uniformity of the crop. This is a critical issue in seed production because a seed crop that is below the acceptable standards would be rejected and this leads to economic loss (Lee *et al*., 2013).

Outcrossing essentially involves the exchange of genetic material between genotypes. Methods that measure gene flow between genotypes are effective for measuring levels of outcrossing between genotypes. Methods for measuring gene flow between genotypes fall into two broad categories namely, direct and indirect methods. Direct methods utilise parentage analysis to estimate gene flow. This involves progeny analysis for the presence of specific markers or traits associated with parents. Frequently dispersal patterns and levels of pollen, seed, gene or unique traits of genotypes are used to estimate levels of gene flow (Messeguer *et al*., 2004; Song *et al*., 2004; Da Silva *et al*., 2005; Lu and Snow, 2005; Endo *et al*., 2009; Kanya *et al.,* 2009; Shivrain *et al*., 2009; Phan *et al*., 2012). Indirect methods use a specific parameter to characterise the distribution of genotypes and then apply population genetics theory to estimate gene flow (Neigel, 1997; Ellstrand, 2003). Recently molecular markers have been increasingly used in outcrossing studies (Song *et al*., 2003; 2004, Rong *et al*., 2007).

Various levels of outcrossing in rice have been reported. Da Silva *et al.* (2005) and Endo *et al.* (2009) used morphological markers to evaluate outcrossing in rice and reported levels less than 1%. However, the use of only one donor and one recipient in these studies was not sufficient to reflect differences due to genotype effect. The study of Rong *et al.* (2007) used molecular markers and different sizes of recipient plots to evaluate outcrossing in three insect resistant GM rice and non-GM isogenic lines. They reported low levels and a dramatic reduction of outcrossing with increased distance from the donor and no effect of plot size on the levels of outcrossing. However, as plot sizes of recipient plots were larger than the donor plot, there was always enough recipient pollen to annul competition from the donor pollen.

There has been a sustained increase in the use of GM crops over recent decades. In 2014 global coverage of GM crops was 181.5 million ha (ISAAA, 2014). There is a wide and diverse range of

research associated with GM crops (Lu and Snow, 2005; Table 2.3). Ongoing research on GM rice focuses on yield increase, including various biotic and abiotic constraints to productivity. GM rice research also looks at quality enhanced rice and consumer preference (high protein and vitamin content). There is also active research on GM rice for the pharmaceutical industry (Chen *et al.,* 2009). GM crops have been reported to have a positive impact on agriculture and the environment (Brookes and Barfoot, 2012). However, there is apprehension over the possibility of undesirable outcrossing between GM and non-GM crops. This would lead to gene exchange and would invariably lead to negative consequences (Ellstrand *et al*., 1999). In rice the concern is the occurrence of undesirable outcrossing between GM rice and non-GM rice and other wild relatives of rice. Several studies have reported outcrossing between GM rice and non-GM rice (Messeguer *et al*., 2001; Chen *et al*., 2004; Rong *et al*., 2006).

Sub-Saharan Africa has also witnessed an increased coverage of rice. This has led to a corresponding increase in demand for quality seed (FAOSTAT, 2013). Rice research in Africa is exploiting new gene pools for crop improvement. Rice breeding is currently focused on developing improved varieties derived from crosses involving *O. glaberrima* and *O. barthii* (AfricaRice, 2011). *Oryza glaberrima* is well adapted to Africa and can withstand biotic and abiotic stresses on the continent. The species can be a source of useful genes for rice improvement (Sarla and Swamy, 2005). Likewise *O. barthii* is well adapted to prevailing environments in Africa with many desirable alleles for improving rice (AfricaRice, 2011). Interspecific hybrid genotypes derived from such crosses could serve as bridge genotypes for gene transfer between species (Ikeda *et al*., 2009). This, however, implies that undesirable traits such as shattering, from wild relatives, could be introgressed into domesticated rice and *vice versa*.

In Africa, it is important to understand the nature and the levels of rice outcrossing under prevailing field and environmental conditions. Such understanding will contribute to formulating strategies to avoid undesirable outcrossing and gene exchange. Knowledge gained from outcrossing studies could be used to recommend temporal mechanisms or to determine minimum isolation distances and buffer zones between varieties or breeding fields to prevent undesirable gene flow.

It is not uncommon to see rice fields in Africa with high levels of admixtures and impurities (Nuijten *et al*., 2009). This could be the result of outcrossing, which is particularly undesirable in the production of quality rice seed. Gene banks, agricultural research institutes and seed institutions have a mandate to maintain and preserve the genetic purity of germplasm in their custody. Maintaining high genetic purity is difficult to achieve when there is undesirable gene flow. Most

practical methods of temporal or spatial isolation are constrained by biological factors and resource availability, respectively (Lu, 2008).

Agro-morphological markers have been extensively used in estimating levels of outcrossing in rice (Da Silva *et al.,* 2005; Endo *et al.,* 2009; Somaratne *et al.,* 2012). In the present study levels of outcrossing in rice were evaluated using agro-morphological markers of red kernel colour and pubescent leaves of the donor against white kernel colour and glabrous leaves of the recipients. Red kernel colour and leaf pubescence have been reported to be dominant (Sweeney *et al.,* 2006; YueHui *et al.,* 2013) and they are easily identifiable on the phenotype.

6.3 Materials and methods

6.3.1 Characterisation trial

Five accessions comprising two intraspecific *O. sativa* Japonica group and three interspecific (*O. glaberrima* x *O. sativa*) hybrids were utilised for the study (Table 6.1). Accessions were chosen based on the presence or absence of dominant traits of leaf pubescence and red kernel colour. This makes it easy to identify hybrids. All plant material was collected from the gene bank of the Africa Rice Center, Cotonou, Benin.

Table 6.1 List of varieties used in the outcrossing study

Variety name	Variety group
WAB96-1-1	Japonica
WAB56-104	Japonica
NERICA 2	Interspecific
NERICA 4	Interspecific
NERICA 7	Interspecific

Prior to selection, the five accessions were amongst 36 that were evaluated for two seasons in field experiments as described in Chapter 3, section 3.3. Selected accessions clustered in different sub-clusters (Figure 5.1). Clustering based on 10 qualitative agro-morphological traits placed NERICA 4 and NERICA 7 in cluster IC; NERICA 2 was in IF and WAB96 was in IE (Figure 5.1). Clustering based on SSR data placed NERICA4 in sub-cluster IA, NERICA7 in sub-cluster IB, WAB56-104 in sub-cluster IE, NERICA 2 in sub-cluster IIB and WAB96-104 in sub-cluster IIA. Initial artificial crosses were made between the donor and each of the recipients to ascertain

compatibility. The F_1 seeds from the different crosses were also planted to determine survival and fertility. They were found to be compatible with viable progeny.

6.3.2 Outcrossing trial

The five accessions used in this study comprised of one *O. sativa* accession (WAB96-1-1) with a red pericarp and pubescent leaves, as the donor and four accessions (with white pericarp and glabrous leaves), comprising one *O. sativa* (WAB56-104) and three interspecific (*O. glaberrima* x *O. sativa*) hybrids namely; NERICA 2, NERICA 4 and NERICA 7 as recipients. Red kernel colour and leaf pubescence are dominant agro-morphological markers (Sweeney *et al.,* 2006; YueHui *et al.,* 2013) and they are easily identifiable on the phenotype. Field experiments were conducted at the Africa Rice Center, Cotonou, Benin. Two experiments were conducted under upland irrigated conditions from August to early November 2013 and from late November 2013 to February 2014.

Figure 6.1 gives a schematic representation of the trial design. A randomised complete block design with three replications was used. The plot size for the pollen donor was 40 m x 40 m and the plot size for the pollen recipient was 3 m x 30 m. The planting method was direct seeding. Three seeds were sown per hole and later thinned to one plant per stand at 14 days after sowing, at a density of 20 cm x 20 cm. The donor cultivar was planted at the prevailing windward side with the receptive cultivars planted downwind.

Using heading data from the characterisation trial of the 36 accessions in Chapter 3, seeding of cultivars was done at different dates in order to synchronise flowering and assure the availability of pollen. In order to prevent pollen flow from outside, an isolation area about 1000 m from any possible pollen source was selected. The site was further surrounded by a physical barrier of maize plants. In addition, there was a distance of 10 m between blocks in order to minimise pollen flow between blocks. Fertilisation, weeding and other cultural practices were done as described in section 3.3.

Entire rows of recipient plants were harvested at distances of 0.2 m, 0.4 m, 0.6 m, 0.8 m, 1.0 m, 1.2 m, 1.4 m, 1.6 m, 1.8 m, 2.0 m, 2.4 m, 2.8 m, 3.0 m 3.2 m, 3.6 m, 4.0 m, 4.8 m, 5.6 m, 6.4 m, 10.0 m, 15.0 m, 20.0 m, 25.0 m and 30.0 m from the donor plot.

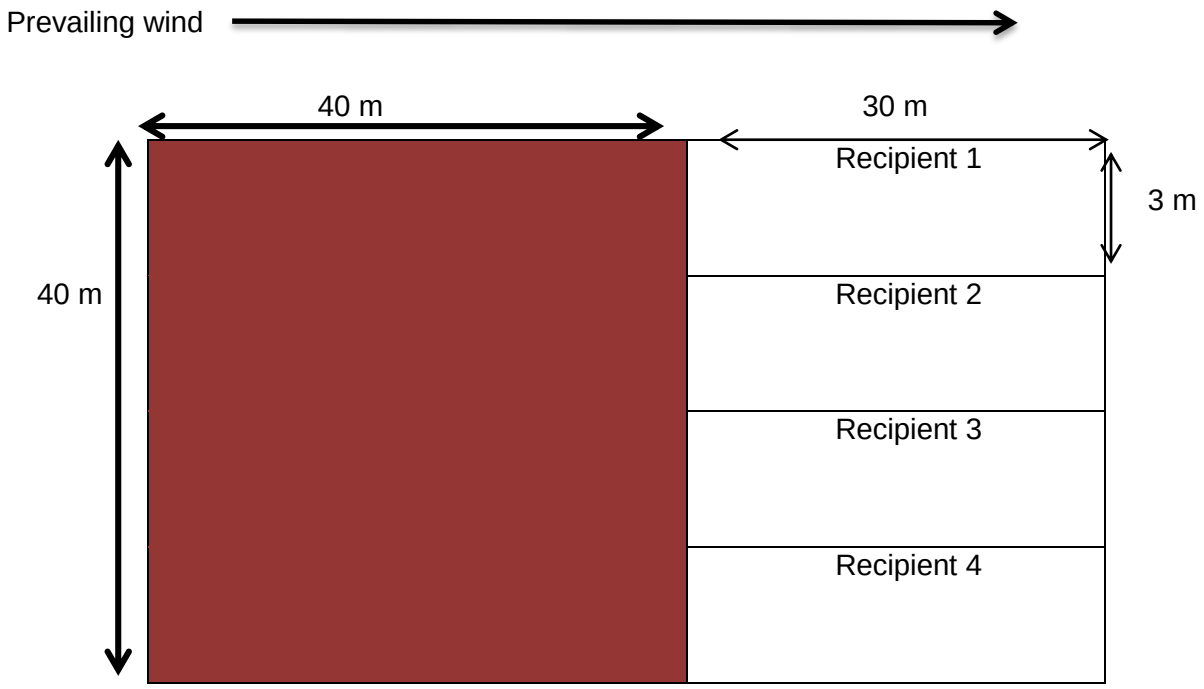


Figure 6.1 Design of field experiment showing pollen donor cultivar on the windward side and receptive cultivars on the leeward side

6.3.3 Progeny testing trial

Four recipient genotypes were used for the progeny testing trial. The location was as described in section 6.3.2. Two experiments were conducted under upland irrigated conditions from January to early April 2014 and from late April to August 2014.

All seeds harvested at a given distance in the outcrossing trial (section 6.3.2) were bulked and planted in a separate plot. Planting method was direct seeding. One seed was sown per hole at a density of 10 cm x 10 cm. Fertiliser application and other cultural practices were as described in section 3.3.

6.4 Data collection

6.4.1 Data collection for characterisation trial

Data collected as described in Chapter 3, section 3.3 was utilised.

6.4.2 Data collection for outcrossing trial

Data on days to heading (initiation, 50% heading and 90% heading) were collected. In addition weather data for wind speed, temperature, relative humidity and rainfall during the heading period was recorded.

6.4.3 Detection of hybrid plants

Figure 6.2 shows the procedure for identification of hybrid plants. Detection of outcrossed plants was done by careful observation of F_1 plants. All plants that produced grains with a red pericarp were then further observed for the presence of leaf pubescence.

