GENETIC DIVERSITY AND SELECTION CRITERIA FOR YIELD AND AROMA IN RICE (*ORYZA SATIVA* L.) GERMPLASM OF THE EASTERN DEMOCRATIC REPUBLIC OF CONGO

PAUL KITENGE KIMWEMWE

MASTER OF SCIENCE

(Plant Breeding)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2024

Genetic Diversity and Selection Criteria for Yield and Aroma in Rice (*Oryza sativa* L.) Germplasm of the Eastern Democratic Republic of Congo

Paul Kitenge Kimwemwe

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Plant Breeding of the Jomo Kenyatta University of Agriculture and Technology.

2024

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature……Date…….

Paul Kitenge Kimwemwe

This thesis has been submitted for examination with our approval as the University Supervisors.

Signature……Date……..

Prof. Edward George Mamati, PhD

JKUAT, Kenya

Signature……Date……..

Dr. Mamadou Fofana, PhD

International Institute of Tropical Agriculture (IITA), DRC

DEDICATION

To my beloved wife, Kisanga Brigitte Esther, and our precious child, Darrel Kimwemwe, you both have been the guiding light throughout my thesis journey.

To my father, Uluwa Kimwemwe Robert, and mother, Amnazo Bisenda Esther, your unwavering support has been my pillar of strength.

This thesis is dedicated to the entire Kimwemwe family, whose love and encouragement have made this achievement possible. Thank you for inspiring me to reach greater heights together as a family.

ACKNOWLEDGEMENT

It is said that a journey of a thousand miles begins with a single step. Thank you, God, for providing me with strength and protection throughout this journey.

I am very grateful to the Great Lakes Integrated Agriculture Development Project (PICAGL), funded by the World Bank Group through the government of the Democratic Republic of Congo (DRC) and implemented by the International Institute of Tropical Agriculture (IITA), for granting me a fellowship.

My heartfelt appreciation goes to Jomo Kenyatta University of Agriculture and Technology (JKUAT) for creating a conducive learning environment and imparting valuable skills. I extend special thanks to my dedicated supervisors, Prof. Edward George Mamati, and Dr. Mamadou Fofana for their unwavering support and mentorship. I acknowledge your immense impact on my academic and personal development.

I would like also to express my deep appreciation to the authorities and staff of the Institut National pour l'Etude et la Recherche Agronomiques (INERA/DRC), including Mr. Dominique Kankonde Ntumba (Director General), Prof. Dr. Ir. Amand Mbuya Kankolongo (former Director General), Mr. Daniel Lunze (Scientific Director), and Prof. Dr. Roger Kizungu Vumiliya (former Scientific Director), for nominating me in the agenda of capacity building.

I am truly thankful to the staff at the International Institute of Tropical Agriculture (IITA) Nairobi, Bukavu, and Kalemie, particularly Dr. Jacob Mingouna, Ms. Julie Lunzihirwa, Dr. Paul Martin Dontsop Ngwezet, Ms. Despine Bamuleke and Mr. Leblanc Bahiga, for their indispensable administrative and technical support throughout this journey.

My gratitude is extended to the staff of the INERA/DRC, Emilingombe station, particularly Mr. Serkali Alimasi Maliba (Station Manager), and the rice program team, Mr. Enock Kaboza Yamba-Yamba, and Mr. Marcel Assani Shabani, for their unwavering support during my research.

I am also thankful to the authorities of Université de Kalemie, particularly Prof. Victor Kalunga Tshikala, Rector of the University, and Prof. Fidel Chuma Mushagalusa, Dean of the Faculty of Agriculture and Environmental Sciences, for nominating and granting me leave to pursue my postgraduate studies.

I would like also to thank the ILRI staff, especially the capacity development, mycotoxin laboratory, and SEQART AFRICA team, for their support during the laboratory works of my research at the International Livestock Research Institute (ILRI).

I sincerely appreciate my country mates, colleagues, and classmates, Chance Bahati Bukomarhe and Mamy Binzunga Makumbu, you have been a second family to me throughout our academic journey in Kenya. Our diverse group of opinions and reflections contributed to my scientific growth. I hope we keep in touch and continue to collaborate in future endeavors.

I am deeply thankful to my family; especially my parents, Mr. and Mrs. Kimwemwe; my in-laws, Mugaruka family; my god-parents, Mr. and Mrs. Kusakana; my brothers, sisters, and friends who have constantly been rooting for me during this journey. Your unwavering support has been instrumental in my journey. Thank you for believing in me always. Thank you for your understanding during certain periods I was overwhelmed and discouraged.

Finally, I would like to thank my wife, Kisanga Brigitte Esther, for her endless love, patience, support, and prayers. I could not have completed this without your support. To our dearly loved son, Darrel Kimwemwe, thank you for coming into our lives and constantly adding bright colors to it. This dissertation should serve as a reminder to you that you can be all you want to be, and the world is yours to explore.

TABLE OF CONTENTS

DECLARATION ..... **ii**

DEDICATION ..... **iii**

ACKNOWLEDGEMENT ..... **iv**

TABLE OF CONTENTS ..... **vi**

LIST OF TABLES ..... **xi**

LIST OF FIGURES ..... **xii**

LIST OF APPENDICES ..... **xiii**

ABBREVIATIONS AND ACRONYMS ..... **xiv**

ABSTRACT ..... **xv**

CHAPTER ONE **1**

INTRODUCTION **1**

1.1 Background Information 1

1.2 Problem Statement 3

1.3 Justification 4

1.4 Research Hypotheses 4

1.5 Objectives of the Study 4

1.5.1 Overall Objective 4

1.5.2 Specific Objectives 4

CHAPTER TWO 6

LITERATURE REVIEW 6

2.1 Domestication and Diversity of Rice 6

2.2 Rice Production in Sub-Saharan Africa 6

2.3 Rice Production in the Democratic Republic of Congo 7

2.4 Rice Production Constraints 8

2.5 Genetic Diversity in Plant Genetic Resources 8

2.5.1 Concept and Importance of Diversity in Crop Improvement 8

2.5.2 Factors Determining Genetic Diversity in Plants 10

2.5.3 Evaluation of Genetic Diversity in Plants 10

2.5.4 Measurement of Genetic Diversity 12

2.6 Aroma in Rice and its Importance 13

2.6.1 Biochemical Basis of Aroma in Rice 14

2.6.2 Genetic Basis of Aroma in Rice 15

2.6.3 Metabolism Pathways of 2-Acetyl-1-Pyrroline in Rice 17

2.6.4 Evaluation of Aroma in Rice 18

2.7 Improvement of Rice Yield 20

CHAPTER THREE 22

MATERIALS AND METHODS 22

3.1 Plant Materials 22

3.2 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm of the Eastern DRC 23

3.2.1 Plant Tissue Preparation and DNA Extraction 23

3.2.2 Single Nucleotide Polymorphism Genotyping 24

3.3 Identification of the Selection Criteria Associated with High Grain Yield among the Rice Accessions 24

3.3.1 Plant Materials and Experiment Sites 24

3.3.2 Experimental Design, Layout, and Field Management 25

3.3.3 Data Collection 26

3.4 Classification of Aromatic Rice Genotypes and Identification Volatile Organic Compound Related to the Aroma Trait 28

3.4.1 Plant Material and Sample Preparation 28

3.4.2 Sensory Evaluation of Aroma in Rice 28

3.4.3 Extraction and Collection of Volatile Organic Compounds 28

3.5 Statistical Analysis 31

3.5.1 Genetic Diversity and Population Structure Analysis 31

3.5.2 Determining Selection Criteria Associated with High Grain Yield Among the Rice Germplasm Accessions 32

3.5.3 Untargeted Metabolomics and Association among Volatile Organic Compounds 34

CHAPTER FOUR 36

RESULTS 36

4.1 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm of the Eastern DRC 36

4.1.1 Characterization of DArTseq-Derived SNP Markers 36

4.1.2 Population Structure 38

4.1.3 Genetic Diversity and Phylogenetic 40

4.2. Identification of the Key Selection Criteria Associated with High Grain Yield among the Rice Germplasm Accessions 44

4.2.1 Analysis of Genetic Parameters 44

4.2.2 Principal Components Analysis 46

4.3.3 Phenotypic Correlation and Path Analysis 49

4.4 Identification of aromatic genotypes and volatile organic compounds associated to the rice grain aroma trait 52

4.4.1 Classification of Rice Genotypes Based on Sensory Test and 2-Acetyl-1-Pyrroline Content 52

4.4.2 Identification of Volatile Organic Compounds Associated with 2-AP 55

CHAPTER FIVE **62**

DISCUSSION **62**

5.1 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm Accessions of the Eastern DRC 62

5.2 Selection criteria for grain yield in rice 66

5.3 Identification of Aromatic Genotypes and Determination of Volatile Organic Compounds Associated with Rice Grain Aroma 69

CHAPTER SIX **72**

CONCLUSION AND RECOMMENDATIONS **72**

6.1 Conclusion 72

6.2 Recommendations 72

REFERENCES **73**

APPENDICES **95**

LIST OF TABLES

Table 2.1: Candidate Genes and their Contributions to Rice Aroma 16

Table 3.1: Sources of Rice Germplasm Used in the Study 22

Table 3.2: Phenotypic Traits, Collection Stage, and Method Used 27

Table 4.1: Proportion of SNP Transitions and Transversions Mutation Types Across the Genomes of the 94 Rice Genotypes 38

Table 4.2: Analysis of Molecular Variance among 94 Genotypes Based on SNPs .. 41

Table 4.3: Genetic Diversity of Five Rice Sub-Populations Based on 8389 SNP Markers 41

Table 4.4: Population's Pairwise Genetic Differentiation Index (F_{st}) 42

Table 4.5: Analysis of Genetic Parameters for Yield and Yield Components among 49 Rice Genotypes 45

Table 4.6: Contribution of 14 Traits to the Variation among 49 Rice Genotypes 47

Table 4.7: Phenotypic Path Analysis of Yield and Yield Related-Traits among 49 Rice Genotypes 51

Table 4.8: Sensory-Based Classification of Rice Genotypes and their 2-Acetyl-1-Pyrroline (2-AP) Contents 53

LIST OF FIGURES

Figure 2.1: QTLs and Associated Candidate Genes for Aroma in Rice 16

Figure 2.2: Putative Biosynthesis of 2-AP 18

Figure 3.1: Field Experimental Layout in 7x 7 Triple Lattice Design. 25

Figure 4.1: SNP Markers' Characteristics 37

Figure 4.2: Population Structure of 94 Rice Genotypes Based on Variation in SNP DArTseq Markers with K = 5 39

Figure 4.3: PCA Illustrating Relatedness and Distribution of the 94 Rice Genotypes Based on 8389 SNP Markers 40

Figure 4.4: Neighbor-Joining Tree of Five Sub-Populations Based on the F_{st} Values 42

Figure 4.5: Phylogenetic Tree Showing Clustering Pattern among 94 Rice Genotypes Based on SNP Markers 43

Figure 4.6: Scatter Plot Showing Contribution of the 14 Variables to the Variation Observed in PC1 and PC2. 48

Figure 4.7: Scatter Plot Showing the Variation Observed among 49 Rice Genotypes Considering PC1 and PC2 48

Figure 4.8: Correlation between Yield and Yield Related-Traits 49

Figure 4.9: Volcano Plot Analysis of Aromatic vs Semi-Aromatic Class of Rice, Depicting Differential Chromatogram Peaks 55

Figure 4.10: Heatmap Matrices 58

Figure 4.11: Chromatogram of an Aromatic Rice Depicting Significant Peaks Positively Associated to 2-AP 60

LIST OF APPENDICES

Appendix I : Rice Genotypes Used for Determining the Selection Criteria for Rice Grain Yield Selection 95

Appendix II: Rice Genotypes Used in Identifying Potential Volatile Organic Compounds Associated to the Rice Grain Aroma Trait 96

Appendix III: Mean Values for Yield and Yield Components of 49 Rice Genotypes Evaluated 97

Appendix IV: Euclidean Genetic Distances among the 94 Rice Genotypes based on 8389 DArTseq SNP Markers 99

Appendix V: Correlation Coefficients between 2AP and the Significant Features 100

Appendix VI: Intensities (Dalton) of Significant Peaks Associated to 2AP Across the Aromatic and Semi-Aromatic Rice 101

Appendix VII: List of Publications and Conference Presentation 107

ABBREVIATIONS AND ACRONYMS

AMOVA	Analysis of Molecular Variance
CAR/DVB/PDMS	Divinylbenzene/Carboxen/Polydimethylsiloxane
CTAB	Cetyltrimethylammonium Bromide
IITA	International Institute of Tropical Agriculture
INERA	Institut National pour l'Etude et la Recherche Agronomiques
IRRI	International Rice Research Institute
DNA	Deoxyribonucleic Acid
DRC	Democratic Republic of the Congo
DArTseq	Diversity Array Technology sequencing
FAO	Food and Agriculture Organization
QTL	Quantitative Traits Loci
SNP	Single Nucleotide Polymorphisms

ABSTRACT

Improving yield and aroma are among significant objectives of many plant breeding programs, since they are two critical parameters that determine farmers' and consumers' variety acceptability and consequently the productivity and market value of rice, respectively. This study aimed to evaluate the genetic diversity and population structure of rice germplasm from the Eastern DRC, determine key parameters associated with high grain yield, classify genotypes and establish the volatile compounds associated with the rice grain aroma. Therefore, 8389 high-quality filtered SNPs generated from 94 rice genotypes using the DArTseq method, were used for genetic diversity and population structure evaluation. The ADMIXTURE program, used for structure analysis, revealed five sub-populations with admixtures. Analysis of molecular variance revealed significant variation between sub-populations (36.09%) and within genotypes (34.04%). The low overall number of migrants (Nm = 0.23) and high fixation index (F_{st} = 0.52) indicated limited gene flow and significant differentiation between the sub-populations. Observed heterozygosity (H_o = 0.08) was lower than expected heterozygosity (H_e = 0.14) because of the high inbreeding (F_{is} = 0.52) nature of rice. A high average Euclidean genetic distance (0.87) revealed the existence of genetic diversity among the 94 genotypes. Selection criteria for grain yield was determined using a set of 49 genotypes. The field experiment was conducted following a 7x7 Triple lattice design in two locations, each with three replicates. Data were collected on fourteen morphological traits. Grain yield reflected a significant and positive correlation with the number of productive tillers/hill, panicle weight, number of primary branches/panicle, number of filled grains/panicle, and number of spikelets/panicle. The sensory test conducted by five panelists, along with the 2-Acetyl-1-Pyrroline (2-AP) profiles classified 49 non-aromatic, 32 semi-aromatic and five aromatic rice genotypes. The classification confirmed Basmati 370, ARS563-425-1-B-2-3 genotypes with strong aroma and high 2-AP contents (1.4 and 1.11 mg/kg, respectively). Furthermore, the untargeted metabolomics revealed 22 major compounds co-detected and positively associated to 2-AP in rice samples. These findings offer additional tools and information to breeding programs for enhancing rice productivity and aroma in the region and beyond, that would ultimately benefit both farmers and consumers.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Rice (*Oryza sativa* L.) is an important cereal crop that serves as a staple food for over 50% of the global population, driving significant agricultural and industrial activities (Mohidem et al., 2022). Its cultivation, processing, and distribution support 140 million of people and contribute to national and international trade, making rice a vital component of economic growth and food security worldwide (Muthayya et al., 2014; Mohidem et al., 2022). It provides over 20% of the calories consumed worldwide, making it a crucial staple food. Rice is cultivated in a wide range of agro-climatic conditions worldwide (Kshirod & Eero, 2017).

In the Democratic Republic of Congo (DRC), rice is the second most-consumed cereal after maize, providing food and income for many households (IFAD, 2019).
For centuries, farmers in DRC have practiced traditional methods of rice cultivation, selecting and preserving seeds from the best-performing plants and others with preferred traits. Over time, these processes have resulted in the development of landraces and locally adapted rice varieties with distinct traits that are well-suited to the region's diverse agro-ecological conditions (Joshi et al., 2023). To address the increasing demands for rice in DRC, the national rice breeding program of the Institut National pour l'Etude et la Recherche Agronomiques (INERA), in consortium with the international research centers; International Institute of Tropical Agriculture (IITA), AfricaRice, International Rice Research Institute (IRRI), has actively selected and introduced new varieties aimed at enhancing rice productivity while preserving the genetic diversity and cultural significance of rice in the region. With its potential; four million hectares of irrigable lowland (Food and Agriculture Organization et al., 2019), DRC could significantly contribute to food security by contributing to increasing rice production and availability in the Sub-Saharan Africa region.

According to various forecasts, the world population is expected to grow significantly and reach 9.7 billion by 2050 (Gu et al., 2021). Among the nine countries that are

projected to have over 50% population growth by 2050, majority are located in tropical regions (Campos & Caligari, 2017). These countries include India, Nigeria, the DRC, Ethiopia, the United Republic of Tanzania, Indonesia, and Egypt (Gu et al., 2021). In these countries, rice is the fastest-growing and preferred food commodity, driven by high population growth, rapid urbanization, and changes in eating habits (Balasubramanian et al., 2007; Seck et al., 2013). To satisfy the growing demand without affecting the resource base adversely, Ahmadi et al. (2014) project a 50% increase in rice production by 2030 in Africa. Effective ways to increase food production are to ensure availability of quality, affordable, and locally produced food (Campos & Caligari, 2017). Therefore, there is an urgent need to develop high-yielding rice varieties with preferred characteristics, to meet the increasing population demand and adapt to the current climate conditions, which is the primary goal of crop breeders (Siddiq & Vemireddy, 2021).

Development of new varieties largely requires availability of genetic variation in the desirable traits within the germplasm accessions (Bhandari et al., 2017). Utilizing plants' genetic diversity, breeders develop new and improved crop varieties with desirable traits, such as nutritional and grain quality, fragrance, resistance to pests and diseases, tolerance to flood and drought, and improved yield, in order to tackle worldwide issues related to food security, sustainability, and adaptability to climate change (Swarup et al., 2021). According to Khan et al. (2015), greater genetic variability enhances the probability of identifying superior genotypes within a population. As pointed out by Novoselović et al. (2016), a limited genetic base is a significant challenge that renders plants more susceptible to biotic and abiotic stress conditions.

Fragrant rice, also known as aromatic rice, holds a prominent position in the global rice market, commanding the highest prices among rice varieties (Calingacion et al., 2014). Prior to even tasting the rice, the aroma quality serves as a crucial parameter that influences its acceptance or rejection (Verma & Srivastav, 2020). Furthermore, for consumers, aroma quality is a primary sensory characteristic that comprises both taste and odor. Traditional rice consumers particularly prioritize aroma as a key selection criterion. It is also widely acknowledged as a crucial property of rice that

indicates its superior quality and market value (Custodio et al., 2019; Twine et al., 2023). Rice fragrance is significantly determined by the composition and proportion of volatile compounds in the rice grain (Hu et al., 2020). Presently, over 250 volatile compounds have been identified in rice. Among these, 2-Acetyl-1-pyrroline (2-AP) holds particular significance as it distinguishes aromatic rice from non-aromatic varieties (Prodhan & Qingyao, 2020).

1.2 Problem Statement

Rice is a staple crop in the Eastern Democratic Republic of Congo (DRC), providing a significant source of nutrition and income for the region's population. It is ranked amongst the top cereals, second after maize. The estimated national consumption of white rice ranges from 7kg to 19.5kg/person/year in 2018 (IFAD, 2019). Demographic dynamics and urbanization are expected to impact food demand with implications on food security and nutrition. In DRC, the demand for rice will continue to rise due to population growth, economic growth, and urbanization (FAO, 2022).

However, rice productivity in the region is low, and crop yields are often impacted by various biotic and abiotic stresses, including pests, diseases, and adverse environmental factors (IFAD, 2019; Iqbal et al., 2018). Moreover, the aroma trait of a rice variety holds a pivotal role in determining its competitiveness and acceptability in the local, regional and global market. Developing new well-adapted high yielding rice cultivars with aromatic characteristics is imperative in addressing the growing demand, and meeting consumers' preferences; thereby enhancing food security in the region. Crop genetic diversity is a pillar for developing new cultivars with desired traits; and enhancing the effectiveness of selection for quantitative traits like yield and aroma necessitates comprehending their correlation with the underlying traits.

There is limited knowledge regarding the genetic diversity, yield-related traits, and aroma-related compounds among the available rice germplasm in Eastern DRC, which is crucial for breeding efforts aimed at enhancing rice production and meeting the increasing demand for high-yielding and aromatic varieties.

1.3 Justification

This research holds significant implications for rice breeding and agricultural practices in the Eastern DRC. By exploring the genetic makeup and selection criteria for high yield and aroma traits of the available germplasm accessions, the study intends to evaluate the potential of the genetic resources for developing new high-yielding rice varieties, with consumers' preferences. Thereby promoting sustainable agriculture and enhancing food security for the growing population of the region. Additionally, identifying the traits that are reliably associated with improved grain yield and aromatic properties will enable breeders and farmers to make informed decisions when selecting and cultivating rice varieties.

1.4 Research Hypotheses

i. There is no significant genetic variation among rice germplasm accessions in the Eastern DRC.
ii. There is no significant correlation among yield components yield among the rice germplasm.
iii. There are no differences among the rice genotypes based on the aroma and no biochemical markers associated with aromatic properties in rice.

1.5 Objectives of the Study

1.5.1 Overall Objective

This study aimed to determine the genetic variation among rice germplasms in the Eastern DRC and establish the attributes that could be associated with the selection criteria for yield and aroma.

1.5.2 Specific Objectives

i. To evaluate the genetic diversity and population structure of rice germplasm accessions in the Eastern DRC.
ii. To determine the selection criteria associated with high grain yield among the rice germplasm accessions.

iii. To classify aromatic genotypes and establish volatile organic compounds associated with the rice aroma.

CHAPTER TWO

LITERATURE REVIEW

2.1 Domestication and Diversity of Rice

Around 10,000 years ago, rice (*Oryza sativa*) was domesticated from distinct gene pools of the wild grass species *Oryza rufipogon*. These gene pools, known as Indica and Japonica, are believed to have diverged between 200,000 and 400,000 years before their domestication (Kovach et al., 2009). Multiple studies have identified five genetically distinct subpopulations within these varietal groups. The Indica group consists of the indica and aus subpopulations, while the Japonica group includes the temperate japonica, tropical japonica, and aromatic subpopulations (Glaszmann, 1987). Initially, it was thought that the aromatic subpopulation was most closely related to the indica types based on grain morphology. However, research using simple sequence repeats revealed that the aromatic subpopulation is genetically more closely related to the japonica subpopulation than to the indica subpopulation (Kovach et al., 2009). Recent re-sequencing studies suggest that the aromatic subpopulation is an ancient admixture between temperate japonica and aus, with a minor presence of indica ancestry (Kovach et al., 2009; McCouch et al., 2016).

2.2 Rice Production in Sub-Saharan Africa

Rice is a traditional staple food in parts of West Africa and Madagascar, and it is increasingly becoming an important staple in East, Central, and Southern Africa. In recent years, the relative growth in demand for rice has been faster in SSA than anywhere else in the world. Demand for rice has increased due to population growth and a shift in consumer preference for rice, especially in urban areas (Arouna et al., 2021).