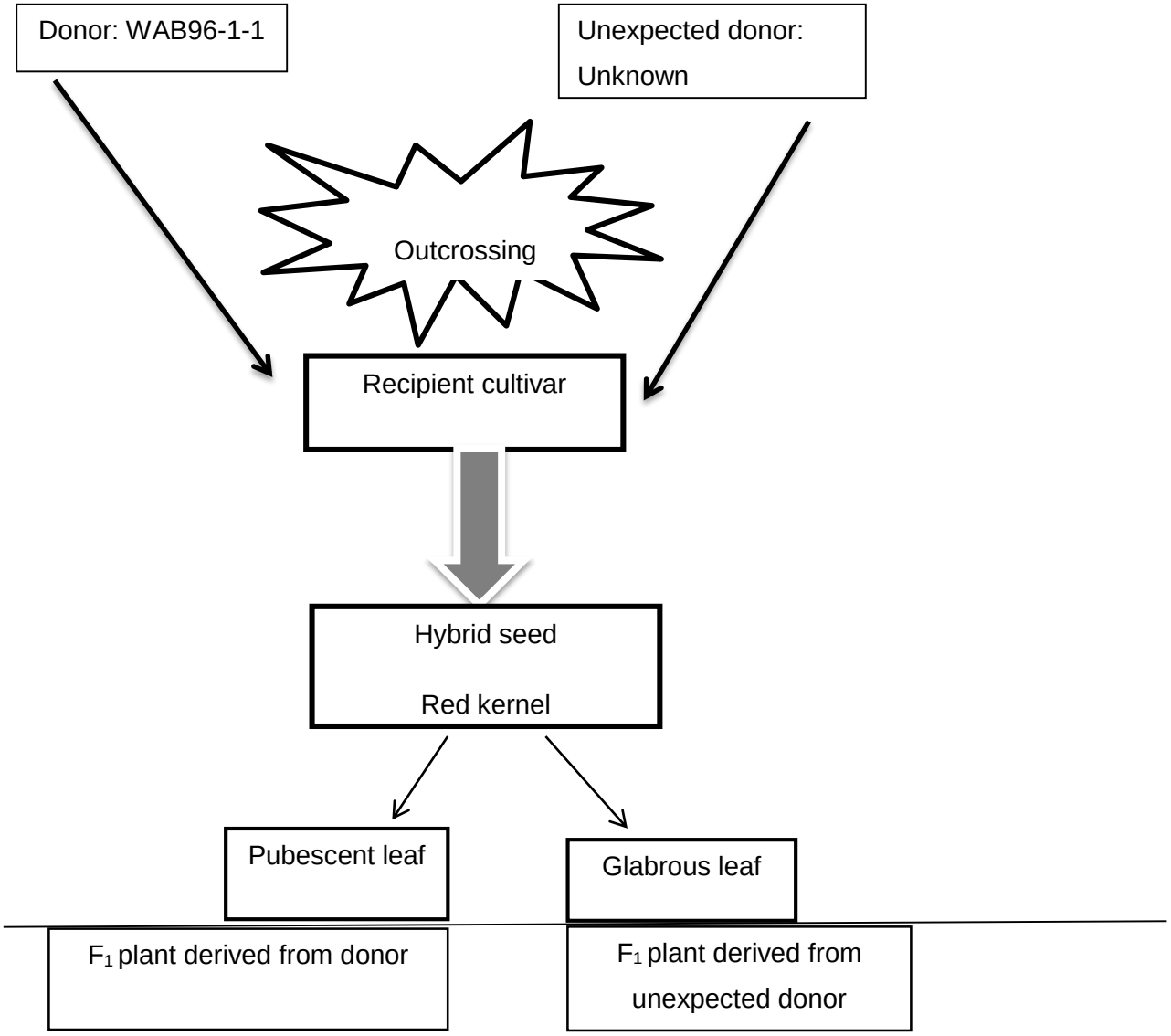


Figure 6.2 Procedure for phenotypic detection of hybrid plants

6.4.4 Data collection and analysis for progeny testing trial

Total plant population for each plot and the number of outcrossed plants, which had the red kernel colour and pubescent leaves, were recorded. Plant height of five representative plants of hybrid and non-hybrid plants was also recorded. XLSTAT Version 3.02 software (Addinsoft, 2008) was used to calculate the correlation coefficient between outcrossing rate and distance.

Outcrossing rate for each treatment was calculated using the following formula:

$$outcrossing\ rate\ (\%) = 100\ x \frac{number\ of\ grains\ \ confirmed\ as\ hybrid\ grain\ from\ intended\ donor\ at\ a\ given\ distance}{total\ number\ of\ grains\ at\ that\ distance}$$

6.5 Results

6.5.1 Characterisation of germplasm

Trait characteristics and summary statistics of donor and recipient germplasm are presented in Table 6.2. Only traits relevant for outcrossing are presented.

Table 6.2 Characteristics of pollen donor and recipients

Variety	Cryopcolor	Lbp	PnEx	PnL	PHt	SOA	SOD	HD	SPKS
WAB96-1-1	Red	Pubescent	41.2	29.0	121.8	29.6	52.3	78.2	155.8
WAB56-104	White	Glabrous	39.5	25.1	105.0	28.1	64.2	62.8	107.6
NERICA 2	White	Glabrous	39.6	26.9	103.6	29.2	47.8	70.3	124.5
NERICA 4	White	Glabrous	39.8	24.9	101.2	29.3	57.3	68.7	119.0
NERICA 7	White	Glabrous	36.8	23.9	123.7	28.8	49.4	68.7	127.7

Cryopcolor = Caryopsis colour; Lbp = Leaf blade pubescence; PnEx = % Panicle exsertion; PnL = Panicle length (cm); PHt = Plant height (cm); SOA = Spikelet opening angle (˚); SOD = Spikelet opening duration (min); HD = Days to 50% heading; SPKS = Spikelet number/panicle

Qualitative traits effectively divided donor and recipient germplasm into distinct groups. The pollen donor had a red kernel colour and pubescent leaves. Pollen recipients had white kernel colour and glabrous leaves. Panicles for the pollen donor and recipients were well exserted and moderately well exserted, respectively. Some of the quantitative traits also divided the germplasm into two distinct classes.

The days to 50% heading divided accessions into medium maturity for the donor and early maturity for the recipients. The number of spikelets/panicle was very good for the donor and good for the recipients (Appendix 2).

Figure 6.3 and Figure 6.4 indicate the general characteristics of hybrid and non-hybrid plants. Hybrid plants defined here as outcrossed plants received pollen from the donor (WAB96-1-1). Non-hybrid plants refer to plants that successfully self-pollinated. Hybrid plants were taller and generally had more tillers than the non-hybrid counterparts. Average number of tillers was 3.3 for hybrids and 3.0 for non-hybrids. Tiller number for hybrid plants ranged from 2.8 (hybrid of NERICA 7) to 3.3 (hybrid of WAB56-104) and for the non-hybrids it ranged from 2.6 (NERICA 7) to 3.2 (WAB56-104). Averages for plant height were 101.4 cm and 93.4 cm for the hybrid and non-hybrid plants respectively. Plant height of hybrid plants ranged from 95 cm (hybrid of NERICA 2) and 109.5 cm (hybrid of NERICA 7) and for non-hybrid plants it ranged from 86 cm (NERICA 2) to 105.3 cm (NERICA 7).

6.5.2 Outcrossing rates under field conditions

Heading initiation date was defined as the date when 10% of all panicles of a cultivar had emerged from the leaf sheath. There were differences in days to heading initiation between the donor and recipient varieties. However, heading and flowering were successfully synchronised between the donor and recipient varieties for the two seasons (Figure 6.5).

In season 1 the difference in heading date ranged from zero (same date as the donor) for NERICA 7 to five days after the donor for NERICA 4. For season 2 the difference ranged from zero for NERICA 4 to four days after the donor for NERICA 2. Though the period of synchronised heading with the donor was generally good, there were differences between varieties (Figure 6.5). In season 1 WAB56-104 initiated heading two days earlier and had an overlap (synchronised) heading of seven out of nine days with the donor. NERICA 2 started one day later than the donor and synchronised heading occurred for six days out of seven with the donor. NERICA 4 started heading five days later than the donor and synchronisation occurred two days out of five with the donor in season 1. For NERICA 7 there was no difference in heading initiation and there was a synchronisation of seven days out of nine.

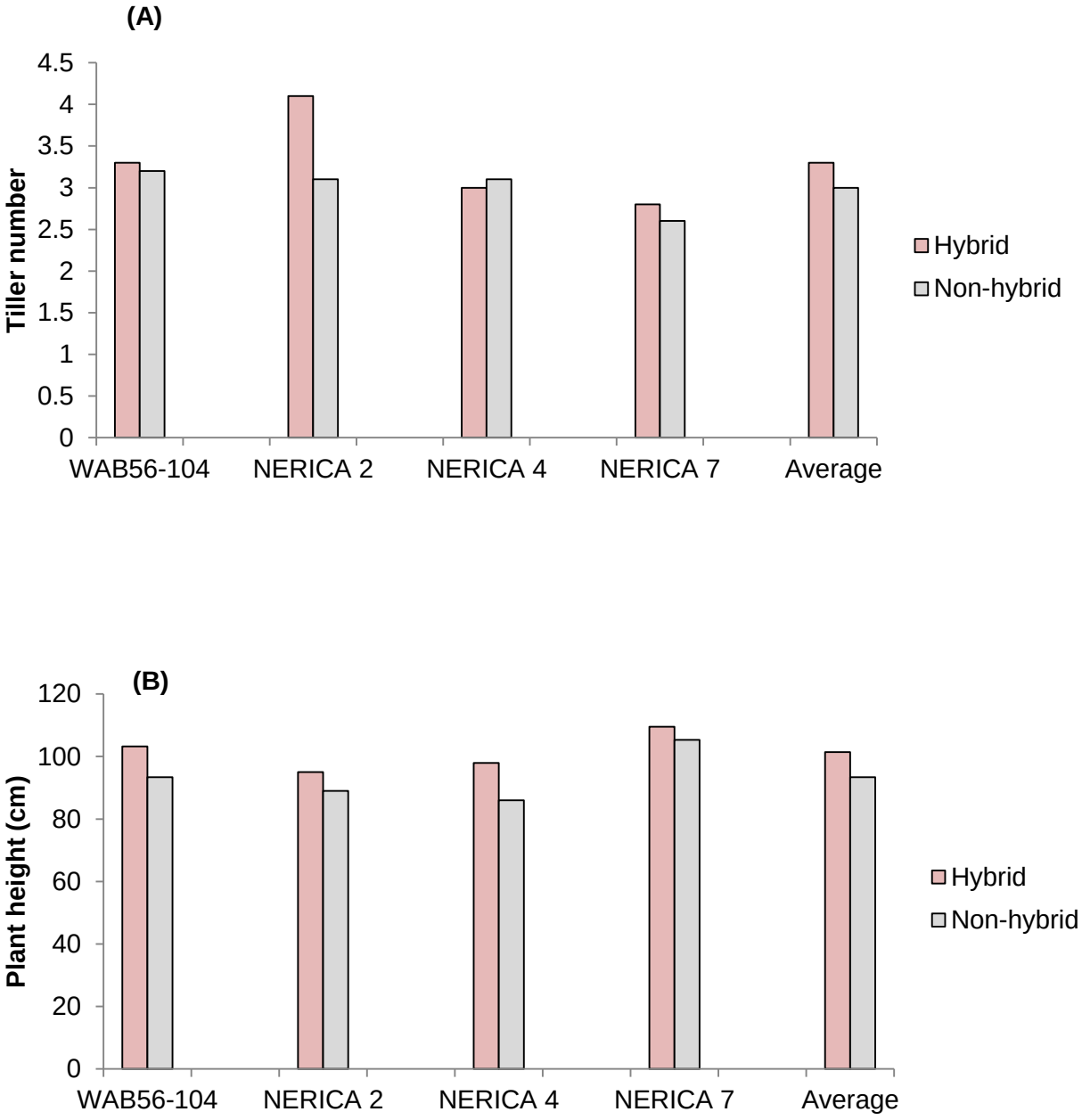


Figure 6.3 Tiller number (A) and plant height (B) of hybrid and non-hybrid plants

Figure 6.4 Hybrid plants in the midst of non-hybrid plants

In season 2, WAB56-104 initiated three days after the donor and synchronised four days out of nine with the donor. NERICA 2 initiated heading three days after the donor and had an overlap of three days out of 11 with the donor. For NERICA 4 there was no difference in heading initiation and there was a synchronisation of seven days out of eight. Heading initiation for NERICA 7 was two days after the donor and there was a synchronisation of five days out of nine with the donor in season 2.

Endo *et al.* (2009) detected outcrossing between different rice cultivars when the difference in heading date was not more than 11 days. In this study, shorter than 11 days heading date differences were observed. The largest difference was five days in season 1 (NERICA4) and four days in season 2 (NERICA2). The heading period was longer (15 days) in season 2 than in season 1 (12 days). However, the donor maintained the same length of seven days for heading period in both seasons. Similarly recipients WAB56-104 and NERICA 7 had a heading period of nine days in each of the seasons. Only NERICA 2 and NERICA 4 showed differences between the two seasons in heading period. For NERICA 2 it was seven days in season 1 and 11 days in season 2. For NERICA 4 the heading period lasted five days in season 1 and eight days in season 2.

Table 6.3 shows the total plant population that was evaluated in the progeny tests. Total number of plants evaluated was 721 134. The lowest for WAB56-104 was 1581 plants at 0.2 m from the donor, whilst the highest total was 15 009 plants 25 m from the donor. For NERICA 4 the lowest population total of 2 330 plants was observed at 30 m from the donor whilst the highest total of 9413 plants was observed at 10 m from the donor.

For NERICA 7 the lowest population total of 3751 plants was observed at 2 m from the donor whilst the highest population count of 18630 was observed at 1.2 m from the donor. For NERICA 2 the lowest population count of 1427 plants was observed at 0.2 m from the donor whilst the highest the population count of 15364 plants was observed at 15 m from the donor.