According to FAO (2022) (www.fao.org/faostat/), Africa produced around 39.87 million tonnes of paddy rice on 16.52 million ha which represents about 5.1 and 10.01 % of the world's total rice production and rice area, respectively. The contribution estimates of African areas on paddy rice production in 2003 are West Africa

accounting for 70.4% (approx. 8.74 million ha) of rice area. The major contributor countries are Nigeria (47.9%), Guinea (5.20%), Côte d'Ivoire (5%), and Mali (4%). East Africa accounts for 16.1% of the rice area; the major contributors are Tanzania (6.0%) and Madagascar (3.19%). Central and southern Africa accounts for 7.5% of the rice area. The major contributors are the Democratic Republic of the Congo (4.05%) and Mozambique (1.8%) (Norman & Kebe, 2004).

2.3 Rice Production in the Democratic Republic of Congo

In Democratic Republic of the Congo, rice is now one of the main staple foods in urban areas, after cassava, maize, and peanuts. Given the high population growth and economic development, it is expected that domestic production will not be able to meet the growing demand for staple starches, such as rice. In 2018, the DRC imported 150,000 metric tonnes of white rice, compared to the demand of 252,000 metric tonnes of white rice in the local market (Bailey et al., 2020). The irrigation potential is estimated at 4 million ha but currently, rice is cultivated on about 450,000 ha. Rainfed rice cultivation is the most practiced and covers 98% of the rice-growing area sown mainly in the forest regions of the central basin (Minister of agriculture and rural development, DRC, 2013). The regions with high rice-growing potential are found in eight out of the 26 provinces and produce 72% of national production distributed as follows: the provinces of Tshiopo, Bas-Uélé, Haut-Uélé, and Ituri (which produce 28% of national production), Mongala and Tshuapa provinces (which produce 13% of national production), Maniema province (with 11% of national production) and Sankuru province (with 20% of national production) (Minister of agriculture and rural development, DRC, 2013). Irrigated rice farming remains marginal, representing only 2% of the total rice cultivation area. This practice is observed in Kinshasa, the Ruzizi Plain in South Kivu, and to a smaller extent in Equateur and Bas-Congo. Around 1945, the colonial authorities decided to develop and enhance the irrigated areas in the Ruzizi Plain, and in 1970, the Chinese mission strengthened the project by installing other irrigated rice areas from the different rivers of the plain (Mirindi, 2017).

2.4 Rice Production Constraints

Rice is cultivated in four ecosystems of SSA: dryland (38% of the cultivated rice area), rainfed wetland (33%), deepwater and mangrove swamps (9%), and irrigated wetland (20%) (Balasubramanian et al., 2007). Abiotic stresses; drought, flood, and variable rainfall; extreme temperatures; salinity; acidity/alkalinity and poor soils, soil erosion, and high phosphorous fixation and biotic constraints; weeds, blast, Rice yellow mottle virus (RYMV), and African rice gall midge (AfRGM) limit rice production on the continent. The changing climate is expected to further aggravate the abiotic constraints and reduce rice yields in all ecosystems (Balasubramanian et al., 2007). Erratic rainfall patterns, prolonged droughts, and increased temperatures can reduce yields and make rice farming less predictable (Balasubramanian et al., 2007).

2.5 Genetic Diversity in Plant Genetic Resources

2.5.1 Concept and Importance of Diversity in Crop Improvement

Genetic diversity refers to the degree of genetic differentiation observed among individuals in a species, whether it is within a specific variety or a broader population (Brown, 1983). The variation in crop plants, whether influenced by natural processes or guided by human intervention, is fundamentally dependent on the genetic variation present within the population. The existing variations found both within and between plant species are the basis driving progress in crop improvement programs (Bhandari et al., 2017).

The significance of genetic diversity increases in the face of climate change and unforeseen events, as it represents a reservoir of numerous novel traits that confer tolerance to diverse biotic and abiotic stresses. Genetic diversity is an anchor of agricultural phenomena like heterosis (hybrid vigor) and transgressive segregation (appearance of extreme phenotypes beyond the range of parental traits). To address shortcomings in commercial varieties and create new ones, diverse lines are required for corrective measures (Bhandari et al., 2017). Therefore, the primary objectives of any crop improvement program revolve around identification of existing diverse lines, establishment of diversity in cases where it is lacking or limited, and subsequent

utilization of this diversity. The genetic diversity found within a population of a plant species, encompassing cultivars, landraces, and wild individuals, represents a vital asset for enhancing food production and promoting the development of sustainable agricultural practices (Esquinas-Alcázar, 2005). It serves as the foundation for both the natural survival of plants and the improvement of crops. The presence of diversity within plant genetic resources creates opportunities for plant breeders to create novel and enhanced cultivars that exhibit desirable traits. These traits encompass characteristics preferred by farmers, such as high yield potential and large seed size, as well as traits favored by breeders, including resistance to pests and diseases, and photosensitivity (Govindaraj et al., 2015).

With the changing climate patterns, breeding of climate-resilient varieties has gained increased importance (Bhandari et al., 2017). The existence of genetic diversity, manifested in the form of wild species, related species, breeding stocks, and mutant lines, among others, serves as a valuable resource for plant breeders in developing climate-resilient varieties (McCouch, 2004). These varieties necessitate the incorporation of novel traits, such as tolerance towards potential new insect pests, diseases, extreme temperatures (both high and low), flood, as well as various air and soil pollutants. To accommodate ever-evolving breeding goals, different genes need to be conserved within cultivated and cultivable crop species through germplasm resources (Govindaraj et al., 2015).

Genetic diversity within and between crop plant species provides breeders with the opportunity to select superior genotypes, whether for direct use as new varieties or as parents in hybridization programs. The presence of genetic diversity between two parents is crucial for achieving heterosis (hybrid vigor) and obtaining transgressive segregants. Genetic diversity also facilitates the development of varieties with specific traits, such as improved quality and tolerance to biotic and abiotic stresses (Bhandari et al., 2017).

2.5.2 Factors Determining Genetic Diversity in Plants

Genetic diversity changes over time due to various factors. Evolutionary forces, such as selection, mutation, migration, and genetic drift, exert continuous influence on populations, leading to changes in the frequency of alleles and ultimately impacting genetic diversity (Bhandari et al., 2017). Domestication and artificial selection favors certain alleles while diminishing others, resulting in an increased prevalence of selected alleles. Consequently, genetic diversity is reduced through domestication when compared to wild populations. Natural selection also plays a significant role in shaping genetic diversity. Directional and stabilizing selection decrease genetic diversity, while disruptive selection increases it. Mutation is another factor that contributes to an increase in genetic diversity (Amos & Harwood, 1998).

The mating system of crop plants also has an impact on genetic diversity. Inbreeding reduces genetic diversity, while outbreeding promotes it. Genetic drift can lead to the loss of rare alleles, thereby diminishing genetic diversity. The spatial distribution of individuals within a species also influences genetic diversity. A larger physical distribution decreases the likelihood of individuals sharing the same genetic makeup. Certain techniques, such as wide-hybridization, hybridization between incompatible types, or introgression from previously isolated populations, can enhance genetic diversity by generating new phenotypes (Bhandari et al., 2017); Conversely, intra-specific hybridization decreases genetic diversity (Brauer et al., 2023). Gene flow within a population, particularly through monoallelic expression, can significantly contribute to increased genetic diversity by introducing new alleles. This process is influenced by various factors, including the presence of concurrent geography, overlapping flowering times, and shared pollinators (Rundle & Nosil, 2005).

2.5.3 Evaluation of Genetic Diversity in Plants

The evaluation of genetic diversity within and among plant populations is a standard practice that involves the use of various techniques. In the pregenomic era, methods such as morphological characterization and biochemical evaluation, specifically allozyme analysis, were commonly employed (Govindaraj et al., 2015).

The morphological markers, especially for traits such as plant height, yield, maturity, and grain color, as well as resistance to insects and diseases have been effectively used for diversity studies (Franco et al., 2001; Ullah et al., 2022). Nevertheless, the exclusive reliance on morphological markers was shown to be unreliable because they are limited in number, susceptible to environmental factors, and influenced by the plant growth stages (Govindaraj et al., 2015; Nadeem et al., 2018). Therefore, molecular markers have become an indispensable tool in genetic research (Deres & Feyissa, 2022), particularly in assessing genetic diversity. Molecular markers allow precise and rapid varietal identification, germplasm characterization, collection, and management. Earlier molecular markers such as random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR) and amplified fragment length polymorphism (AFLP) have been frequently used for fingerprinting and characterization of varieties and germplasm accessions of different crop species (Deres & Feyissa, 2022). Despite the potential benefits of these molecular markers in plant breeding, their usefulness is limited due to their reliance on prior sequence information, which can be costly and time-consuming to obtain (Deres & Feyissa, 2022). Genotyping-by-sequencing (GBS) techniques such as diversity array technology sequencing (DArTseq) is a low-cost and rapid genotyping method that enables the screening of hundreds of highly polymorphic markers without previous genome sequence information for the detection of loci (Nadeem et al., 2018; Wenzl et al., 2004). DArTseq produces both dominant (SilicoDArT) and co-dominant (SNP) markers that have successfully been applied for genetic diversity and population structure analysis in crops such as cassava (Adu et al., 2021), taro (Fufa et al., 2022), rice (Mogga et al., 2018), rye (Niedziela & Bednarek, 2023), sorghum (Mudaki et al., 2023), and wheat (Novoselović et al., 2016). The use of DArTseq markers enables a comprehensive analysis of genetic diversity, complete genome profiling, and high-density mapping of complex traits, all of which are crucial for marker-based breeding (Deres & Feyissa, 2022).

2.5.4 Measurement of Genetic Diversity

To understand the impact of selection, mating systems, and other breeding interventions in the field of population genetics, it becomes crucial to elucidate and quantify the extent of genetic diversity within a population and the distribution of this genetic diversity among different populations (Nadeem et al., 2018).

The assessment of genetic variation can be accomplished at multiple levels, such as the variations in the DNA sequence among individuals within a population (Swarup et al., 2021). Molecular markers serve as potent instruments for investigating genetic diversity. They offer a means to quantify and analyze genetic variation, gain insights into population dynamics, and make well-informed decisions across multiple domains, including breeding, conservation, and evolutionary biology (Nadeem et al., 2018). Various methods and concepts are applied to determine genetic structure and genetic variability between and within populations of organisms (Singh, 2005); Allele frequencies at specific loci are counted to measure genetic variation. The Mean Number of Alleles (MNA) is used as an indicator of genetic variation within populations. Low MNA suggests low genetic variation due to isolation or historical events, while high MNA indicates allelic diversity due to factors like crossbreeding (Yang et al., 2010).

Heterozygosity measures genetic variation within a population. It can be estimated using observed and expected heterozygosity values at specific loci and across all populations (Eltaher et al., 2018; Luo et al., 2019).

Molecular data can be used to estimate inbreeding values, which reflect the extent of homozygosity within populations. Inbreeding coefficients are averaged to estimate average inbreeding coefficients for each population (Matessi & Jayakar, 1973).

Genetic differentiation between populations is assessed by examining the independence of allelic composition among populations. Statistical tests help determine if populations are genetically distinct. Diagnostic alleles unique to certain populations are used to assess genetic purity, introgression, and genetic composition.

Genetic admixture proportions can be estimated to identify gene flow between populations (Wright, 1965).

Methods like “Structure” are used to infer population structure, assign individuals to populations, and identify admixed individuals using multi-locus genotype data. Population structure analysis is essential for understanding the genetic diversity, distribution, and evolutionary dynamics of populations, which have significant implications for conservation, management, and breeding strategies (Schierenbeck, 2017). Linkage disequilibrium (LD) is tested to check for non-random associations between different loci. This can indicate factors like population admixture or selection. Genetic diversity within and between populations is quantified using Wright’s F statistics and Analysis of Molecular Variance (AMOVA). AMOVA is a commonly used statistical method for evaluating the extent to which different levels of population structure contribute to genetic variation patterns (Fitzpatrick, 2009).

These methods assess differentiation at various hierarchical levels of population structure (Fitzpatrick, 2009).

2.6 Aroma in Rice and its Importance

The aroma of rice is an important parameter of quality that plays a crucial role in enhancing its value in the rice market. The aroma of rice and its derivatives products holds significant importance as a quality characteristic. Various types of rice are cultivated worldwide, and many of them are renowned for their distinct fragrance. Certain rice varieties, including jasmine-type and basmati-type, are particularly renowned for their distinct aroma which make them highly favored among consumers (Verma & Srivastav, 2020). Prior to even tasting the rice, aroma serves as a crucial parameter that can influence its acceptance or rejection (Verma & Srivastav, 2020). Furthermore, for consumers, the aroma is a primary sensory characteristic that comprises both taste and odor. Traditional rice consumers particularly prioritize the aroma as a key criterion. It is also widely acknowledged as a crucial property of rice that indicates its superior quality and market value (Custodio et al., 2019; Twine et al., 2023). Therefore, when compared to non-aromatic rice, the aromatic rice holds significant potential to captivate rice consumers due to its distinctive taste and

deliciousness. This attraction contributes to its higher price, which in turn can lead to improved socio-economic conditions for farmers in developing countries involved in rice cultivation (Verma & Srivastav, 2022).

2.6.1 Biochemical Basis of Aroma in Rice

Cooked rice contains over 200 different types of volatile organic compounds (VOCs), from the chemical classes; alkanes, aldehydes, alcohols, esters, ketones, phenols, fatty acids, benzyl derivatives, enones, furans, furanones, monoterpenoids, sesquiterpenoids, naphthalenes, xylenes, pyridines, and pyrroles compounds being the primary constituents (Hashemi et al., 2013; Hu et al., 2020; Imran et al., 2023). However, only some of these volatiles have been associated with the overall perceived aroma of rice. Among the aroma compounds are 2-Acetyl-1-pyrroline (2-AP), aldehydes, heterocyclics, and alcohols (Hu et al., 2020). The 2-AP has been recognized as a key factor in determining the flavor of aromatic rice. It has often been employed as a reference compound to differentiate between aromatic and non-aromatic rice varieties.

In the analysis of cooked rice, various aroma attributes have been examined. These include characteristics such as bland-like, bran-like, brown rice, burned-like, buttery-like, cold-steam-bread-like, corn-like, corn-leaf-like, cracker-like, dusty-like, earthy-like, fermented-sour-like, floral-like, gasoline aroma-like, grainy-like, hot-steam-bread-like, musty-like, nut-like, paint-like, pandan-like, pearl-barley, plastic-like, popcorn-like, potato-like, rancid-like, raw-dough-like, rice milk-like, smoky-like, spicy-like, sulfur-like, tortilla-like, vegetable-like, and white glue-like (Verma & Srivastav, 2018). The popcorn-like aroma, sometimes referred to as pandan (*Pandanus amaryllifolius*) aroma, in rice is primarily attributed to 2-AP. This aroma compound holds paramount significance as it has been extensively utilized for identifying and distinguishing different rice cultivars based on their aroma profiles. The concentration of 2-AP varies among different varieties of cooked rice. While 2-AP can be detected in all aromatic rice varieties after cooking, only specific types of non-aromatic rice have the ability to release 2-AP during the cooking process (Kasote et al., 2021).

2.6.2 Genetic Basis of Aroma in Rice

The research on rice aroma has identified several candidate genes associated with the synthesis of aromatic compounds (Table 2.1). One of the most important aroma genes is OsBadh2, located on chromosome 8, consisting of 15 exons and 14 introns, which codes for a protein with 503 amino acids. By comparing the significant variations between aromatic and non-aromatic rice, Bradbury et al., (2005) detected major mutations in the badh2 gene, such as an 8 bp deletion and 3 SNPs in exon 7. Another aroma-related locus, aro4-1, on chromosome 4 contains the OsBadh1 gene, which is assumed to have a similar function to OsBadh2. Additionally, eight genes clustered with OsBadh1 on the aro4-1 locus may also contribute to grain aroma expression. Aro3-1, identified on chromosome 3, represents a minor aroma QTL, and a differentially expressed gene (*LOC_Os03g21040*) in this region has been linked to grain aroma (Amarawathi et al., 2007). On chromosome 5, a major aroma QTL harbors the *OsGlyI* gene, indicating the involvement of methylglyoxal in rice aroma. Further integration mapping suggests that OsBadh2, *OsBadh1*, *OsGly*, and *OsP5CS* genes are potential candidates responsible for aroma in rice (Amarawathi et al., 2007; Bradbury et al., 2005; Pachauri et al., 2014; Talukdar et al., 2017). The OsBadh1 gene, with a high homology to *OsBadh2*, might also influence rice aroma. Moreover, *OsBadh1* is associated with a BADH1 protein haplotype (PH2) with specific substitutions that affect substrate binding. Another candidate gene, *OsGlyII*, involved in methylglyoxal detoxification, has been identified on chromosome 3 (Prodhan & Qingyao, 2020). Figure 2.1 illustrates the integrated information and positions of various potential QTLs (Quantitative Trait Loci) associated with rice aroma, along with candidate genes located on different chromosomes.

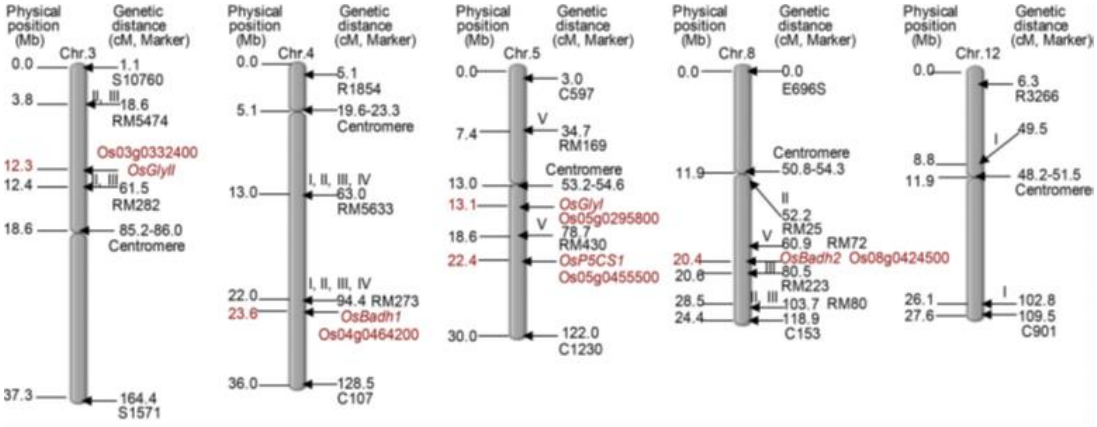

Figure 2.1: QTLs and Associated Candidate Genes for Aroma in Rice

Source: (Prodhan & Qingyao, 2020).

Table 2.1: Candidate Genes and their Contributions to Rice Aroma

Gene	ID	Tissue specificity (Expression)	Response to stress	Splicing	Protein interaction
OsP5CSI	Os05g0455500 LOC_Os05g38150	Flower buds, milk grains	Osmoregulation, salinity, anoxia	Introns 3, 19; Termination	Ferredoxin-dependent glutamate synthase
OsP5CS2	Os01g0848200 LOC_Os01g62900	Flowers, flower buds	Osmoregulation, salinity, anoxia	All introns; Termination	Glutamate synthase
OsGlyl	Os05g0295800 LOC_Os05g22970	Leaves before flowering flower buds	Salinity, anoxia	-	-
OsGlyll	Os03g0332400 LOC_Os03g21460	Flower buds, flower	Salinity, anoxia	-	Glyoxalase
OsGlyIII	Os01g0667200 LOC_Os01g47690	Roots and leaves before flowering	Salinity, anoxia	Introns 7, 8	Ferredoxin-nitrite reductase
OsBadhl	Os04g0464200 LOC_Os04g39020	Roots before flowering, flowers	Salinity, anoxia, submergence	Intron 4	Glutamate synthase
OsBadh2	Os08g0424500 LOC_Os08g32870	Flowers, flower buds	Salinity, anoxia, submergence	-	Glutamate synthase

Source: (Prodhan & Qingyao, 2020)

2.6.3 Metabolism Pathways of 2-Acetyl-1-Pyrroline in Rice

2-AP biosynthesis is a product of the polyamine degradation pathway (Prodhan & Qingyao, 2020). Polyamines are organic compounds that contain multiple amino groups. Initially, the involvement of the polyamine pathway was proposed in 2-AP synthesis (Behera & Panda, 2023). Within this pathway (Figure 2.2), polyamines, such as arginine, putrescine, and ornithine, undergo conversion to γ-amino butyraldehyde (GABald), which serves as a precursor for γ-aminobutyric acid (GABA). GABald then spontaneously cyclizes to Δ1-pyrroline, which acts as a precursor for 2-AP and plays a crucial role in regulating its biosynthesis (Chen et al., 2008). In non-aromatic rice varieties, GABA is formed from GABald through the action of the functional BADH2 enzyme, encoded by the OsBadh2 gene, which inhibits the biosynthesis of 2-AP. On the other hand, in aromatic rice varieties, the non-functional BADH2 enzyme, encoded by osbadh2, fails to convert GABald to GABA, leading to accumulation of GABald and the production of 2-AP (Bradbury et al., 2005). Therefore, 2-AP can be enzymatically synthesized through both the glycolysis and polyamine degradation pathways or non-enzymatically through direct formation.

In the direct synthesis pathway, which is independent of BADH2, glutamate is converted to a precursor of proline, which then reacts with methylglyoxal to generate 2-AP (Huang et al., 2008). In aromatic rice, proline, ornithine, and glutamate serve as nitrogen sources, while GABald, 1-pyrroline, and Δ1-pyrroline act as sources of the pyrroline ring. Methylglyoxal provides the necessary carbon for the synthesis of 2-AP, as discussed by Prodhan & Qingyao, (2020).

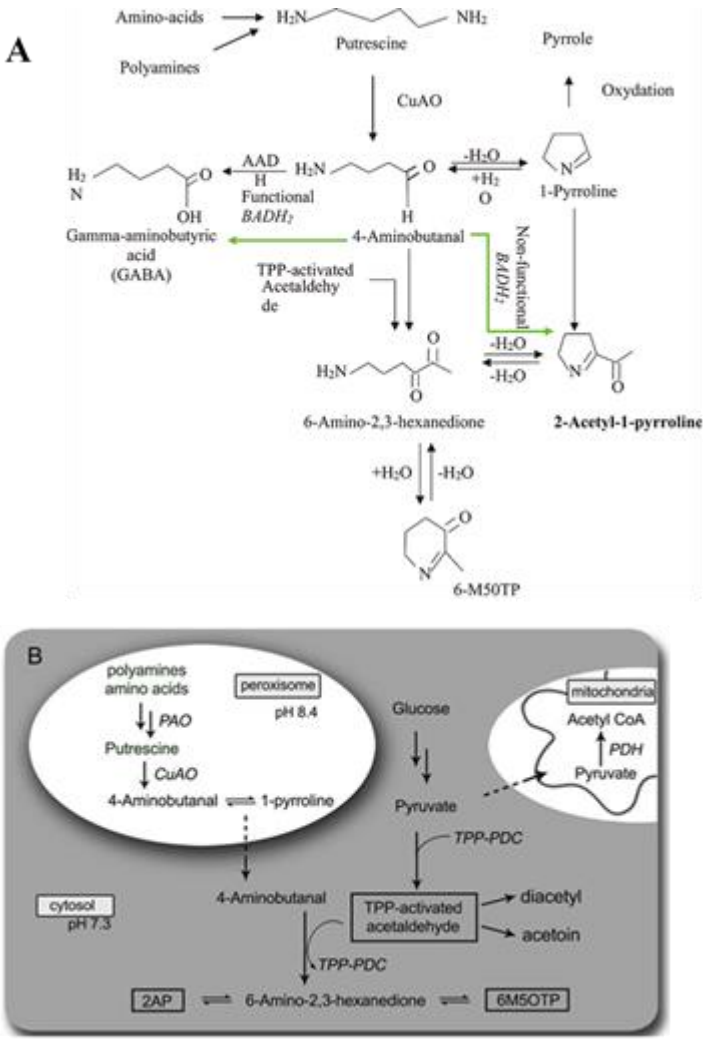

Figure 2.2: Putative Biosynthesis of 2-AP

Key: (A) TPP-catalyzed transfer of a C2 unit to 4-aminobutanal. The reaction with 4-aminobutanal forms an unstable diketone 6-amino2,3-hexadione, which cyclizes to either 2-AP or 6M5OTP. (**B**) Cellular localization of the reactions leading to the formation of 2-AP and 6M5OTP; CuAO: copper amine oxidase; PAO: polyamine oxidase; AADH – amino aldehyde dehydrogenase; PDC – Pyruvate decarboxylase; PDH – pyruvate dehydrogenase complex (Daygon et al., 2017).