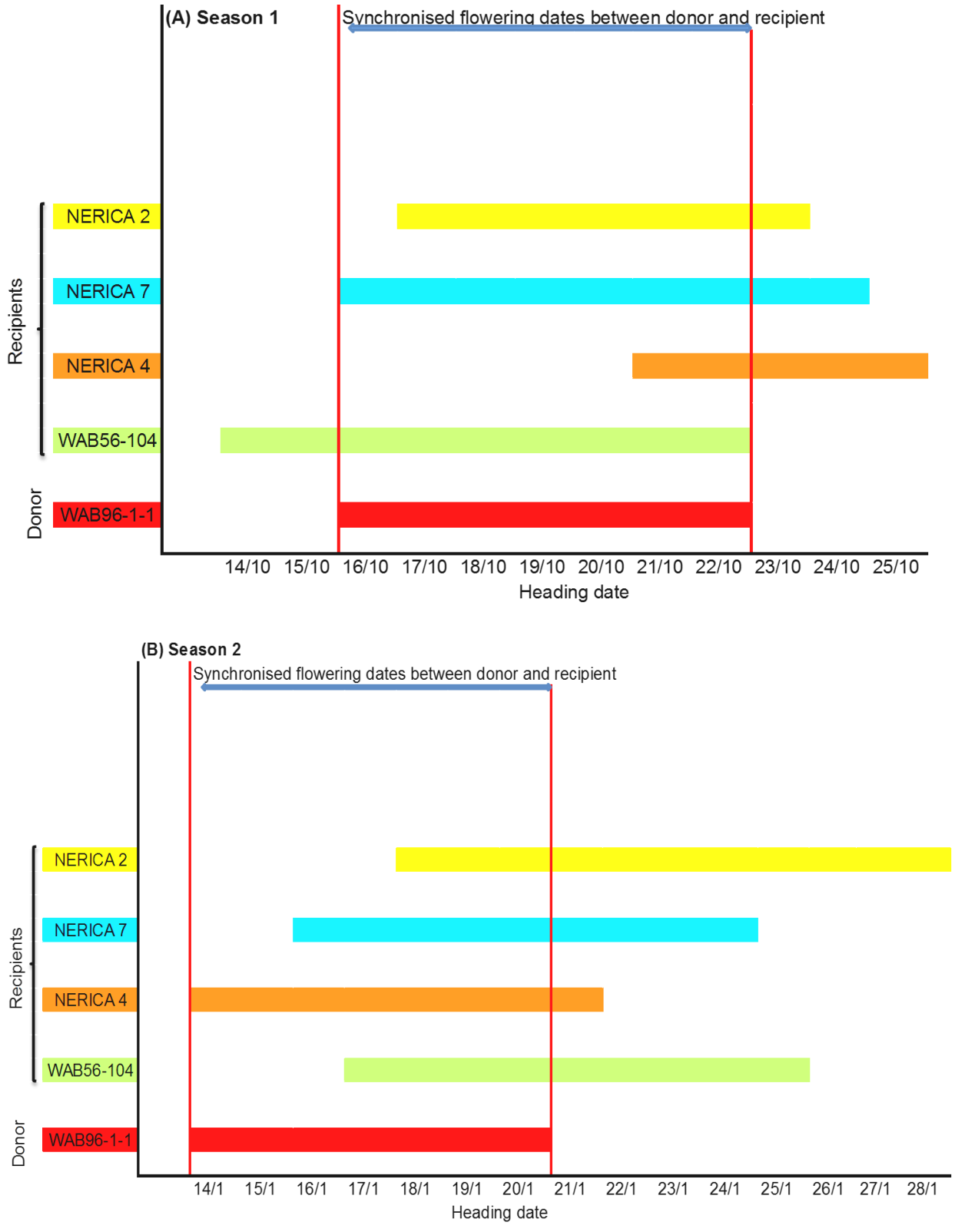


Figure 6.5 Heading initiation dates and period of synchronised heading between donor and recipients. A = Season 1; B = Season 2

Table 6.3 Total plant population for different distances and cultivars used in progeny tests

Distance (m)	WAB56-104	NERICA 4	NERICA 7	NERICA 2	Distance total
0.2	1581	2654	6320	1427	11982
0.4	3659	2844	7741	4668	18912
0.6	6385	4357	10454	5955	27151
0.8	10551	7940	10438	4863	33792
1.0	6824	7912	11325	5665	31726
1.2	10114	3682	18630	9906	42332
1.4	8097	7084	7299	7060	29540
1.6	7015	4833	10845	5043	27736
1.8	6187	3670	12139	9422	31418
2.0	3907	5563	3751	7886	21107
2.4	8190	6288	12051	7297	33826
2.8	7818	4845	12432	6276	31371
3.0	4897	7174	9887	5109	27067
3.2	6140	5594	12451	5994	30179
3.6	6099	5480	11484	8059	31122
4.0	8068	5697	6691	4326	24782
4.8	11496	5992	14550	4127	36165
5.6	8941	7828	9426	4827	31022
6.4	8512	7168	10299	7084	33063
10	8066	9413	10366	6629	34474
15	10246	9304	8457	15364	43371
20	6415	7579	8262	5513	27769
25	15009	6336	12665	8790	42800
30	4014	2330	8297	3786	18427
Variety total	178231	141567	246260	155076	721134

Outcrossing rates between cultivars for both seasons combined are presented in Table 6.4. There was an average outcrossing rate of 0.7% which varied from 0.05% at 25 m to 2.45% at 0.2 m. In general outcrossing rates decreased with increasing distance from the donor. WAB56-104 and NERICA 4 had similar average outcrossing rates of 0.9% $\pm$ 0.52 and 0.9% $\pm$ 0.70 respectively.

NERICA 2 was different from other NERICA varieties as very low outcrossing rates were observed. The lowest average outcrossing rate of 0.2% was for NERICA 2.

Table 6.4 Outcrossing rates over two seasons in relation to distance

Distance (m)	Outcrossing (%)				
	WAB56-104	NERICA 4	NERICA 7	NERICA 2	Average
0.2	2.45	3.16	2.95	1.24	2.45
0.4	1.33	1.94	1.62	0.43	1.33
0.6	1.29	1.75	1.11	0.10	1.07
0.8	1.10	1.18	1.07	0.14	0.87
1	1.42	1.33	1.27	0.13	1.04
1.2	1.33	0.76	0.82	0.24	0.79
1.4	0.52	1.65	1.15	0.16	0.87
1.6	0.79	1.15	0.46	0.06	0.62
1.8	0.93	1.13	0.50	0.25	0.70
2	0.86	0.56	0.79	0.75	0.74
2.4	0.54	0.93	0.46	0.14	0.52
2.8	0.97	0.82	0.79	0.03	0.65
3	0.64	0.88	0.76	0.05	0.58
3.2	1.15	1.05	0.55	0.06	0.70
3.6	1.33	0.82	0.46	0.08	0.67
4	1.07	0.37	0.46	0.01	0.48
4.8	0.39	0.87	0.43	0.04	0.43
5.6	0.49	0.33	0.34	0.04	0.30
6.4	0.52	0.61	0.29	0.01	0.36
10	0.51	0.21	0.07	0.02	0.20
15	0.42	0.16	0.09	0.02	0.17
20	0.17	0.16	0.05	0.00	0.10
25	0.11	0.03	0.03	0.04	0.05
30	0.15	0.11	0.03	0.01	0.08
Average	0.9 $\pm$ 0.52	0.9 $\pm$0.70	0.7 $\pm$ 0.63	0.2 $\pm$ 0.28	0.7 $\pm$ 0.51

Figure 6.6 shows total outcrossing rates for each recipient cultivar for the two seasons. The highest outcrossing rate in season 1 was observed for WAB56-104 with a value of 1.2% and the lowest was for NERICA 2 with a value of 0.1%.

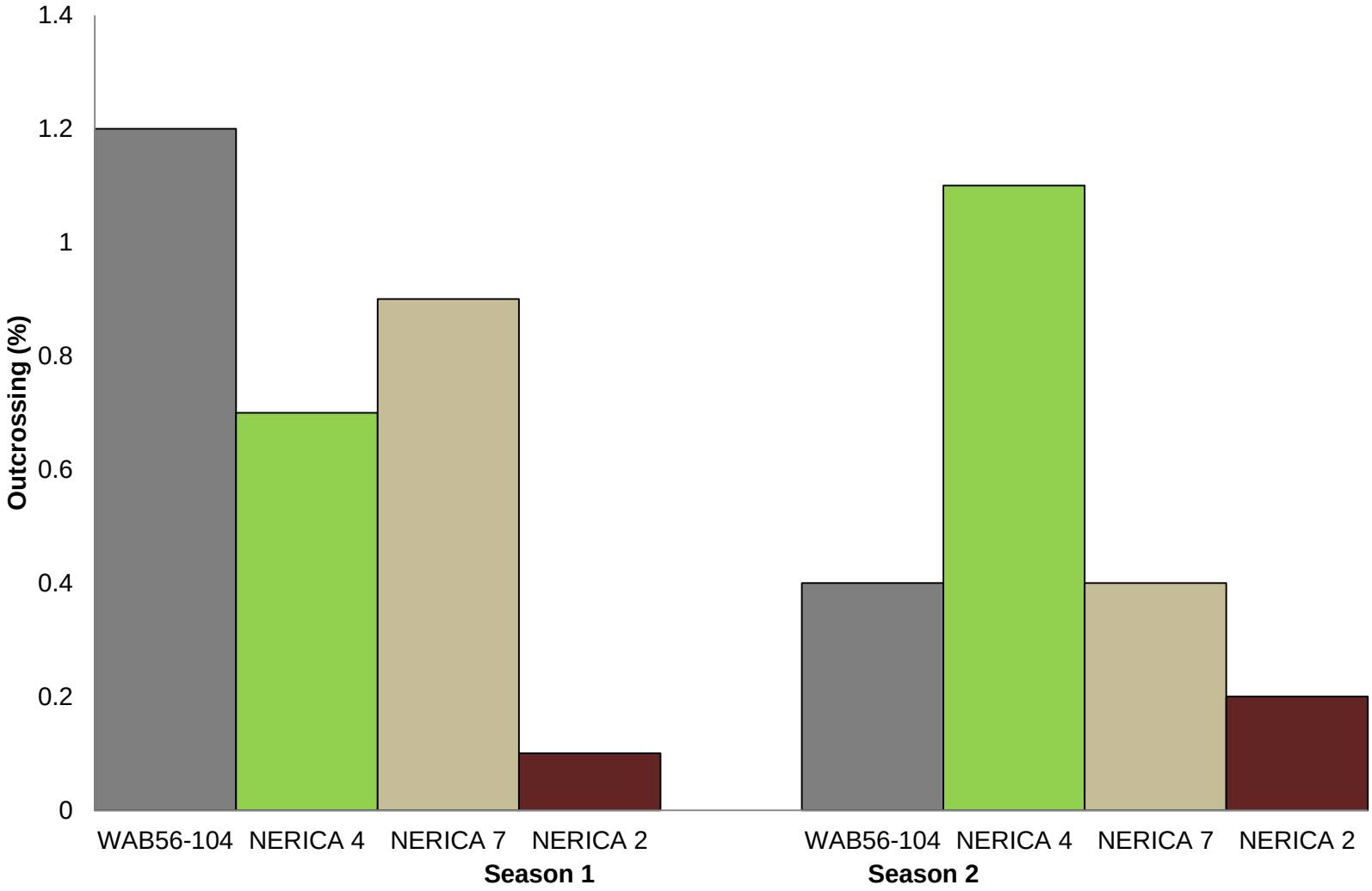


Figure 6.6 Overall outcrossing rates of accessions on a plot basis

For NERICA 4 and NERICA 7 observed outcrossing rates were 0.7% and 0.9% respectively. In season 2 the highest outcrossing rate was observed for NERICA 4 with a value of 1.1% and the lowest was for NERICA 2 with a value of 0.2%. Outcrossing rates in season 2 for NERICA 7 and WAB56-104 were 0.4% each. Overall, outcrossing rates were higher for NERICA 2 and NERICA 4 in season 2, and lower for WAB56-64 and NERICA 7. NERICA 2 showed the lowest outcrossing rate with the donor in each of the seasons. The accessions WAB56-104 and NERICA 4 that had highest outcrossing rates in season 1 and season 2 respectively, also had high levels of synchronised flowering with the donor.

Outcrossing rates for each season at specific distances from the pollen donor are presented in Figure 6.7. There were differences between cultivars in levels of outcrossing within and between seasons. In season 1 (Figure 6.7A) the highest outcrossing rate of 3.9% was observed for NERICA 4 at a distance of 0.2 m from the pollen donor and the lowest was by NERICA 2 with zero outcrossing at 20.0 m from the pollen donor. Outcrossing rates for WAB56-104 ranged from 0.14% at 25.0 m to 2.2% at 3.6 m from the donor. In general, outcrossing rates for WAB56-104 were relatively high (most values ≈ 2%) for up to 4.0 m from the donor. Outcrossing rates for NERICA 4 ranged from 0.04% at 30.0 m to 3.9% at 0.2 m from the donor. NERICA 7 had relatively high levels of outcrossing for up to 1.4 m from the pollen donor. Generally, outcrossing rates were less than or equal to 0.6% at and beyond 6.4 m from the donor. For NERICA 2 the levels of outcrossing by the donor were the lowest, ranging from zero to 0.47%.

In season 2 (Figure 6.7B) the highest outcrossing rate of 3.3% was observed for WAB56-104 at a distance of 0.2 m from the pollen donor and the lowest was for NERICA 2 with zero outcrossing. Outcrossing rates for WAB56-104 ranged from 0.08% at 25 m to 3.3% at 0.2 m from the donor. Outcrossing rates for NERICA 4 ranged from zero to 2.4% at 25 m and 0.2 m from the donor. Outcrossing rate for NERICA 7 ranged from 0.02% at 20.0 m to 3.02% at 0.2 m from the pollen donor. For NERICA 2 the levels of outcrossing by the donor ranged from zero to 2.04%.

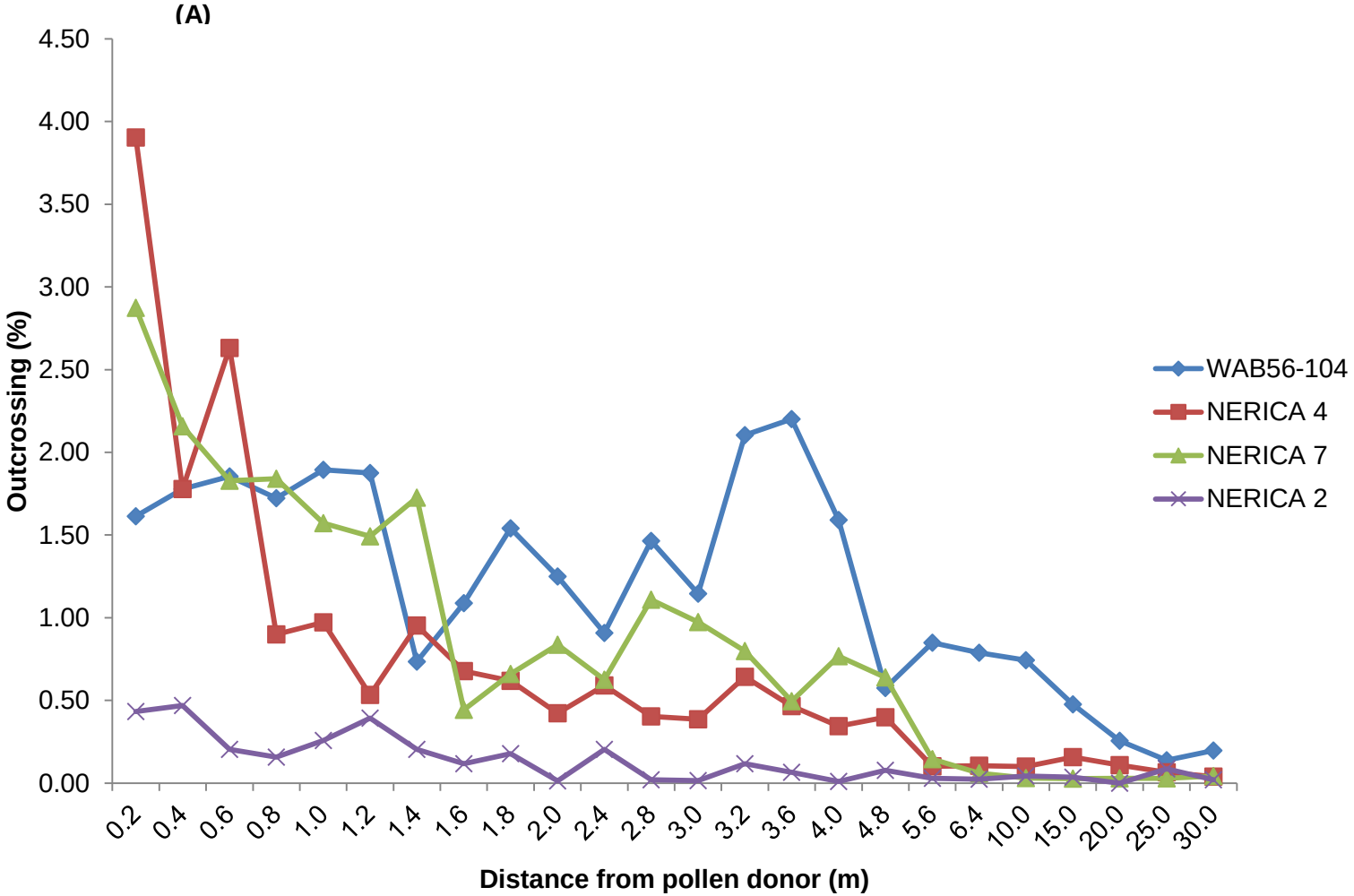


Figure 6.7A Outcrossing in relation to distance for season 1

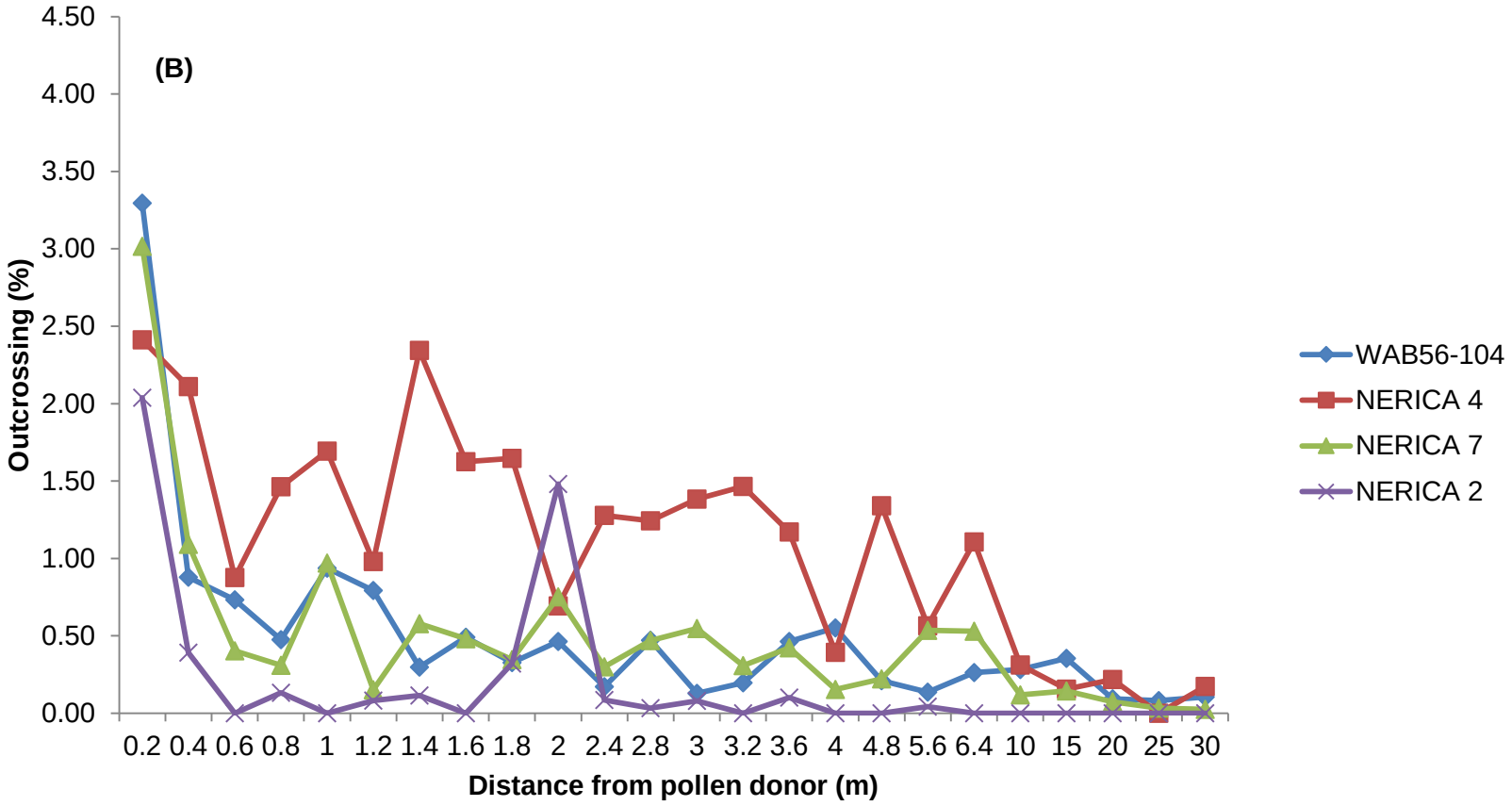


Figure 6.7B Outcrossing in relation to distance for season 2

Figure 6.8 shows average outcrossing rates at specific distances for both seasons. In general outcrossing rates tended to drop with increasing distance from the donor (Figure 6.8A). Outcrossing rates dropped dramatically up to 1.6 m from the donor. There was a gradual drop between 1.6 m and 20 m which stabilised beyond 20 m. However, there were varietal differences. For WAB56-104 outcrossing rates dropped dramatically up to 1 m from the donor (Figure 6.8B). There was a gradual drop between 1.2 m and 4.8 m which stabilised at distances beyond 20 m from the donor. For NERICA 4 outcrossing rates dropped dramatically up to 1.2 m from the donor (Figure 6.8C). There was a gradual drop between 1.2 m and 4 m which stabilised at 10 m from the donor. For NERICA 7 outcrossing rates dropped dramatically up to 0.8 m from the donor (Figure 6.8D). There was a gradual drop between 1 m and 6.4 m which stabilised from 10 m from the donor. For NERICA 2 outcrossing rates dropped dramatically up to 0.6 m and then stabilised beyond that. The relationship between distance and levels of outcrossing was determined using the Pearson correlation matrix. There was a negative correlation between distance and outcrossing rates. The correlation coefficient was -0.63 at $p \leq 0.001$. This implies that outcrossing decreased as distance from the pollen source increased. The coefficient of determination (R^2) for distance and outcrossing was 0.396 at $p \leq 0.001$.

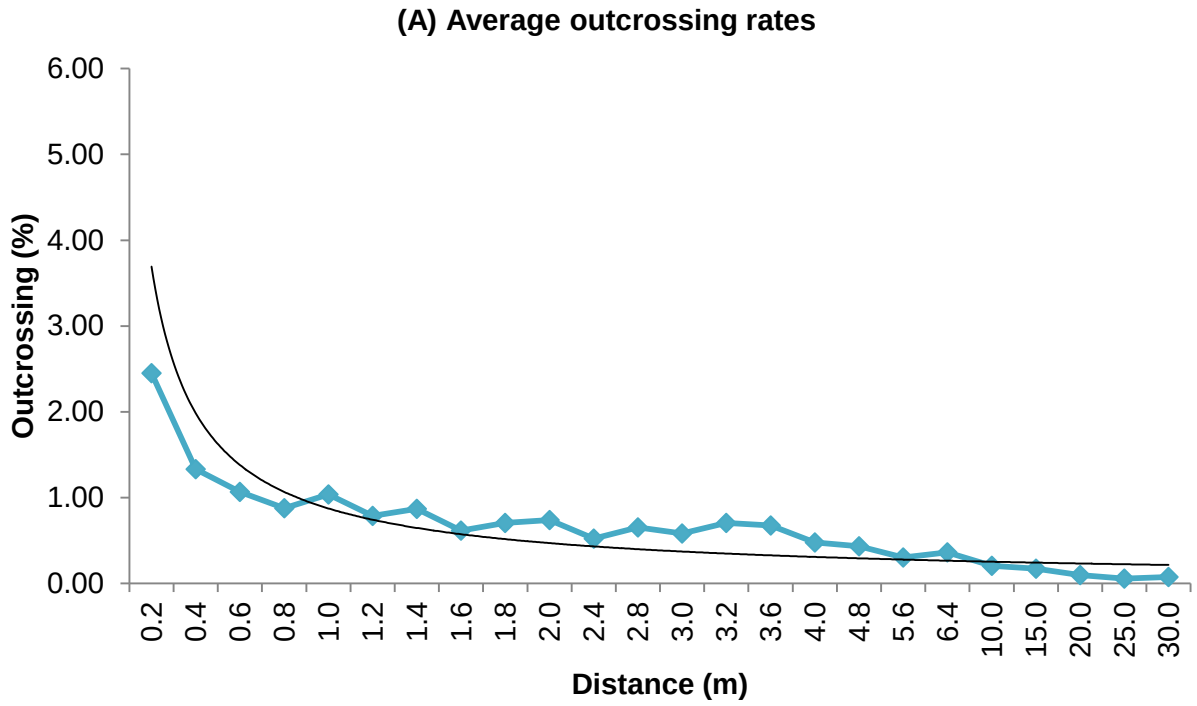


Figure 6.8A Outcrossing rates (blue squares) irrespective of variety with trend line (smooth line) in relation to distance for both seasons

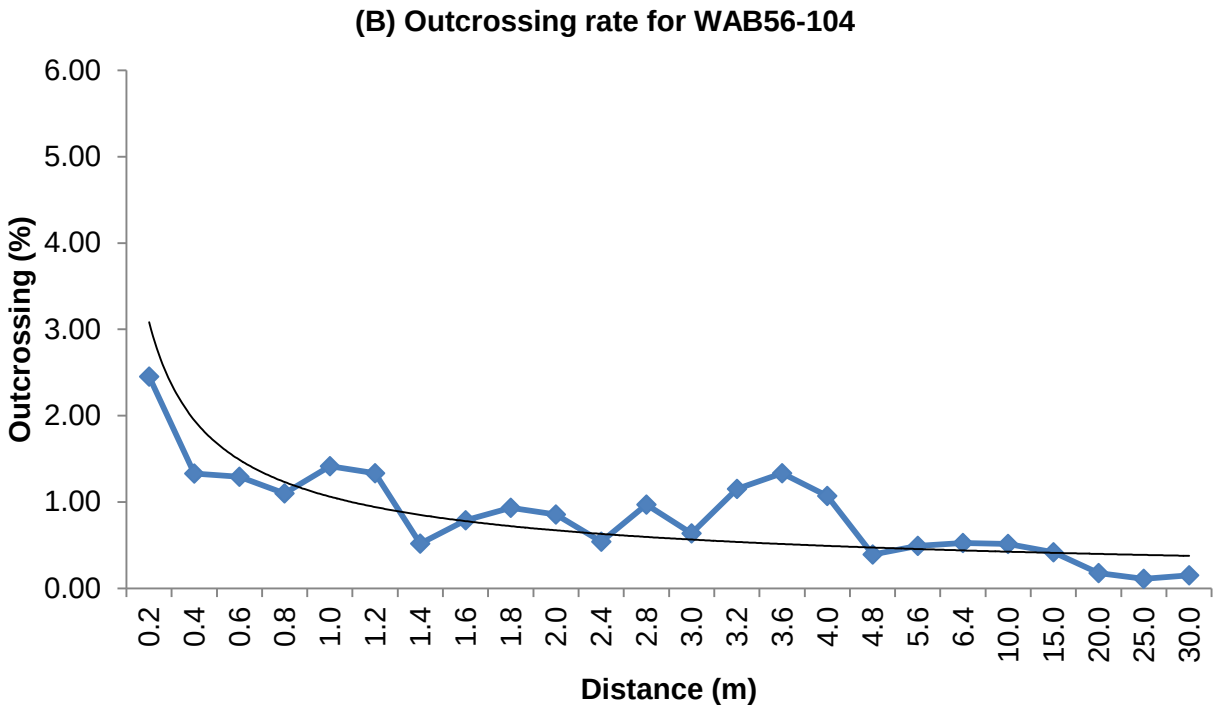


Figure 6.8B Outcrossing rates (square-line) with trend line (smooth line) in relation to distance for both seasons for WAB56-104