2.6.4 Evaluation of Aroma in Rice

Identification of characteristic volatile compounds in rice is of importance to rice aroma analysis. The methods used to detect rice aroma can be categorized into two types: sensory evaluation and instrumental analysis (Zheng et al., 2022).

2.6.4.1 Sensory Evaluation Method

The sensory evaluation using the human nose as the detector offers direct, effective, unique, and intuitive information about rice aroma. This method relies on the final human sense and is regarded as the most direct and intuitive approach. The human nose possesses a theoretical odor detection limit of approximately 10^{-19} mol (Wilkie et al., 2004), making sensory evaluation a valuable and sensitive method for rice aroma analysis and odor active volatiles analysis. This evaluation is typically done by trained panel members who consume specific foods to maintain a neutral taste. The taste, aroma, and overall quality of the cooked rice are evaluated. Statistical analysis is employed to ensure experimental accuracy. Honma et al. (2019) used the descriptive approach to assess sensory properties of different cooked brown rice samples by Eight trained panelists. Multiple attributes, including aroma, showed significant differences among the samples. In the study conducted by Lapchareonsuk & Sirisomboon (2015), the aim was to develop a method using visible and shortwave near-infrared (NIR) spectroscopy to assess the sensory characteristics of cooked rice. The trained sensory panel evaluated qualities such as adhesiveness, hardness, stickiness, dryness, whiteness, and aroma. The findings indicated a significant correlation between these sensory attributes and the visible and shortwave NIR spectral data. Partial least squares regression was employed to establish models for predicting the sensory qualities of cooked rice using both NIR spectroscopy techniques. The prediction results for sensory qualities exhibited coefficient of correlation ranging from 0.837 to 0.918.

2.6.4.2 Instrumental Analysis Methods

Instrumental analysis techniques include gas chromatography (GC), electronic nose (E-nose), and metal oxide semiconductor (MOS) sensors, among others. The results of the detection process include both qualitative analysis and quantitative comparison. Qualitative analysis helps identify and differentiate various volatile organic compounds (VOCs), allowing for identification or distinction of different rice varieties. On the other hand, quantitative analysis provides information on the concentration of specific VOCs (Zheng et al., 2022).

The gas chromatography (GC) method has been widely employed as the primary instrumental analysis technique for analyzing rice aroma. It is an effective method for studying volatile compounds and is commonly utilized for both qualitative and quantitative analysis of volatiles in rice. The GC method involves two crucial steps: sample pretreatment and final detection. These steps play a significant role in ensuring accurate and reliable analysis of volatile components in rice samples (Hu et al., 2020).

Various sample preparation methods and analytical instrumentation techniques have been developed for the extraction, identification, and quantification of 2-AP from both aromatic and non-aromatic rice samples. Sample preparation methods include purge and trap (PTM), simultaneous distillation extraction (SDE), solvent extraction followed by direct injection (SEfbDI), solid-phase microextraction (SPME), headspace analysis (HSA) and supercritical fluid extraction (SFE). On the other hand, major analytical instrumentation techniques involve gas chromatography (GC) coupled with detectors such as nitrogen-phosphorus (NPD), flame ionization (FID), mass spectrometry (MS), or olfactometry. These methods and techniques provide valuable tools for the extraction, identification, and quantification of 2-AP in rice samples (Verma & Srivastav, 2022).

These methods are coupled with Gas Chromatography-Mass Spectrometry (GC-MS), Gas Chromatography-Flame Ionization Detector (GC-FID) and/or Gas chromatography olfactometry (GC-O) have been used widely and are ideally suited for readily analyzing 2-AP found in rice due to its versatile and sensitive detection at the range of ppb concentration (Verma & Srivastav, 2022).

2.7 Improvement of Rice Yield

Increasing rice yield has become a major objective of breeders and growers for ensuring food security in many countries. However, for yield improvement in rice, breeders need a deep understanding of the relationship between yield and yield-related traits (Li et al., 2014).

Yield is an important and complex trait. It is controlled both by several genes known as quantitative trait loci and highly affected by external environmental factors. Yield

in rice is influenced directly by traits like panicle number per unit area and/or per plant, and 1000 grain-weight, filled grains per panicle as well as indirect traits like growth period, plant height, panicle length, tillering ability, seed setting rate, seed length, and grains per panicle (Li et al., 2019).

Correlation and path analysis provide a good measure to estimate the relationship between traits and help to identify the most important traits to be taken into consideration for effective selection for increasing yield (Akhi et al., 2016).

Several strategies to improve yield potential include conventional hybridization, heterosis breeding, ideotype breeding, wide hybridization, molecular marker-assisted breeding and genetic engineering (Khan et al., 2015). Conventional breeding is still a widely utilized approach for developing new crop varieties with higher yield potential. The integration of genomic research, molecular biology, transgenic breeding, and molecular marker applications with conventional plant breeding techniques has created the basis for molecular plant breeding and will certainly accelerate rice improvement programs across the world. Molecular-assisted selection (MAS) is now an integral component of germplasm improvement. Various genes for a large number of traits have been tagged with molecular markers to apply MAS for trait selection and improvement. Map-based cloning has resulted in the isolation of several genes for resistance to biotic and abiotic stresses as well as yield-related traits (Khan et al., 2015).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Plant Materials

Ninety-four rice genotypes (Table 3.1) available at the rice breeding program of the National Institute for Agronomic Research and Study (INERA) in the Eastern Democratic Republic of Congo were used in this study. This collection comprised of genotypes from different research centers including AfricaRice (43), International Rice Research Institute (IRRI)-Burundi (25), IRRI-Kenya (10), INERA (12) and local landraces (5).

Table 3.1: Sources of Rice Germplasm Used in the Study

Entry No.	Genotype name	Source/ program	Entry No.	Genotype name	Source/ program
1	Komboka	IRRI-Burundi	27	D20-ARS-3-2	AfricaRice
2	IR64	IRRI-Burundi	28	IR96279-33-3-1-24	IRRI-Burundi
3	IBEI6	AfricaRice	29	ARS134-B-1-1-5-B	AfricaRice
4	GIZA128	IRRI-Burundi	30	Orylux7-1	AfricaRice
5	Nipponbare	IRRI-Burundi	31	WAHX14N-926	AfricaRice
6	Jasmine	IRRI-Burundi	32	MR254	AfricaRice
7	NL59	AfricaRice	33	Golmy	AfricaRice
8	FKR	AfricaRice	34	ARS848-15-3-2-4	AfricaRice
9	08FAN10	IRRI-Burundi	35	IR93348:32-B-15-3-B-B-B-1	IRRI-Burundi
10	WAB2066-TGR2	AfricaRice	36	ARS168-3-B-1-B	AfricaRice
11	WAB2066-TGR3	AfricaRice	37	IR88638	IRRI-Burundi
12	IR99084-B-B-13	INERA-DRC	38	ARICA12	AfricaRice
13	IR127229	IRRI-Burundi	39	ARICA3	AfricaRice
14	IR106172-78 :1-B-B	INERA-DRC	40	IR64-sub-1	IRRI-Burundi
15	ARS848-15-3-2-3	AfricaRice	41	HHZSAL6	AfricaRice
16	IR106364-B-B-CNUS	INERA-DRC	42	ARS755-3-3-1-B	AfricaRice
17	ARS844-24-10-2-B	AfricaRice	43	ARS134-B-1-1-5	AfricaRice
18	ARS168-1-B-3-B	AfricaRice	44	IR990-48-B-B-12	IRRI-Burundi
19	ARS851-1-3	AfricaRice	45	IR64-biofortified	IRRI-Burundi
20	IR87638-10-2-2-4	INERA-DRC	46	IR107015-37	IRRI-Burundi
21	IR98419-B-B-11	INERA-DRC	47	ARS79-5-11-11	AfricaRice
22	IR97071-24-1-1-1	INERA-DRC	48	V18/RRS126-48-1-13-2	AfricaRice
23	ARS803-4-5-4-3	AfricaRice	49	Orylux11	AfricaRice
24	IR93856-23-1-1-1	INERA-DRC	50	ARS134-B-B-B	AfricaRice
25	ARS790-5-11-1-1	AfricaRice	51	Magoti	Local
26	IR17015-6-5-3-B1	INERA-DRC	52	Runingu	Local
53	IR106359-B-18-5	INERA-DRC	74	ARS169-2-B-3-B	AfricaRice
54	IR95624-B-138-3	INERA-DRC	75	ARS134-B-1-1-4	AfricaRice
55	IR13A461	IRRI-Burundi	76	IR82574/643-1-2	IRRI-Burundi
56	Mugwiza	IRRI-Burundi	77	Orylux5	AfricaRice
57	Vuninzara	IRRI-Burundi	78	SAHEL210	AfricaRice
58	IR97045-24-1-1-1	IRRI-Burundi	79	IR841	IRRI-Burundi
59	Kigoma	Local	80	ARS39-145/EP-3	AfricaRice
60	Makasane	IRRI-Burundi	81	ARS101-4-B-1-1-B	AfricaRice
61	Rukaramu	Local	82	ARS101-4-B-1-3	AfricaRice

Entry No.	Genotype name	Source/ program	Entry No.	Genotype name	Source/ program
62	Musesekara	IRRI-Burundi	83	NERICA-L-19-Sab-1	AfricaRice
63	Yasho-Yasho	Local	84	ARS756-1-1-3-B-2-2	AfricaRice
64	Kasozi	IRRI-Burundi	85	ARS563-425-1-B-2-3	AfricaRice
65	IR7525	IRRI-Burundi	86	ARICA4	IRRI-Kenya
66	Orylux7	AfricaRice	87	ARICA17	IRRI-Kenya
67	ART29	INERA-DRC	88	Basmati370	IRRI-Kenya
68	Sipi	INERA-DRC	89	IRAT109	IRRI-Kenya
69	CRS36	IRRI-Kenya	90	NERICA1	IRRI-Kenya
70	ARICA2	AfricaRice	91	NERICA2	IRRI-Kenya
71	NL19	AfricaRice	92	NERICA10	IRRI-Kenya
72	NL14	AfricaRice	93	NERICA12	IRRI-Kenya
73	NL17	AfricaRice	94	PAN84	IRRI-Kenya

3.2 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm of the Eastern DRC

3.2.1 Plant Tissue Preparation and DNA Extraction

Twenty seeds from each of the 94 genotypes were pre-germinated by soaking in water and incubating at 28°C in a growth chamber for 48 hours. The sprouted grains were sown on seedling nursery trays using cocopeat as substrate with hoagland solution and raised in a greenhouse at Jomo Kenyatta University of Agriculture and Technology (JKUAT) for two weeks. Leaf samples were sent to SEQART AFRICA (https://www.seqart.net/) at the International Livestock Research Institute (ILRI), Nairobi, where genotyping was conducted following their in-house procedure. Briefly, the samples were freeze-dried and stored until DNA extraction process. Total genomic DNA was extracted using the NucleoMag® plant genomic DNA extraction kit (Macherey-Nagel™, Dueren, Germany) as described by the manufacturer. Briefly, plant tissue was lysed using CTAB-Lysis Buffer MC1. Adjusting the binding conditions of nucleic acid with Binding Buffer MC2 and addition of paramagnetic beads were carried out simultaneously. After magnetic separation and removal of supernatant, the paramagnetic beads were washed with Wash Buffers MC3, MC4, and 80 % ethanol to remove contaminants and salt. Then the ethanol from previous wash steps was removed by the Wash Buffer MC5. Finally, highly purified DNA was eluted with low salt Elution Buffer MC6 and directly used for downstream applications. The quality and quantity of DNA were assessed using 0.8% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific™, Waltham, MA, USA), respectively.

3.2.2 Single Nucleotide Polymorphism Genotyping

Genotyping was done by sequencing using the DArTSeq method, which relies on genome complexity reduction using enzymes (Kilian et al., 2012; Sansaloni et al., 2011). Briefly, the extracted genomic DNA was digested with two restriction enzymes (*PstI* and *MseI*) simultaneously, which cleave the DNA at specific recognition sites. The resulting fragments were then ligated to adapters that contained unique barcodes for each sample (Wenzl et al., 2004). The ligated fragments were then amplified by PCR to generate a library of DNA fragments for sequencing. The library fragments were then sequenced using single read sequencing runs of 77 cycles by the Illumina Hiseq2500.

DArTseq markers scoring was achieved using DArTsoft14 implemented in KDCompute plug-in system (https://kdcompute.seqart.net/kdcompute/), which is an in-house software scoring pipeline for GBS analysis performing polymorphism identification and calling as well as producing a number of quality parameters allowing for selection of polymorphic markers with the optimal quality. Single nucleotide polymorphisms (SNPs) markers were scored for presence or absence.

3.3 Identification of the Selection Criteria Associated with High Grain Yield among the Rice Accessions

3.3.1 Plant Materials and Experiment Sites

A sample of forty-nine rice genotypes (Appendix I) of the populations were selected from the germplasm and used in the field experiment aimed at determining the selection criteria for high grain yield.

The experiments were carried out in farmer's rice fields at two locations: Taba-Congo (S05°50'19" and E029°17'20", at an altitude of 775m) and Kabimba (S05°34'19.5", E029°20'03.3", at an elevation of 781m above sea level), located at 12km and 65km away from Kalemie, respectively, in the Tanganyika province of the Democratic Republic of Congo. The experimental fields were set up on clayey loam soils.

3.3.2 Experimental Design, Layout, and Field Management

The field experiments at each site were conducted using a 7 x 7 Triple lattice design with three replications, as presented on the layout in Figure 3.1. Each replication comprised 49 plots, representing 49 rice genotypes distributed in seven blocks, with each block containing seven plots. The experiment plot covered a total area of 40m x 22m, with each replication measuring 11m x 22m and each experimental plot measuring 2m x 1m. A 1m-wide path was left between the replicates and between the blocks to facilitate easy movement during data collection. The 49 rice genotypes were randomly assigned within each of the three replications using the Plant Breeding Tools (PBTools, version 1.4., http://bbi.irri.org/products).

Figure 3.1: Field Experimental Layout in 7x 7 Triple Lattice Design.

Each experimental plot consisted of 32 plants spaced at 25cm x 25cm, with one plant per hole. The experiments were conducted under irrigated conditions in rice farmer's fields from July to December 2021. Prior to sowing, the rice seeds were soaked in warm water (28°C) for one day and then kept under warm conditions for another day to accelerate germination. Subsequently, the germinated seeds were sown in the nursery, and after 21 days, the seedlings were transplanted to the prepared muddy fields. The fields were kept dry for up to two weeks to promote good root development,

after which they were flooded with water for three days. Throughout the crop cycle, the fields were weeded four times, with the first weeding carried out eighteen days after transplanting, followed by subsequent weeding at monthly intervals. Before any weeding operation, the fields were drained to facilitate the process. For fertilization, urea (46% N) was applied in two fractions, with the first fraction added 20 days after transplanting at the start of the tillering phase, at a rate of 60 kg ha^{-1}, and the second fraction applied during panicle initiation, also at a rate of 60 kg ha^{-1}. During the reproductive and maturation stages, the fields were protected against birds using bird mist nets and scarecrows.

3.3.3 Data Collection

Data were collected on fourteen morphological traits following the Descriptors for Rice (IRRI, 1980) as shown in Table 3.2. These traits included days to flowering, days to maturity, plant height, panicle length, the number of productive tillers/hill, the number of primary branches/panicle, the number of spikelets/panicle, the number of filled grains/panicle, panicle weight, thousand grains weight, and grain yield. Measurements were taken from ten selected plants or panicles per hill from the middle of each plot. After harvest, a sample of 10 grains from each genotype was shelled to assess the physical grain quality, i.e., grain length and grain width, and the ratio of grain length to grain width was calculated. The mean values of data collected from the two locations were then subjected to statistical analysis.

Table 3.2: Phenotypic Traits, Collection Stage, and Method Used

Traits	Code	Collection stage	Collection method
Days to flowering	DTF	Flowering	Recorded by counting days from seedling to the time when 50% of plants on a plot have flowered
Days to maturity	DTM	Maturity	Recorded as the duration in days from seeding to the time when more than 80% of the grains on the panicles are fully ripened.
Plant height	PH	Reproductive	Measured (cm) from soil surface to tip of the plant using the measuring tape. The average of ten plants per plot was considered
Panicle Length	PL	Near maturity	Length of main axis of panicle measured from the panicle base to the tip. Record the average of ten representative plants.
Number of productive tillers/hill	NPTH	Maturity	Obtained by counting the productive tillers per hill and the average of ten plants was recorded
Number of primary branches/panicle	NPBP	Maturity	Obtained by counting the primary branches on a panicle, and the average of ten panicles was recorded
Number of spikelets/panicle	NSP	Maturity	Number of spikelets from ten randomly selected plants in a plot and the average was recorded
Number of filled grains/panicle	NFGP	Maturity	Filled grains/panicle was counted from ten randomly selected panicles and the average was recorded
Panicle weight	PW	Maturity	A random sample of ten panicles were weighted and the average was deduced for each plot
1000 grain weight	ThGW	Post-harvest	A random sample of 1000 well-developed, whole grains dried to 13% moisture content were weighed on a precision balance.
Grain length	GL	Post-harvest	Measured in millimeters as the distance from the base of the lowermost sterile lemma to the tip (apiculus) of the fertile lemma or palea, whichever is longer using the caliper.
Grain width	Gw	Post-harvest	Measured in millimeters as the distance across the fertile lemma and the palea at the widest point using the caliper.
Grain length/width ratio	GLGwR	Post-harvest	Obtained by dividing the grain length by the grain width.
Grain yield (t/ha)	GY	Post-harvest	Harvested grains were weighted (g) for each plot and extrapolated to tons per hectare.

Source: (IRRI, 1980)

3.4 Classification of Aromatic Rice Genotypes and Identification Volatile Organic Compound Related to the Aroma Trait

3.4.1 Plant Material and Sample Preparation

A set of 85 diverse rice genotypes including breeding lines and local landraces (Appendix II), maintained at the Institut National pour l'Etude et la Recherche Agronomiques, were grown under irrigated conditions in Kalemie (S05°50'19" and E029°17'20", at an altitude of 775m), from September 2021 and harvested in January 2022. The harvested paddy samples were separately sun-dried to about 14% of moisture content. A sample of 100 grams (g) of each genotype, packed and sealed in polypropylene bag, transferred to the mycotoxin laboratory at the International Livestock Research Institute (ILRI), Nairobi, Kenya, for analyses. All samples were stored at - 20°C until further processing. Fifteen g of all paddy samples were de-hulled and milled with a Toss Sample Mill machine (Cyclotec™ 1093) to obtain brown rice flour.

3.4.2 Sensory Evaluation of Aroma in Rice

A sensory test was carried to identify aromatic genotypes based on the human's nose appreciations. Five hundred (500) mg of brown rice flour from each sample were placed in a 20 mL covered headspace vial, then 1 mL of distilled water was added and cooked in a DK 20 Digester (Velp Scientifica™, Via Stazione, Italy) for 20 min at 80ºC. Each rice sample were cooked and smelled by five panelists. Three known rice varieties belonging to the class of aromatic (Basmati 370), semi-aromatic (Komboka) (Ng'endo et al., 2022) and non-aromatic (Nipponbare) (Tabanao et al., 2021), grown in the same conditions, were utilized as checks in the classification. The smell strength of the test samples was compared to that of each one of the checks, which allowed the panelists to classify the genotypes into aromatic, semi-aromatic and non-aromatic.

3.4.3 Extraction and Collection of Volatile Organic Compounds

The classified aromatic genotypes were further utilized to investigate Volatile Organic Compounds (VOCs). A method reported by Xie et al. (2019) with modifications was

applied for isolation, detection and quantification of 2-acetyl-1-pyrroline (2AP) and other VOCs in rice using headspace solid phase-microextraction (HS-SPME) method coupled with the gas chromatography-mass spectrometry (GC-MS). Before use, the SPME fiber (50/30 μm CAR/DVB/PDMS, Supelco®, Bellefonte, PA, USA) was conditioned, at 260ºC for 30min as described by the manufacturer.

Three hundred (300) milligrams of brown rice flour samples were weighed in 20 mL headspace vial and sealed with a cap (Magnetic precision metal screw cap and a polytetrafluoroethylene (PTFE)/silicone Septa). 10 μl 2 ng/μl 2,4,6-trimethylpyridine (TMP) with 99% purity (Sigma-Aldrich®, USA) was added into the vial as an internal standard. The extraction of volatile compounds was achieved by incubating the samples at 80ºC for 10min and then shake the vial followed by extraction at 80ºC for 30min, in the DK 20 Digester (Velp Scientifica™, Via Stazione, Italy). During the extraction phase, the SPME fiber was inserted into the headspace vial via the septum and exposed at about 1cm above the sample matrix to absorb the VOCs. A blank run using GC-MS was also executed after every tenth sample to eliminate any residual traces from previous runs in the GC column.

3.4.3.1 Gas Chromatography-Mass Spectrometry Analysis

The adsorbed VOCs were desorbed by inserting the SPME fiber into the inlet of a 7890A GC (Agilent Technologies, USA) and exposing the fiber for 10 min at a constant temperature of 260°C. The oven temperature was held for 1 min at 50°C, raised to 120°C at 5°C/min, then increased at 20°C/min to 260°C, and finally held there for 5 min. The separation was achieved with a VF-5MS column (30 m × 0.25 μm, 0.25 μm film thickness, Agilent Technologies, USA) using helium as a carrier gas at a constant flow rate of 1.0007 $mL.min^{-1}$.

The mass spectrometer (MS) was operated at SCAN mode with a source temperature of 260°C at a speed of 10,000 within a mass range from m/z 35 to 500, and a run was completed in 33min. The full-scan mode (MS1) was used for the volatile profiling.

3.4.3.2 Data Acquisition and Processing

The selected ion monitoring (SIM) mode was run to select the confirmation and quantification ions for each volatile. The ions for 2AP and TMP were first monitored to enable selection of genotypes with detectable level of 2AP, a known major biochemical marker for aroma in rice. Two groups of ions were monitored (Lee et al., 2019; Xie et al., 2019), the first related to 2AP had an m/z of 68, 83, and 111 with retention time around 6.5 min, and the second with regards to TMP had an m/z of 106, and 121 with retention time around 8.39. Peak areas were determined by integrating an ion unique to for 2AP and TMP.