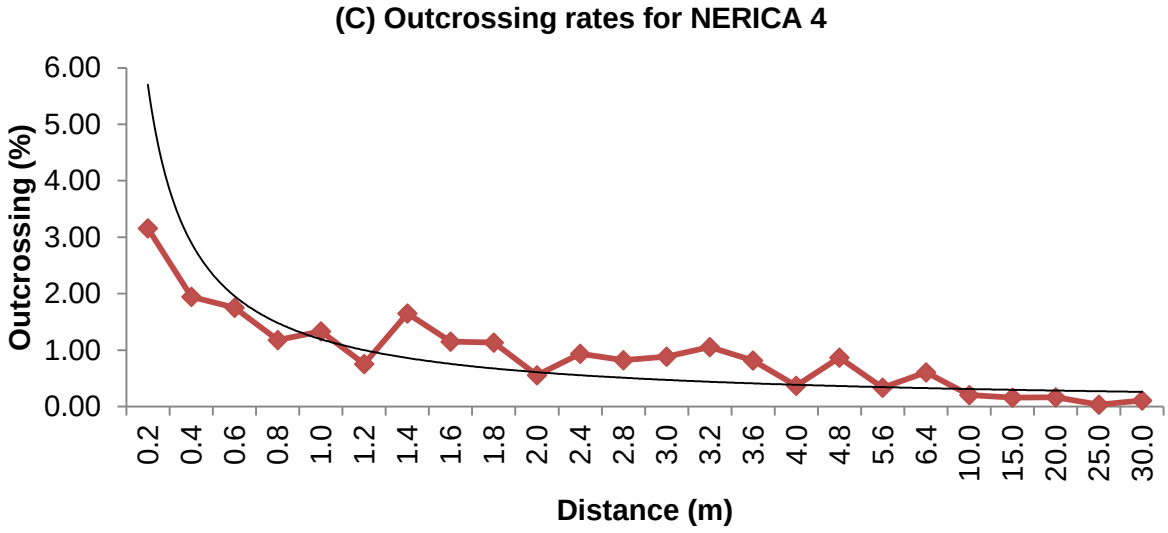


Figure 6.8C Outcrossing rates (square-line) with trend line (smooth line) in relation to distance for both seasons for NERICA 4

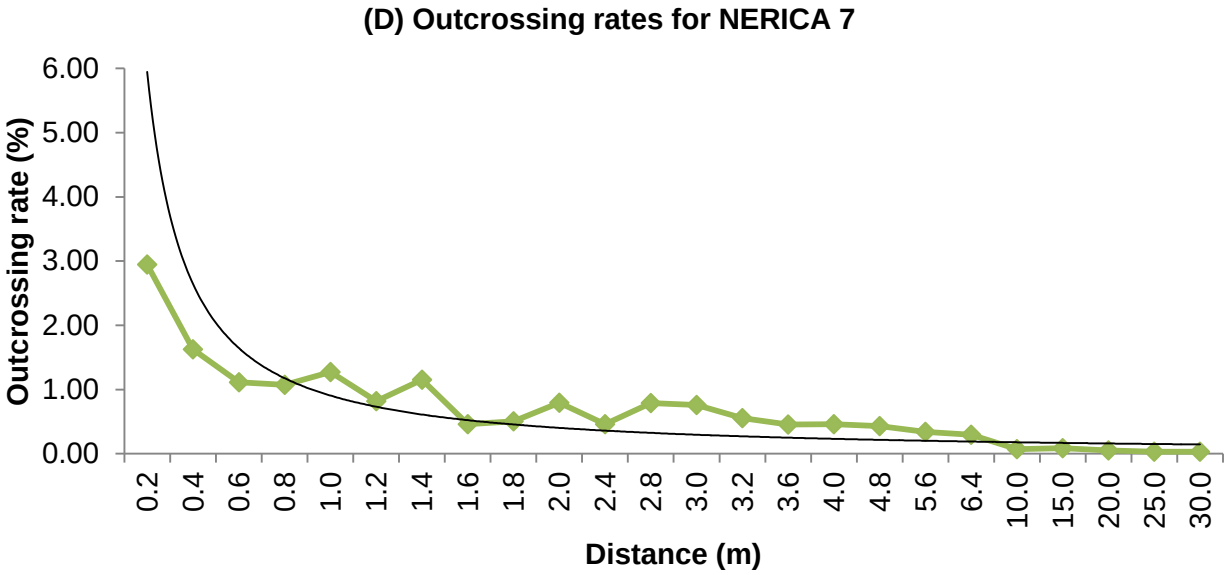


Figure 6.8D Outcrossing rates (square-line) with trend line (smooth line) in relation to distance for both seasons for NERICA 7

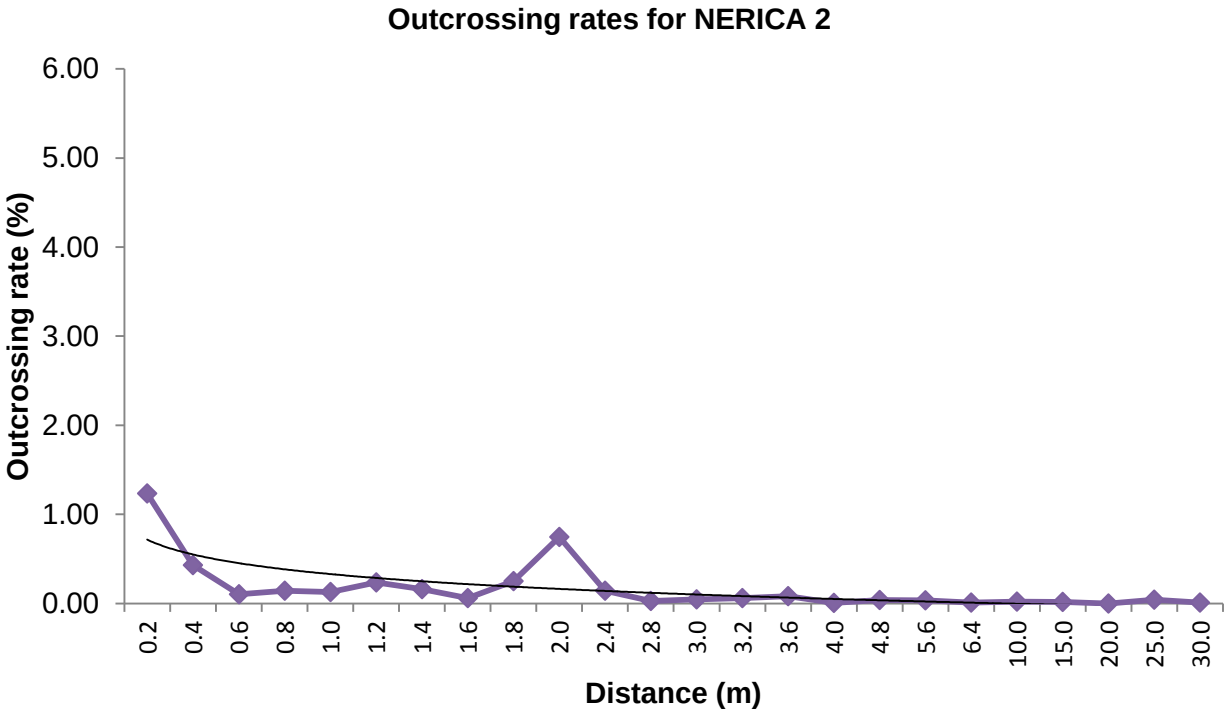


Figure 6.8E Outcrossing rates (square-line) with trend line (smooth line) in relation to distance for both seasons for NERICA 2

6.5.3 Weather conditions during the flowering period

Weather conditions (temperature, relative humidity and wind speed) during the flowering period of cultivars in the two seasons are presented in Figure 6.9 and Appendix 4. The flowering period for season 1 was from October 14–25, 2013 and for season 2 it was from January 14–28, 2014. Temperatures were relatively stable with an average of 28ºC for each season. Relative humidity for season 1 ranged from 75–97% with an average humidity of 83%. Relative humidity was lower in season 2 ranging from 30–71% with an average of 56%. Relative humidity was also more stable in season 1. In season 2 the relative humidity progressively dropped with the length of flowering period showing a dramatic reduction towards the end of the flowering period. Wind speed was similar for the two seasons.

6.6 Discussion

6.6.1 Outcrossing rates

Several studies have reported a relationship between distance and outcrossing rate or pollen dispersal in rice (Messeguer *et al*., 2004; Rong *et al*., 2004; 2007; Song *et al*., 2004; Endo *et al*., 2009; Kanya *et al*., 2009). However, other factors have been reported to also influence outcrossing in rice. Gealy *et al.* (2003) reported that temperature, radiation and humidity influenced the behaviour of floret opening and pollen survival and therefore also influenced outcrossing. A relationship between weather conditions and outcrossing has been reported in several other studies (Matsui *et al.*, 2000; Messeguer *et al.*, 2004; Jagadish *et al.*, 2007). In this study, weather conditions were similar for the two seasons under evaluation and could not have greatly influenced outcrossing. Floral characteristics have also been reported to influence levels of outcrossing in rice (Matsui and Kagata, 2003). This study confirmed the relationship between distance and outcrossing levels.

Outcrossing rates between the donor WAB96-1-1 and recipient WAB56-104 were generally high. This could be due to a high level of flowering compatibility between the two cultivars. For instance there was a high level of flowering synchronisation between them. Endo *et al.* (2009) detected outcrossing between different rice cultivars when the difference in heading date was not more than 11 days. Similar and relative high outcrossing rates were observed for interspecific rice cultivars NERICA 4 and NERICA 7. This could also be due to high flowering synchronisation between them and the donor.

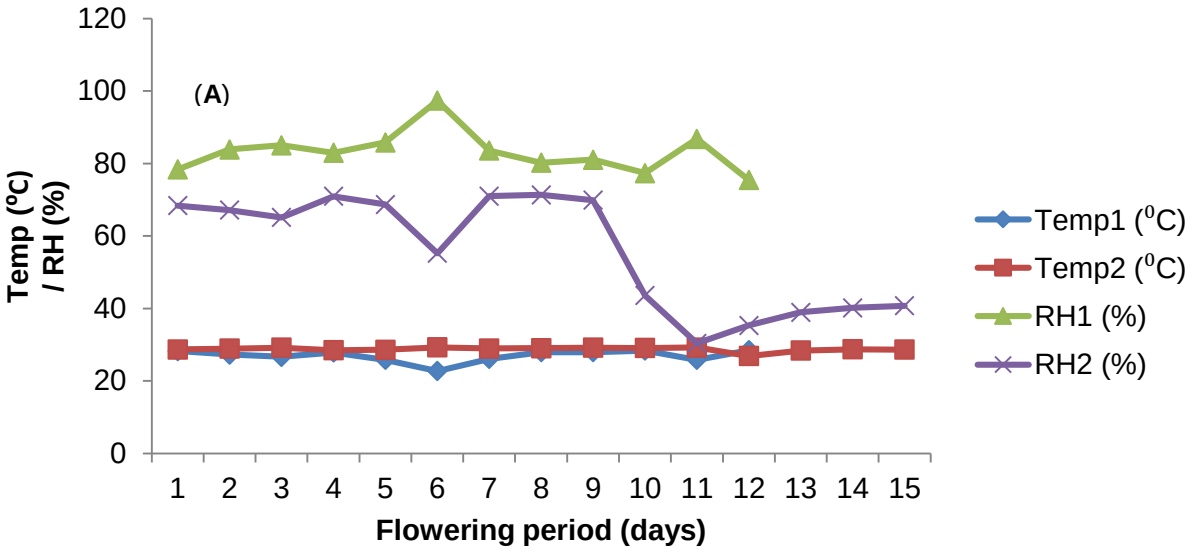


Figure 6.9A Temperature and relative humidity during the flowering period in season 1 (14-25 October, 2013) and season 2 (14-28 January, 2014)

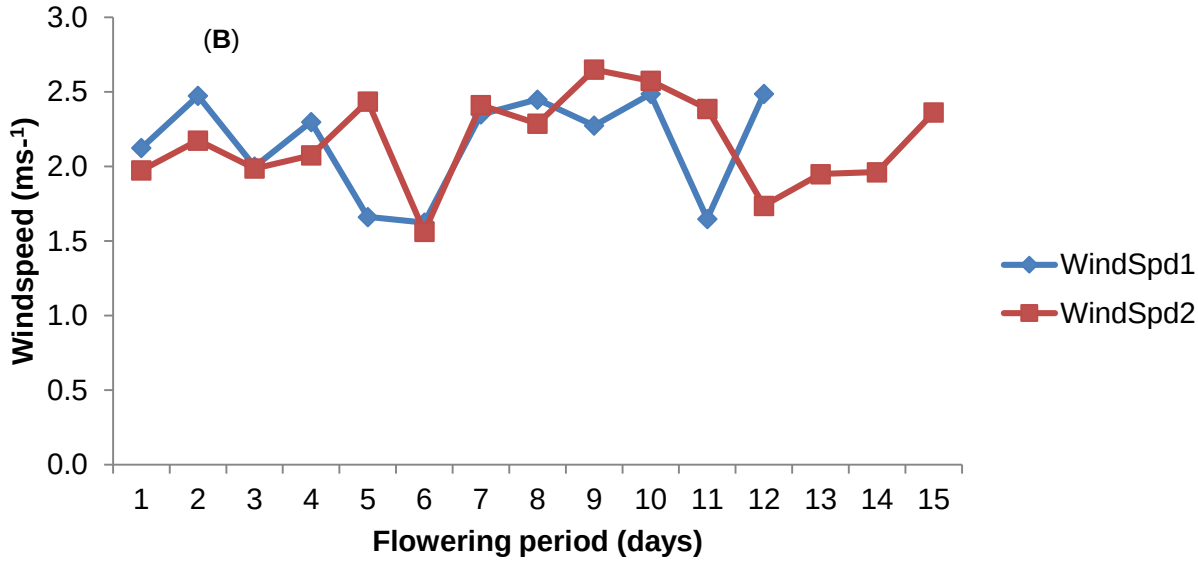


Figure 6.9B Wind speed during the flowering period in season 1 (14-25 October, 2013) and season 2 (14-28 January, 2014)