The concentration of 2AP was calculated as an equivalent weight of TMP (Equation (1)), using the formula proposed by Tanchotikul & Hsieh (1991):

$$2AP(mg/kg) = \frac{R*T}{RRT} * \frac{1}{[W*(1-M)]} \quad (1)$$

Where R: the ratio of the combined peak areas of the 111, 83, and 68 m/z ions of 2AP to the peak area of the 106 m/z ion of TMP; T: the amount of TMP used as the internal standard (μg); RRF: the relative recovery factor; W: the wet weight of the rice sample (g), and M: the moisture content of the rice sample.

The samples from the aromatic and semi-aromatic classes with or without detectable level of 2AP content were retained and used for investigating other VOCs using untargeted metabolomics path (Yao et al., 2019).

The acquired raw GC-MS data files were converted to Computable Document Format (.cdf) using OpenChrom Community Edition 1.2.0 (Alder) (Wenig & Odermatt, 2010). Data processing for feature detection and alignment was performed as described by Yao et al. (2019), using the XCMS package implemented in R environment (Smith et al., 2006). Chromatographic peaks were filtered and identified using the following parameters, (a) feature detection using centWave method and adjusted parameters such as maximal ppm m/z deviation in consecutive scans at 100, noise filter at 6, peak width in seconds at minimum of 3 and maximum of 10, prefilter mass trace and only retain the ones with at least 3 peaks with intensity ≥ 100, integration type at 2, and

Signal/Noise ratio at 6; (b) Retention time correction was used with the Obiwarp method, plottype=deviation; (c) alignment of peaks across samples was set as m/z width = 0.25, bw = 10, minfraction = 0.5, and minsamp = 1. The normalization technique median fold change (medFC) was used to adjust peak intensities between samples to a common scale.

3.5 Statistical Analysis

3.5.1 Genetic Diversity and Population Structure Analysis

The generated SNP markers were aligned to the rice reference genome, *Oryza sativa* V7.0 (http://rice.plantbiology.msu.edu), to determine their positions along the 12 rice chromosomes. The criteria for data filtration were as follows: non-informative monomorphic markers were removed, markers with a call rate >95% and minor allele frequency >5% were retained. The VCFtools V0.1.13 software (Danecek et al., 2011) was used for SNP markers filtration. To analyze the characteristics and distribution of the markers along the 12 rice chromosomes, parameters such as polymorphic information content (PIC), reproducibility, and call rate were determined using the dartR package in R (Gruber et al., 2018). Additionally, the same package was utilized to calculate the proportion of mutation types, including transversion (Tv) and transition (Ts), responsible for the observed polymorphism.

Population structure was analyzed to gain insight into the evolution of the 94 rice genotypes. The filtered SNP markers were utilized for structure analysis using ADMIXTURE V1.3.0 (Alexander & Lange, 2011) which uses a model-based maximum likelihood estimation. The optimal number of sub-populations was determined by evaluating the cross-validation errors based on K values (K = 1–10). As recommended by Alexander & Lange (2011), the K value associated with the lowest cross-validation error was considered the ideal number of sub-populations. The population structure was then visualized using R software, Version 4.3.0 (R Core Team, 2023). Additionally, principal component analysis (PCA) was conducted using the adegenet package in R (Jombart & Ahmed, 2011) to investigate genetic relatedness patterns among the sub-populations. The resulting PCA scores were exported and used

to generate a 3D PCA plot using the SRPLOT online platform (http://www.bioinformatics.com.cn/srplot, accessed on 24 April 2023).

Analysis of molecular variance (AMOVA) as described by Nei (1972), was used to partition the total variance into among and within populations using poppr package of R (Kamvar & Grunwald, 2021). To assess the genetic differentiation among the sub-populations identified in the structure analysis, Nei's pairwise fixation indices (F_{st}) (Nei, 1987) were generated using package hierfstat of R (Goudet, 2005).

The gene flow (Nm) among the sub-populations was estimated using the formula, $Nm = (1-F_{st})/4F_{st}$, as suggested by (Wright, 1965a). The results obtained from ADMIXTURE were used to calculate genetic diversity indices, including observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) for the sub-populations. The calculations were performed using the adegenet package in R (Jombart & Ahmed, 2011).

A neighbor-joining (NJ) phylogenetic tree was constructed using the ape package in R (Paradis et al., 2004) to visualize the genetic differentiation among the sub-populations. The relationships among individuals were analyzed by generating a pairwise genetic distance matrix using the Euclidean distance method implemented in R. The resulting phylogenetic tree was created using the hclust function in R and exported in Newick format using the ape package for annotation in the interactive tree of life (iTOL) version 6.5.2 (https://itol.embl.de/, accessed on 19 April 2023) (Letunic & Bork, 2021).

3.5.2 Determining Selection Criteria Associated with High Grain Yield Among the Rice Germplasm Accessions

The mean values of the evaluated traits for the 49 genotypes are shown in Appendix III. Genetic parameters were estimated to understand the nature of variation in yield and its components, in determining genetic and environmental effects on the expression of the studied traits. The genotypic variance ($\sigma^2 g$) (Equation (2)), environment variance ($\sigma^2 e$) (Equation (3)), phenotypic variance ($\sigma^2 p$) (Equation (4)) were determined by the formula suggested by Johnson et al. (1955).

Genotypic variance ($\sigma^2 g$):

$$\sigma^2 g = \frac{MSg - MSe}{r} \quad (2)$$

Where MSg is the mean square of genotypes, MSe is the mean square of error, and r is the number of replications.

Environmental variance ($\sigma^2 e$) :

$$\sigma^2 e = MSe \quad (3)$$

Phenotypic variance ($\sigma^2 p$):

$$\sigma^2 p = \sigma^2 g + \sigma^2 e \quad (4)$$

Genotypic and phenotypic coefficients of variation (GCV and PCV, respectively) were calculated according to the formula suggested by Johnson et al. (1955) as shown in Equations (5) and (6), respectively.

$$GCV = \frac{\sqrt{\sigma^2 g}}{X} x100 \quad (5)$$

$$PCV = \frac{\sqrt{\sigma^2 p}}{X} x100 \quad (6)$$

Where x is the experimental mean.

Heritability (Broad sense):

Heritability (Broad sense) was computed following the Standard broad-sense heritability method described by Johnson et al. (1955) as presented in Equation (7).

$$H^2 bs = \frac{\sigma^2 g}{\sigma^2 p} x100 \quad (7)$$

Genetic advance:

Estimation of genetic advance (GA) and genetic advance as percentage of the mean (GAM) were estimated following the formulas Johnson et al. (1955) as shown in Equation (8) and (9), respectively.

$$GA = \frac{k \times \sqrt{\sigma^2 p} \times \sigma^2 g}{\sigma^2 p} \quad (8)$$

Where GA = Expected genetic advance; k = Standardized selection differential at 5% selection intensity (K = 2.063).

$$GAM = \frac{GA}{X} \times 100 \quad (9)$$

Where GAM = Genetic advance as percentage of mean; GA = Expected genetic advance; X = Grand mean of a character.

Principal component analysis was carried out to determine the contribution of the individual genotypes and variables respectively to the total variation observed. This was carried out using R software (Version 4.3.1).

Phenotypic correlation analysis was performed to assess the relationship among the studied traits whereas path analysis was conducted to determine the direct and indirect effects of each trait on yield. These were performed in R software (Version 4.3.1), using the variability package.

3.5.3 Untargeted Metabolomics and Association among Volatile Organic Compounds

The resultant normalized data file obtained from the XCMS package containing details of each chromatogram peak detected (mass to charge ratio, retention time and intensity for each variable in every sample), was uploaded into MetaboAnalyst5.0 (https://www.metaboanalyst.ca/, accessed on 18 August 2023), a comprehensive online platform dedicated for metabolomics data analysis (Xia et al., 2009).

The Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) was employed along with the Classical Volcano Plot (Li, 2012) to identify features

that differ significantly (p-value <0.05) between the two groups of rice (Aromatic and Semi-aromatic) to segregate systematic variation based on linearity and orthogonality, and to extract significant features that can effectively distinguish between different sample groups (Wang et al., 2020). The Variable Importance in Projection (VIP) score was employed to pinpoint features with the highest discriminatory potential (VIP score > 1.0). Features meeting this criterion (VIP > 1) and having a p-value < 0.05 underwent further analysis, including fold change assessment (cutoff > 2.0). The relationship among these volatile compounds were elucidated by performing the correlation analysis. A heatmap hierarchical cluster analysis (HCA) of the genotypes based on the significant volatile metabolomics data were performed to provides a simplified data visualization, using the Euclidean distance and Ward methods (Mascellani et al., 2021).

The significant features which showed positive correlation with the 2AP were then identified by comparing mass spectra and retention index (RI) with the National Institute of Standards and Technology library (NIST) 2014 library (https://www.nist.gov/nist-research-library, accessed 21 August 2023). The RI values were obtained from the analysis of a mixture of n-alkane (C7-C26) using MS-DIAL software version 5.1.230719 (Tsugawa et al., 2015).

CHAPTER FOUR

RESULTS

4.1 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm of the Eastern DRC

4.1.1 Characterization of DArTseq-Derived SNP Markers

A total of 31,366 SNP markers were generated from the 94 rice genotypes using DArTseq method, out of which 27,831 markers (88.7%) were mapped to the reference genome. After filtering, 8389 informative SNP markers were retained for structure and diversity analyses. The distribution of markers on the 12 rice chromosomes, along with their characteristics after filtering, are presented in Figure 4.1.

The number of SNP markers ranged from 475 (chromosome 10) to 1014 (chromosome 1), with an average of 699 markers per chromosome (Fig 4.1a). The 8389 informative SNP markers had an average polymorphic information content (PIC) of 0.25, ranging between 0 and 0.5 (Fig 4.1b). The reproducibility of SNP markers varied from 87% to 100%, with a mean of 98%. Approximately 88% of SNP markers exhibited ≥95% reproducibility (Fig 4.1c). The call rate ranged from 96% to 100%, with an average of 97% (Fig 4.1d).