Temp1 = Temperature for season 1, Temp2 = Temperature for season 2, RH1 = Relative humidity for season 1, RH2 = Relative humidity for season 2, WindSpd1 = Wind speed for season 1, WindSpd2 = Wind speed for season 2

The low outcrossing rate observed for NERICA 2 might be due to both genetic and morphological factors. Clustering and the similarity matrix based on SSR data indicated that NERICA 2 in comparison to other recipient accessions was the most closely related to WAB56-104. Studies have indicated outcrossing levels are higher between distantly related genotypes than closely related ones (Messeguer *et al*., 2001; 2004; Song *et al*., 2003; Chen *et al*., 2004). The presence of awns in NERICA 2 might lead to low outcrossing. The presence of awns in closed panicles has been reported to disturb the free movement of anthers and stigmas on flowering spikelets, which results in reduction of outcrossing rates (Ishii *et al*., 2013). Morphological and genetic differences were also observed between NERICA 2 on the one hand and NERICA 4 and NERICA 7 on the other. NERICA 2 is awned whilst NERICA 4 and NERICA 7 are not. Clustering based agro-morphological traits placed NERICA 4, NERICA 7, WAB56 and NERICA 2 in separate sub-clusters. However, NERICA 2 was further apart. The dendrogram based on SSR data clustered NERICA 2 separately from NERICA 4, NERICA 7 and WAB56-104.

It is difficult to compare outcrossing rates of different studies as they are influenced by the type of genotypes used, the environment in which the study was done and GxE interactions. This study was unique as it looked at outcrossing between cultivars of domesticated conventional rice including interspecific (*O. sativa* x *O. glaberrima*) hybrids. Accessions were mainly improved cultivars for the upland ecology. Most studies on rice outcrossing investigated at crop-to-wild outcrossing. Others that involved crop-to-crop gene flow mainly investigated transgene-to-crop gene flow. However, results from this study are generally in line with previous studies on outcrossing in rice. Messeguer *et al*. (2004) reported outcrossing rates of 0.036% and 0.086% for red rice and conventional rice, respectively. Gealy *et al. (*2003) reported natural outcrossing rates to be less than 1%. Langevin (1990) reported rates of 1–52%. In agreement with the present study, other studies have also reported levels of outcrossing that varies depending on type of genotypes used and the environment in which studies where done. In addition, outcrossing rates reduce as the distance between the recipient and the pollen donor increases (Gealy *et al.,* 2003; Messeguer *et al*., 2004; Hoyle and Cresswell, 2007; Endo *et al*., 2009).

6.6.2 Implications of gene flow on germplasm conservation and management

Outcrossing occurs as a result of gene flow between genotypes. Whilst not all gene flow is manifested in the phenotype, agro-morphological markers still remain invaluable in assessing levels of gene flow between genotypes. This study did not look at gene flow between cultivated rice *O. sativa* and the weedy rice (red rice) or wild relatives (like *O. barthii)*. Implications of results

are, however, equally true for such genotypes. Hybridisation due to outcrossing invariably leads to incorporation of new genes into the gene pool of the recipient population from the source population. This could lead to significant evolutionary changes in the recipient population. The nature of the impact of the change depends on whether gene flow was neutral, detrimental or beneficial to the recipient population (Ellstrand *et al.*, 1999). *Oryza glaberrima* has several alleles, many useful, that can be used for rice improvement (Jones *et al*., 1997; Sarla and Swamy, 2005; AfricaRice, 2011) but which is constrained by the sterility barrier of F_1 hybrids derived from crosses involving *O. glaberrima* and other rice species (Ikeda *et al*., 2009). As observed in the current study, there was gene flow between interspecific hybrid rice and *O. sativa* Japonica group. Therefore genotypes of interspecific rice could serve as a bridge for the transfer and exchange of desirable alleles between *O. glaberrima* and *O. sativa*.

Gene banks, in their routine operations, carry out regeneration of germplasm. It is not uncommon to see large numbers of different populations being grown in close proximity (at a spacing of less than 1 m between different genotypes). Populations may include cultivated species and their wild relatives. When populations of related genotypes are grown in close proximity, gene flow is likely to occur between them. When this happens it eventually leads to the alteration of gene pools of recipient populations. Such a situation defeats the purpose of germplasm conservation where the primary objective is to maintain the genetic makeup of accessions or populations being regenerated. In some cases the recipient germplasm progressively assumes the characters of the donor. For instance in Asia, hybridisation with cultivated rice was cited as being responsible for the near extinction of the taxon *O. rufipogon* ssp*. formosana* (Ellstrand *et al*., 1999).

The world has recently witnessed the increased use of transgenes (ISAAA, 2012). Herbicide tolerant rice has been introduced to counter the serious constraint of weed infestation in rice cultivation. A possible crop-to-weed gene flow will result in the target weeds developing tolerance to herbicides. This will result in the evolution of more aggressive weeds with associated practical and economic consequences. Hybrids have been found to be fitter and more aggressive than either of the parents. Langevin (1990) reported hybrids that were taller with longer and wider flag leaves and with more tillers than the red rice parents. Lee *et al.* (2013) also reported that off-type rice plants were taller with longer leaves than cultivated rice. Similarly, in the present study, observed hybrids were taller and had more tillers than the non-hybrids. Such aggressiveness of hybrids could be an adaptive strategy for survival.

Therefore to achieve the desired results, germplasm conservation practices and the use of transgenic crops should be mindful of the possibilities and consequences of gene flow. Appropriate measures derived from experimentation that mimic natural conditions on spatial and temporal isolation could help reduce the negative consequences.

6.6.3 Implications of gene flow on quality seed production

In the current study, outcrossing occurred up to 30 m from the donor. Results indicated relatively high levels of outcrossing for distances of up to 4.0 m from the donor. Some level of outcrossing took place up to 20 m from the donor. It is not uncommon to see rice fields in close proximity in sub-Saharan Africa. In addition, rice is largely cultivated on land bordering waterways and inland valleys. Cultivated rice therefore shares the same ecological niche with wild and other weedy relatives. Such situations are conducive for gene flow between the different genotypes. Farmers in sub-Saharan Africa also use the same rice seed over long periods, usually selecting seed for the next season from the previous crop. Thus even when the initial seed stock met the acceptable certification standards, the number of outcrossed plants is bound to progressively increase with time. Farmers would select seed from vigorous plants from within the plot. Such plants could be transgressive segregants resulting from outcrossing. In explaining the possible scenarios for interspecific hybrid development in farmers' fields in West Africa, Nuijten *et al.* (2009) indicated that fertile hybrids could develop within two backcross generations.

Certified seed programmes recommend a minimum isolation distance of 10 m for breeder seed (and about 3 m for other seed classes) between a rice crop for seed production and sources of possible pollen contamination. Results in this study showed that outcrossing occurred at longer distances. Whilst the levels at longer distances might be low, they should be interpreted within the context of the impact that such levels would have on seed purity in later generations of seed or crop production. The amount of outcrossed plants could increase substantially over the next generations, invariably resulting in off-types which are undesirable in seed production. A field of rice is mainly composed of homozygotes and the level of homozygosity is reduced with the presence of off-types. This eventually leads to reduction of the genetic purity of the seed stock.

Breeders at research institutes observe much smaller distances (less than the standard 3 m) between different populations when evaluating them in breeding blocks and field experiments. This would result in undesirable gene flow and outcrossing during the evaluation of breeding lines, which can delay or complicate the fixing of such materials for release as cultivars. Contamination by undesired pollen should be avoided in conventional plant breeding, especially for hybrid rice.

Segregating material cannot meet the standards for compulsory DUS testing necessary for variety release. Material that is not fixed before release will segregate in subsequent generations in farmers' fields.

In the current study there were differences in levels of outcrossing between genotypes. It is only at distances beyond 20 m that levels of outcrossing fell to very low values and stabilised.

6.7 Conclusions

Agro-morphological markers of red kernel colour and leaf pilosity were effective in assessing outcrossing rates between the donor and recipient varieties. In general, agro-morphological markers are valuable tools in plant diversity studies. They combine simplicity and ease of scoring. They have been used in several studies to estimate levels of outcrossing in rice (Da Silva *et al*., 2005; Endo *et al*., 2009; Somaratne *et al.,* 2012). The ideal marker is dominant and visibly expressed on the phenotype.

Outcrossing or gene flow *per se* is not a hazard as much as it is an evolutionary process (Lu, 2008). However, with the exploitation of novel techniques and the introduction of hitherto unlikely gene combinations in rice cultivars, the risks associated with undesirable outcrossing would escalate. Field experiments remain a valuable tool for assessing these risks and proposing containment strategies however, they do not always sufficiently mimic future reality prior to widespread adoption of such cultivars or technologies (Wolfenbarger and Phifer, 2000). Outcrossing is prevalent in rice and associated containment measures should be put in place to minimise undesirable gene flow. Outcrossing could have a negative impact on the maintenance of genetic purity in germplasm conservation and the production of good quality seed due to exchange and introgression of genes from the donor. In addition, gene flow between domesticated rice and wild relatives could lead to weeds of increased fitness as some alleles in domesticated rice may confer a selective advantage in the wild.

Pollen flow also reduces with increased distance (Song *et al*., 2004; Endo *et al*., 2009; Kanya *et al*., 2009). In another study, Devos et al. (2006) reported the effectiveness of pollen barriers as deterrent to pollen flow and outcrossing. Equally farmers that continually use seed for long periods should be made aware of the degeneration of genetic purity with time. In this regard the rate and nature of genetic purity degeneration over time should be investigated. It is common practice for many farmers to recycle rice seed for many generations before renewal. ResuIts could be used

to advise farmers and other stakeholders as to how long seed stocks should be used before renewal.

6.8 Recommendation

Results of this study indicated that distance is an important factor influencing outcrossing in rice. However, there are other factors that also significantly influence rice outcrossing. Matsui and Kagata (2003) reported a relationship between floral characteristics and outcrossing. Weather conditions (Gealy *et al.*, 2003; Song *et al.,* 2004) and field orientation (Hoyle and Cresswell, 2007) have also been indicated as important factors for outcrossing. It is important that factors that were not covered in this study be investigated in order to have a more holistic understanding of outcrossing rates of the cultivars. Such knowledge could be a useful guide for maintaining genetic purity in germplasm conservation and in the production of quality rice seed. It could be used to guide policy makers in the deployment of new cultivars with novel gene combinations that could have a potential impact on the environment. Physical isolation like minimum isolation distance (based on more meaningful experimental data) as indicated by this study and also by the study of Rong *et al.* (2007), could be effective barriers to outcrossing. Pollen flow also reduces with increased distance (Song *et al.,* 2004; Endo *et al.,* 2009; Kanya *et al.,* 2009). In another study Devos *et al.* (2006) reported the effectiveness of pollen barriers as deterrent to pollen flow and outcrossing.

Equally, farmers that continually use seed for long periods should be made aware of the degeneration of genetic purity with time. In this regard the rate and nature of genetic purity degeneration over time should be investigated. The results could be used to advise farmers and other stakeholders as to how long seed stocks should be used before renewal.

6.9 References

Addinsoft. 2008. XLSTAT Version 3.02. http://www.xlstat.com/en/. (Accessed 18/07/2015)

AfricaRice (Africa Rice Center). 2011. Boosting Africa's Rice Sector: A Research for Development Strategy 2011–2020. Cotonou, Benin: ii + 77 pp.

Brookes G, Barfoot P. 2012. Global impact of biotech crops: Environmental effects, 1996–2010. GM Crops and Food: Biotechnology in Agriculture and the Food Chain 3:129–137.

Chen H, Lin YJ, Zhang QF. 2009. Review and prospect of transgenic rice research. Chinese Science Bulletin 54:4049–4068.

Chen LJ, Lee DS, Song ZP, Suh HS, Lu BR. 2004. Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Annals of Botany 93:67–73.

Da Silva EF, Da Silva LM, Montalvan R. 2005. Crossing rate and distance in upland rice. Bragantia, Campinas 64:197–201.

Devos Y, Reheul D, Schrijver AD. 2006. Considerations of cross-fertilization between GM and non-GM maize. ISB News Report March 2006. http://www.isb.vt.edu/articles/mar0603.htm. (Accessed 15/06/2014)

Ellstrand NC. 2003. Current knowledge of gene flow in plants: implications for transgene flow. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 358:1163–1170.

Ellstrand NC, Prentice HC, Hancock JFE. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539–563.

Endo T, Sato H, Yamaguchi M, Kataoka T, Nakagomi K, Ito T, Mori K. 2009. Estimate of outcrossing rates in a rice plant (*Oryza sativa* L.) under field conditions using a purple grain rice cultivar, *Okunomurasaki*. Breeding Science 59:195–202.

FAOSTAT 2013. http://faostat3.fao.org/faostat-gateway/go/to/download/C/CC/E. (Accessed 11/03/2014)

Gealy D, Mitten D, Rutger N. 2003. Gene flow between red rice and herbicide resistant rice: Implications for weed management. Weed Technology 17:627–645.

Hoyle M, Cresswell, JE. 2007. The effect of wind direction on cross-pollination in wind-pollinated GM crops. Ecological Applications 17:1234–1243.

Ikeda R, Sokei Y, Akintayo I. 2009. Seed fertility of F_1 hybrids between upland rice NERICA cultivars and *Oryza sativa* L. or *O. glaberrima* Steud. Breeding Science 59:27–35.

IRGSP (International Rice Genome Sequencing Projcet). 2005. The map-based sequence of the rice genome. Nature 436:793–800.

ISAAA 2012. ISAAA Brief 44. Executive Summary: Global Status of Commercialized Biotech/GM Crops. http://www.isaaa.org/resources/publications/briefs/44/executivesummary/ (Accessed 17/05/2015)

ISAAA 2014. ISAAA Brief 49. Executive Summary: Global Status of Commercialized Biotech/GM Crops. http://www.isaaa.org/resources/publications/briefs/44/executivesummary/ (Accessed 17/05/2015)

Ishii T, Numaguchi I, Miura K, Yoshida K, Thanh PT, Htun TM, Yamasaki M, Komeda N, Matsumoto T, Terauchi R, Ishikawa R, Ashikari M. 2013. *OsLG1* regulates a closed panicle trait in domesticated rice. Nature Genetics 45:462–465.

Jagadish SVK, Craufurd PQ, Wheeler TR. 2007. High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). Journal of Experimental Botany 58:1627–1635.

Jones MP, Dingkuhn M, Aluko GK, Semon M. 1997. Interspecific *Oryza sativa* L x *O. glaberrima* Steud. progenies in upland rice improvement. Euphytica 92:237–246.

Kanya JI, Kinyamario JI, Amugune NO, Hauser TP. 2009. Dispersal distance of rice (*Oryza sativa* L.) pollen at the Tana River delta in the coast province, Kenya. African Journal of Biotechnology 8:2265–2270.

Langevin SA. 1990. The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa* L.). Evolution 44:1000–1008.

Lee J, Park JH, Koh HJ. 2013. Morphological and genetic characterization of off-type rice plants collected from farm fields in Korea. Journal of Plant Biology 56:160–167.

Lu BR. 2008. Transgene escape from GM crops and potential biosafety consequences: An environmental perspective. Collection of Biosafety Reviews 4:66–141.

Lu BR, Snow AA. 2005. Gene flow from genetically modified rice and its environmental consequences. BioScience 55:669–678.

Matsui T, Kagata H. 2003. Characteristics of floral organs related to reliable self-pollination in rice (*Oryza sativa* L.). Annals of Botany 91:473–477.

Matsui T, Omasa K, Horie T. 2000. High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (*Oryza sativa* L.). Plant Production Science 3:430–434.

Messeguer J, Fogher C, Guiderdoni E, Marfa V, Catala MM, Baldi G, Mele E. 2001. Field assessments of gene flow from transgenic to cultivated rice (*Oryza sativa* L.) using a herbicide resistance gene as trace marker. Theoretical and Applied Genetics 103:1151–1159.

Messeguer J, Marfà V, Català MM, Guiderdoni E, Melé E. 2004. A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. Molecular Breeding 13:103–112.

Neigel JE. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology and Systematics 28:105–128.

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P. 2009. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers' fields. PLoS ONE 4:1-9.

Phan PDT, Kageyama H, Ishikawa R, Ishii T. 2012. Estimation of the outcrossing rate for annual Asian wild rice under field conditions. Breeding Science 62:256–262.

Rong J, Lu BR, Song Z, Su J, Snow AA, Zhang X, Sun S, Chen R, Wang F. 2006. Dramatic reduction of crop-to-crop gene flow within a short distance from transgenic rice fields. New Phytologist 173:346–353.

Rong J, Xia H, Zhu Y, Wang Y, Lu BR. 2004. Asymmetric gene flow between traditional and hybrid rice varieties (*Oryza sativa*) indicated by nuclear simple sequence repeats and implications for germplasm conservation. New Phytologist 163:439–445.

Sarla N, Swamy BPM. 2005. *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. Current Science 89:955–963.

Shivrain VK, Burgos NR, Sales MA, Mauromoustakos A, Gealy DR, Smith KL, Black, HW, Jia M. 2009. Factors affecting the outcrossing rate between Clearfield™ rice and red rice (*Oryza sativa*) Weed Science 57:394–403.

Somaratne LHMYK, Abayawickrama ASMT, Wickramasinghe IP, Samarasinghe WLG. 2012. Estimating Out-Crossing Rate of Bg 379-2 Using morphological markers and confirmation by molecular markers. Rice Science 19:166–168.

Song Z, Lu BR, Chen J. 2004. Pollen flow of cultivated rice measured under experimental conditions. Biodiversity and Conservation 13:579–590.

Song ZP, Lu BR, Zhu YG, Chen JK. 2003. Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657–665.

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. 2006. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. The Plant Cell 18:283–294.

Wolfenbarger LL, Phifer PR. 2000. The ecological risks and benefits of genetically engineered plants. Science 290:2088–2093.

YueHui Z, YongSheng Z, Ling L, HongGuang X, JianFu Z, HuaAn X. 2013. Genetic analysis and fine mapping of the pubescence gene *GL6* in rice (*Oryza sativa* L.). Chinese Science Bulletin 58:2992–2999.

Chapter 7

General conclusions and recommendations

7.1 Conclusions

Rice is a major food crop that serves as a source for food and livelihood for about 3.5 billion people around the world. There are two cultivated species namely *O. sativa* and *O. glaberrima* for which there is ongoing debate about the domestication history. One model proposes a single domestication event whilst another proposes multiple domestication events.

As a model for research, diversity studies have been carried out on the crop. Rice diversity studies use different types of markers including agro-morphological and molecular markers. Both marker types are effective in evaluating diversity in rice germplasm. Information from diversity studies is vital for planning outcrossing trials and the maintenance and conservation of germplasm, identifying parents for breeding purposes and planning crosses. The choice of marker types depends on the objectives of the study and available resources. Agro-morphological markers combine simplicity and rapidity, whilst molecular markers like microsatellites are more robust.

In this study agro-morphological and microsatellite markers were used separately and in combination to assess the level of diversity in a collection of 36 rice accessions. Accessions comprised *O. sativa*, *O. glaberrima* and interspecific hybrids (*O. sativa* x *O. glaberrima*). Using clustering and principal component analysis the collection was divided into two major clusters. However, the composition of clusters differed between markers and clustering methods.

Rice is a self-pollinating plant with some amount of cross pollination occurring under natural conditions. Outcrossing itself is part of evolution and is not a hazard as such. However, with the introduction of novel combinations, that may have an impact on the environment, there is the need for caution. Outcrossing leads to gene flow from the donor to the recipient. The impact of gene flow depends on its nature, whether it is neutral, beneficial or deleterious. In rice, outcrossing occurs between domesticated rice and its wild relatives; it also occurs between domesticated rice and weedy rice (red rice), and crop-to-crop. Most studies on crop-to-crop gene flow involved transgenic rice where various levels of outcrossing have been reported. There has been little research on outcrossing of interspecific hybrid rice. The present study has confirmed that there is crop-to-crop gene flow in domesticated rice involving modern improved cultivars of intraspecific (*O. sativa* and *O. glaberrima*) and interspecific hybrids (*O. sativa* x *O. glaberrima*). Outcrossing in rice has implications for germplasm conservation and maintenance. It also has implications for

managing crosses and field evaluation trials in rice breeding programmes. Equally important is the impact it has on the production of quality rice seed. All of these activities should take cognisance of outcrossing and make use of mitigating measures that have been derived from experimental data that mimic natural conditions.

7.2 Recommendations

Africa, precisely the Niger River Basin, is the centre of diversity of *O. glaberrima,* one of the two species of cultivated rice. This species, with its associated wild relatives, are highly adapted to prevailing conditions in Africa. However, this potential is yet to be exploited, as most of the rice coverage in Africa is with *O. sativa*. There is the need to further study African rice and its related wild and weedy species to tap into its potential for increased rice productivity. Outcrossing in rice occurs within and between species under field conditions. It is important to know the nature and levels of outcrossing that occurs in Africa. The number of *O. glaberrima* accessions used in the study was too small to be representative. Therefore future studies could use a larger sample size for *O. glaberrima* accessions.

The use of microsatellites and associated applications of marker-assisted selection (MAS) are invaluable as a breeding tool. They are practical and efficient for simple traits that are difficult to score. Many economically important traits like yield are complex and establishing MAS systems may be expensive and cumbersome. Therefore diversity evaluation should use molecular techniques in combination with conventional approaches of agro-morphological traits.

Outcrossing as a natural phenomenon occurs widely in the plant kingdom. This study investigated outcrossing in modern improved cultivars of domesticated rice. There were various levels of outcrossing that were influenced by the genotype, the distance from the donor, and flowering synchronisation between donor and recipient genotypes. Whilst the study indicated that outcrossing has an impact on the maintenance of genetic purity, it did not look at the nature of this change. It would be useful to design a study that will look at the degeneration or change in purity levels of rice at different generations (seed classes) of rice. Future studies could also look at the nature of gene flow between domesticated rice and wild or weedy relatives of rice under prevailing conditions in Africa. There is little understanding of outcrossing rates of *O. glaberrima.* A study designed to look at the nature of outcrossing between *O. glaberrima* and other rice species would contribute largely to the understanding of rice as a self-polling outcrossing plant.

Morphological markers effectively estimated outcrossing in rice in this study. However, with the associated drawback of morphological markers being incapable of detecting neutral variation a future study on outcrossing could use a combination of morphological and molecular markers. The latter is more robust and can be used to reliably confirm hybrids.

The study showed outcrossing to occur up to 30 m from the donor. A genetic purity level of about 99.9% is considered adequate for breeder seed and nuclear or parental seed stock. In order to achieve that level of purity based on findings of the current study, at least 20 m of isolation distance between different genotypes is required.

Summary

Keywords: Agro-morphological, diversity, interspecific rice, microsatellites, *O. glaberrima*, *O. sativa,* outcrossing

Rice is an important food crop with two domesticated species namely *Oryza glaberrima* Steud. also known as African rice and *O. sativa* L. also known as Asian rice. Rice is a self-pollinating plant that shows some level of outcrossing under field conditions. Understanding the levels of gene flow is important for managing the conservation and maintenance of germplasm for gene banks and plant breeding programmes. It is equally important for the production of quality rice seed. There is also the concern of gene flow between genetically modified rice and other rice species; wild relatives and weedy (red) rice. This study used agro-morphological and microsatellite markers to evaluate the diversity in a collection of 36 accessions consisting of intraspecific (*O. sativa* and *O. glaberrima*) and interspecific (*O. sativa* x *O. glaberrima*) hybrid genotypes.

The Shannon-Weaver diversity index was low measured by agro-morphological markers. The highest level of diversity was detected in the Indica accessions and the lowest diversity in the *O. glaberrima* accessions. Japonica accessions had brown or green apiculus and were awnless. *O. glaberrima* and landrace accessions had purple apiculus and some were awned. The Japonica group comprised mostly of improved varieties with a white pericarp whilst *O. glaberrima* accessions and landraces had a red pericarp. Selected interspecific hybrids combined traits of *O. sativa* and *O. glaberrima* to varying degrees. Microsatellite markers showed a total of 321 alleles with an average of 11.9 alleles per locus and an average major allele frequency of 0.29 per locus. The average gene diversity value was 0.81 and polymorphism information content was 0.80 per locus. Diversity indices for interspecific hybrids were intermediate between *O. sativa* and *O. glaberrima*, but closer to the *O. sativa* Japonica group. Two populations were revealed which corresponded to the *O. sativa* Indica group and *O. sativa* Japonica group. Interspecific hybrid accessions were dispersed between the two groups. Microsatellite data detected higher diversity between accessions in comparison to agro-morphological data. There was correlation between the different matrices, indicating that they all reflected similar patterns in the structure of diversity of the collection.

Agro-morphological markers of red kernel colour and leaf pilosity of the pollen donor was also used to estimate outcrossing rates in four genotypes of rice consisting of improved cultivars of *O. sativa* Japonica group (one accession) and interspecific hybrids (three accessions). The dominant

markers of red kernel colour and leaf pilosity were effective in estimating outcrossing in rice. There was an average outcrossing rate of 0.7% ± 0.51, with a potential outcrossing rate of 2.45% ± 0.86. Outcrossing rates decreased with increase in distance. It ranged from 2.45% at 0.2 m from the donor to 0.05% at 25 m from the donor. Gene flow was influenced by type of genotype, the distance of recipient from the pollen donor and flowering synchronisation. The study suggests that natural outcrossing could be partly responsible for the appearance of non-uniform and off-type plants in rice fields in sub-Saharan Africa. The study recommends a minimum distance of 20 m between genotypes to minimise outcrossing to acceptable levels.

Opsomming

Sleutelwoorde: Agro-morfologies, diversiteit, interspesifieke rys, mikrosatelliete, rys, *O. sativa, O. glaberrima*, uitkruising

Rys is ‘n belangrike voedselgewas met twee aangeplante spesies naamlik *Oryza glaberrima* Steud., ook bekend as Afrika rys en *O. sativa* L., ook bekend as Asiese rys. Rys is ‘n selfbestuiwende gewas met ‘n persentasie kruisbestuiwing onder veldtoestande. ‘n Begrip van die hoeveelheid geenvloei is belangrik vir die bestuur en onderhoud van kiemplasma vir geenbanke en planteteelt programme. Dit is net so belangrik vir die produksie van kwaliteit ryssaad. Daar is ook kommer oor geenvloei tussen GM rys en ander rys spesies; wilde verwantes en onkruidagtige (rooi) rys. Hierdie studie het agro-morfologiese en mikrosatelliet merkers gebruik om diversiteit in ‘n versameling van 36 inskrywings bestaande uit intraspesifieke (*O. sativa* L. en *O. glaberrima* Steud.) en interspesifieke (*O. sativa* x *O. glaberrima*) rys genotipes te bepaal. Resultate van kwalitatiewe en kwantitatiewe eienskappe het verskille in diversiteit tussen die inskrywings getoon. Die waargenome diversiteit in *O. glaberrima* was laer as in *O. sativa*. Interspesifieke rys het intermediêre vlakke van diversiteit getoon.

Die Shannon-Weaver diversiteitsindeks was laag gemeet aan agro-morfologiese merkers. Die hoogste diversiteitsvlak is gesien in die Indica inskrywings en die laagste diversiteit in die *O. glaberrima* inskrywings. Japonica inskrywings het bruin of groen aarpunte gehad en was baardloos. *O. glaberrima* en landrasinskrywings het pers aarpunte gehad, en sommige het baarde gehad. Die Japonica groep het meestal bestaan uit verbeterde variëteite met ‘n wit perikarp terwyl *O. glaberrima* inskrywings en landrasse ‘n rooi perikarp gehad het. Die interspesifieke basters het eienskappe van *O. sativa* en *O. glaberrima* gekombineer in verskillende grade. Mikrosatelliet merkers het ‘n total van 321 allele getoon met ‘n gemiddeld van 11.9 allele per lokus en ‘n gemiddelde major alleel frekwensie van 0.29 per lokus. Die gemiddelde geendiversiteitswaarde was 0.81 en die polimorfisme informasie inhoud was 0.80 per lokus. Diversiteitsindekse vir interspesifieke basters was intermediêr tussen *O. sativa* en *O. glaberrima*, maar nader aan die *O. sativa* Japonica groep. Twee populasies is aangetoon wat ooreenkom met die *O. sativa* Indica groep en *O. sativa* Japonica groep. Interspesifieke basterinskrywings was versprei tussen die twee groepe. Mikrosatelliet data het meer diversiteit tussen inskrywings uitgewys in vergelyking met agro-morfologiese data. Daar was korrelasie tussen die verskillende matrikse, wat gewys het dat hulle dieselfde patrone in die struktuur van die diversiteit van die versameling wys.

Die agro-morfologiese merkers vir rooi saadkleur en blaarharigheid van die stuifmeelskenker is ook gebruik om die hoeveelheid kruisbestuiwing in vier genotipes van rys, bestaande uit verbeterde cultivars van *O. sativa* ssp. *japonica* (een inskrywing) en interspesifieke rys (drie inskrywings), te bepaal. Die dominante merkers van rooi saadkleur en blaarharigheid was effektief om die hoeveelheid uitkruising in die rys te bepaal. Daar was ‘n gemiddelde uitkruisingsvlak van 0.7% $\pm$ 0.51, met ‘n potensiële uitkruisingsvlak van 2.45% $\pm$ 0.86. Uitkruisingsvlakke het gewissel van 2.45% by 0.2 m van die donor tot 0.05% by 25 m van die donor. Geenvloei is beïnvloed deur die tipe genotipe, die afstand van die ontvanger vanaf die donor en blom sinkronisasie. Die kruisbestuiwingsvlakke was gewoonlik die hoogste by afstande naaste aan die skenker. Die studie het getoon dat natuurlike uitkruising gedeeltelik verantwoordelik is vir die voorkoms van nie-uniforme en af-tipe plante in ryslande in sub-Sahara Afrika.

Appendix 1 List and pedigree of varieties used in the study

S/N	Variety name	Pedigree	Taxon/group
1	Afhikari	Nankai128*///IR64/Hohoemi*//Kiri-Miyazaki	Japonica
2	PL 87-3	TOG5681/TOG7291	*glaberrima*
3	ITA123	Mutant of OS 6S (R 55 / VL 921)	Indica
4	ITA150	63-83/(DOURADO PRECOCE,ROK1,SE363G)	Indica
5	WAB56-104	IDSA 6 / IAC 164	Japonica
6	Moroberekan	Landrace	Japonica
7	NERICA 4	WAB 56-104 / CG 14	Interspecific
8	WAB100-B-B-B-B-21-H2	ITA 257/WABA	Japonica
9	WAB128-B-B-13-HB	WABC 4/DOURADO PRECOCE	Japonica
10	WAB176-8-HB	ITA 257 // IDSA 6/LAC 20	Japonica
11	WAB217-B-B-2-HB	IDSA 6/ROK 16 // SEL IRAT 194/GNONKONSOKA	Japonica
12	WAB224-8-HB	IDSA 6/LAC 20 // IDSA 6/IAC 164	Japonica
13	WAB272-B-B-1-H1	3290 / WABC 165	Japonica
14	WAB285-B-B-4-L2-L1-LB	ITA 150 / WABIS 764	Japonica
15	WAB306-B-B-6-L2-L1-LB	ITA 311 / WABYS 93	Japonica
16	WAB307-B-B-B1-L3-L1-LB	ITA 311 / WABIS 701	Japonica
17	NERICA 7	WAB 56-104 / CG 14	Interspecific
18	WAB337-B-B-7-H4	ITA 135 / WABC 165	Japonica
19	WAB365-B-2-H1-HB	ITA 257 / WAB 56-57	Japonica
20	WAB384-B-11-H2-H1-HB	ITA 184 / ROK 16	Japonica
21	NERICA 12	WAB 56-50 / CG 14	Interspecific
22	NERICA 13	WAB 56-50 / CG 14	Interspecific
23	WAB506-125-3	WAB 56-50 / DR 2	Japonica
24	WAB519-55-3	WAB 56-14 / DR 2	Japonica
25	NERICA 2	WAB 56-104 / CG 14	Interspecific
26	WAB96-1-1	ITA 257 / YS 121	Japonica
27	IRAT104	IRAT 13 / MOROBEREKAN	Indica
28	B6144F-MR-6-0-0	-	Indica
29	NERICA 14	WAB 56-50 / CG 14	Interspecific
30	NERICA 15	WAB 56-50 / CG 14	Interspecific
31	NERICA 16	CG 14 / WAB 181-18	Interspecific
32	NERICA 18	CG 14 / WAB 181-18	Interspecific
33	TOS 15505	Landrace	Indica
34	TOS 15729	Landrace	Indica
35	TOS 8076	Landrace	Indica
36	CG14	Landrace	*glaberrima*

S/N serial number

Appendix 2 Classification of agro-morphological traits

Quantitative traits	
Flag leaf length (cm)	**Spikelet opening angle (°)**
1 = Short (<25); 3 = Intermediate (29-33); 5 = Long (>36)	1 = narrow (21–26); 3 = moderately open (26.1–30); 5 = well open (30.1–35)
Flag leaf width (cm)	**Spikelet opening duration (min)**
3 = Narrow (<1); 5 = Intermediate (1 – 2) 7 = Broad (>2)	1 = short duration (36–45); 3 = medium duration (45.1–55.4); 5= long duration (55.5–60)
Leaf blade length (cm)	**Tiller number**
1 = Very short (<21); 3 = Short (~30); 5 = Intermediate (~50); 7 = Long (~70); 9 = Very long (>80)	3 = Low (<10 culms); 5 = Intermediate (~15 culms); 7 = High (>20 culms)
Leaf blade width (cm)	**50% Heading (days)**
3 = Narrow (<1); 5 = Intermediate (1–2) 7 = Broad (>2)	3 = early (50–69); 5 = medium (70–90); 7 = late (90+)
Panicle exsertion (%)	**Spikelets per panicle**
1 = Exserted (34–37); 3 = Moderately well exserted (38–40); 5 = well exserted (40+)	1 = Very poor (<50); 3 = Poor (51–75); 5 = Fairly good (76–100); 7 = Good (101–150); 9 = Very good (> 150)
Panicle length (cm)	**Grain filling percentage**
1 = Very short (<11); 3 = Short (~15); 5 = Medium (~25) 7 = Long (~35) 9 = Very long (>40)	1 = Completely sterile (0%); 2 = Highly sterile (1–49%); 3 = Partly sterile (50–74%); 4 = Fertile (75–90%); 5 = Highly fertile (>90%)
Panicle number per plant	**1000-grain weight (g)**
3 = Low (2–5 panicles); 5 = Intermediate (6–10 panicles); 7 = High (> 10 panicles)	1 = Light (<25); 3 = Intermediate (25-30); 5 = Heavy (>30)
Plant height (cm)	**Grain yield tha^{-1}**
1 = Very short (<50); 2 = Very short to short (51–70); 3 = Short (71–90); 4 = Short to intermediate (91–105); 5 = Intermediate (106–120); 6 = Intermediate to long (121–140) 7 = Long (141–155); 8 = Long to very long (156–180); 9 = Very long (>180 cm)	1 = Very low (<1); 3 = Low (1.1-2); 5 = Medium (2.1-3); 7= High (>3)

Appendix 3 List of microsatellite primers used in this study

Marker	Motif	Forward primer	Reverse primer	Tm (°C)	Expected product size
RM11	(GA)17	TCTCCTCTTCCCCCGATC	ATAGCGGGCGAGGCTTAG	55	140
RM19	(ATC)10	CAAAAACAGAGCAGATGAC	CTCAAGATGGACGCCAAGA	55	226
RM60	(AATT)5AATCT(AATT)	AGTCCCATGTTCCACTTCCG	ATGGCTACTGCCTGTACTAC	55	165
RM85	(TGG)5(TCT)12	CCAAAGATGAAACCTGGATTG	GCACAAGGTGAGCAGTCC	55	107
RM167	(GA)16	GAACATAAACCATGCGGGAG	AGCTAGTGGCAAAAGTGTGC	55	128
RM222	(CT)18	CTTAAATGGGCCACATGCG	CAAAGCTTCCGGCCAAAAG	55	213
RM248	(CT)25	TCCTTGTGAAATCTGGTCCC	GTAGCCTAGCATGGTGCATG	55	102
RM264	(GA)27	GTTGCGTCCTACTGCTACTTC	GATCCGTGTCGATGATTAGC	55	178
RM286	(GA)16	GGCTTCATCTTTGGCGAC	CCGGATTCACGAGATAAACTC	55	110
RM542	(CT)22	TGAATCAAGCCCCTCACTAC	CTGCAACGAGTAAGGCAGAG	55	113
RM1227	(AG)15	CATGGTAGCACACACCCTTG	CATCGACATGTGGACCACTC	55	176
RM3341	(CT)15	AGGACAGTCCACTCCCACTG	TCGTCGCCATCATTGGTATC	55	145
RM3529	(CT)39	CGCGCCACCTCGATATATAC	GCTCAGGTTAACCAAGGTGG	55	148
RM3907	(GT)20	GGAGGCCAAGGAAGAGGTAG	CGTCAATGGGGTAGGTCTTG	55	135
RM5463	(TC)19	ACCCTTGCAGACAACGTACC	ATATACCAGCAGCTGCATGC	55	176
RM5590	(TG)34	TGGATAAGCGATTGAGGTAG	CGTTATAATGAGGGAGGGAG	55	182
RM5812	(GAG)11	CGCTGACATCTTGCCCTC	GTAGGACCCACGTGTCATCC	55	144
RM6314	(CTT)11	GATTCGTGTCGGTTGTCAAG	GGTTCAGGGACGAATTTCAG	60	169
RM6673	(TAA)10	CATCGCATCGTATCGTATCG	GCTTCAAACACGCCTTCTTC	55	147
RM6840	(TCT)17	TACCAAGACTCCGCTATGGC	GAAGAAGGGATCATGGATCG	55	191
RM6842	(TCT)21	TAAATCGAAGGAGGGGGAAG	GGAAGAAGGAGGAGGAGGTG	61	196
RM10018	(GCT)7	ACTAGTACACCTCAACTTCACTCC	CCTTTAGTTTGCTTGTGACC	55	147
RM15281	(GA)25	CGGGCTTATATCTTTGGCAAATGG	GCCTCCTCCCTCCTTTCTCG	55	166
RM18452	(ACAT)10	AGACAACTAGAGGTAGCACATCTTCC	GTCGGGATTAAATTCTCGCTAGG	55	221
RM23662	(GGC)10	GAGAGGACGATGGCACTATTGG	CGAGGAACTTGATTCGCATGG	55	149
RM24035	(GGC)7	GCTCCAGTTTCTAGTGGGCTTGC	ATGCGGCAGTCAATCAACAGG	55	230
RM27973	(GCT)7	CCACACTGCCCAGGATTTAAGC	CTGTTCCCATCATCCAAATGACC	55	287

Appendix 4 Weather conditions during flowering

Flowering period	Temp (^{0}C)	RH (%)	WindSpd (ms^{-1})
Season 1			
14/10	28.7	78.4	2.0
15/10	28.9	84.2	2.2
16/10	29.2	85.2	2.0
17/10	28.5	83.0	2.1
18/10	28.6	86.3	2.4
19/10	29.3	97.3	1.6
20/10	29.0	84.0	2.4
21/10	29.1	80.4	2.3
22/10	29.2	81.2	2.7
23/10	29.1	77.6	2.6
24/10	29.2	87.4	2.4
25/10	26.9	75.4	1.7
Season 2			
14/1	29	68	2.0
15/1	29	67	2.2
16/1	29	65	2.0
17/1	29	71	2.1
18/1	29	69	2.4
19/1	29	55	1.6
20/1	29	71	2.4
21/1	29	71	2.3
22/1	29	70	2.7
23/1	29	44	2.6
24/1	29	30	2.4
24/1	27	35	1.7
26/1	28	39	2.0
27/1	29	40	2.0
28/1	29	41	2.4

Temp = Temperature; RH = Relative humidity; WindSpd = Wind speed