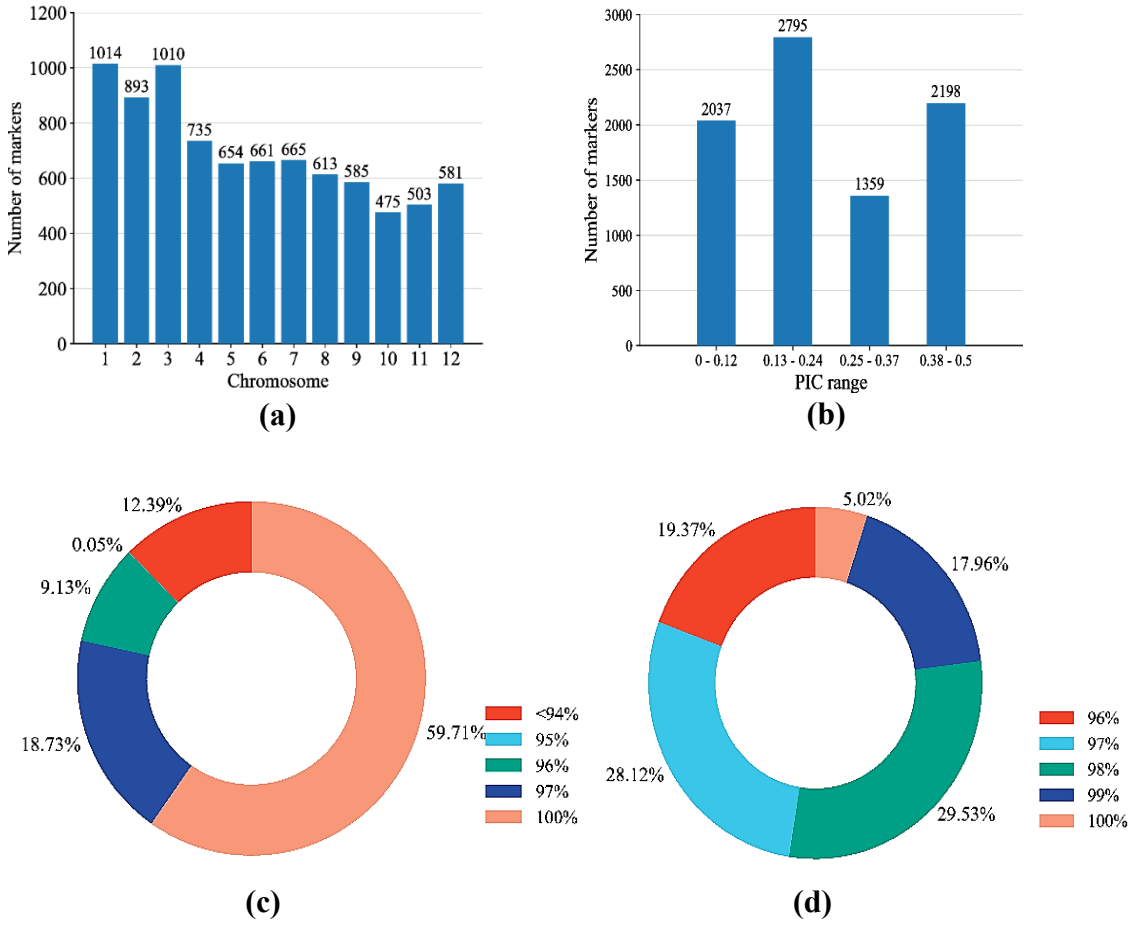

Figure 4.1: SNP Markers' Characteristics

Key: (**a**) distribution of SNP markers on the 12 rice chromosomes, (**b**) PIC range values of the SNP markers, (**c**) reproducibility of the SNP markers, and (**d**) call rate of the SNP markers.

The SNP mutation types are summarized in Table 4.1, showing the frequency of transitions (Ts; i.e., A/G, T/C substitutions) and transversions (Tv; i.e., A/T, A/C, T/G or C/G substitutions). The proportion of polymorphisms due to different transitions ranged from 31.06% (A/G) to 31.54% (T/C).

The proportion of polymorphisms due to transversions ranged from 7.5% (C/G) to 10.2% (A/T). Overall, among the SNP variations, transitions (62.6%) were more frequent than transversions (37.4%), with a Ts/Tv ratio of 1.67.

Table 4.1: Proportion of SNP Transitions and Transversions Mutation Types Across the Genomes of the 94 Rice Genotypes

	Transitions (Ts)		**Transversions (Tv)**				**Ts/Tv**
Substitution	A/G	T/C	A/T	A/C	T/G	C/G	
Number of alleles	2606	2646	859	829	816	633	
Frequency (%)	31.06	31.54	10.2	9.88	9.7	7.5	1.67
Total	**5252 (62.6%)**		**3137 (37.4 %)**				

4.1.2 Population Structure

Population structure based on a filtered set of 8389 SNP DArTseq markers gave five distinct sub-populations (herein referred to as Pop1, Pop2, Pop3, Pop4 and Pop5) (Figure 4.2) across the 94 genotypes. Considering K = 5, Pop1 was comprised of four rice genotypes out of which two were local landraces (Magoti, Runingu), and two (Jasmine and IR127229) were obtained from IRRI-Burundi. Pop2 was made up of 24 genotypes including 13 from AfricaRice, six from IRRI-Burundi and five from INERA-DRC. Pop3 was a group of 21 rice genotypes among which 10 were from AfricaRice, five from IRRI-Burundi, five from INERA-DRC and one from IRRI-Kenya. Pop4 (n = 37) was formed by genotypes from all the five sources though mainly comprised of AfricaRice genotypes (19) and IRRI-Burundi genotypes (12). Additionally, three genotypes were the local landraces, two from INERA-DRC and one from IRRI-Kenya. Pop5 was composed of eight genotypes out of which seven were IRRI-Kenya genotypes and one AfricaRice genotype.

PCA was used to further explore genetic relationship among the 94 rice genotypes (Figure 4.3) confirm the patterns of admixture populations (Liu et al., 2020; Ma & Amos, 2012). Based on the SNP markers, 36.3% of the total genetic variation was explained by the first three axes of the PCA, grouping the 94 genotypes into 3 clusters. The first comprised only 2 landraces; the second cluster gathered together most of the genotypes from AfricaRice, IRRI-Burundi and INERA-DRC, some from IRRI-Kenya and landraces. The third cluster were mainly composed by genotypes from IRRI-Kenya.

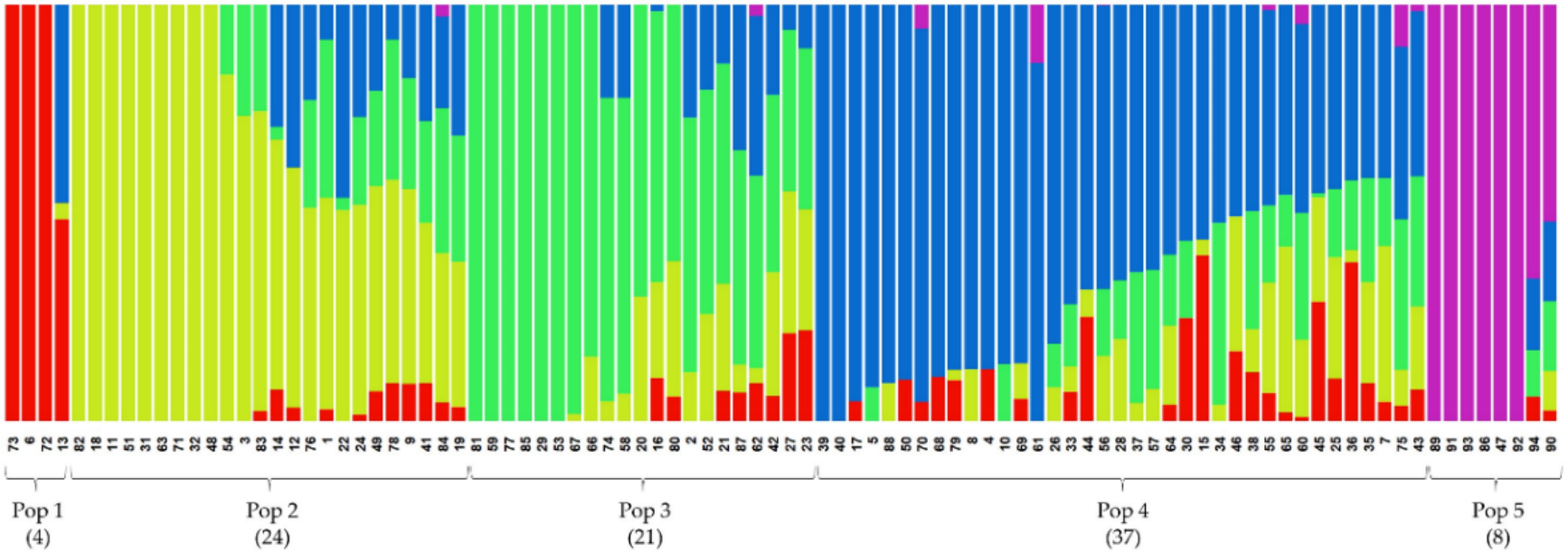

Figure 4.2: Population Structure of 94 Rice Genotypes Based on Variation in SNP DArTseq Markers with K = 5

Key: Each genotype is represented by a vertical bar that is segmented into K colors, indicating the likelihood of membership to each cluster. The genotype codes used are the same as those in Table 3.1.

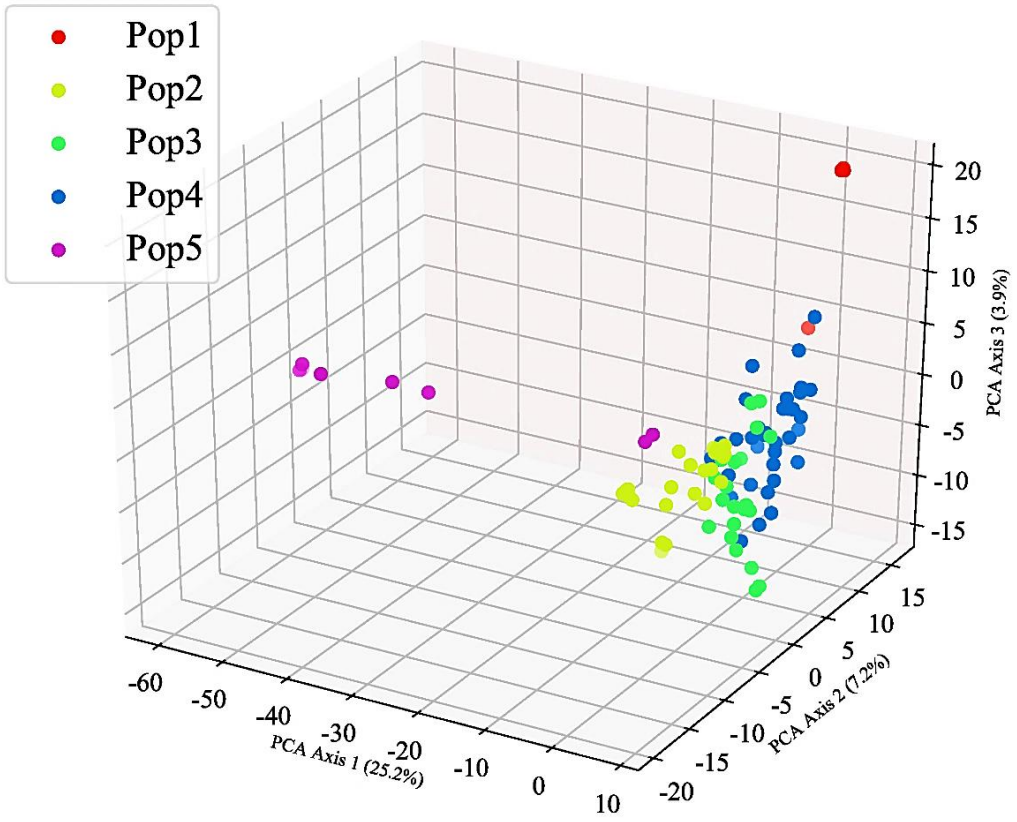

Figure 4.3: PCA Illustrating Relatedness and Distribution of the 94 Rice Genotypes Based on 8389 SNP Markers

Key: Each color corresponding to a specific sub-population from the ADMIXTURE results.

4.1.3 Genetic Diversity and Phylogenetic

The analysis of molecular variance (AMOVA) conducted on the 94 rice genotypes revealed highly significant genetic differences ($p < 0.001$) between sub-populations and within genotypes. However, no significant difference ($p > 0.001$) was observed between genotypes within sub-populations, as shown in Table 4.2. Among the total genetic variations observed in the 94 rice genotypes, 36.09% was attributed to genetic differentiation between the sub-populations, 34.04% to genetic differentiation within the genotypes, and the remaining proportion (29.87%) was due to genetic differences between genotypes within the sub-populations. Additionally, the overall fixation index (F_{st}) and number of migrants (Nm) among the sub-populations were 0.52 and 0.23, respectively (Table 4.2).

Table 4.2: Analysis of Molecular Variance among 94 Genotypes Based on SNPs

Source	DF	Estimated variance	Proportion of variation (%)	p-value	F_{st}	Nm
Between sub-populations	4	892.93	36.09	<0.001	0.52	0.23
Between genotypes within sub-population	89	738.89	29.87	>0.001		
Within genotypes	94	842.2	34.04	<0.001		
Total	**187**	**2474.03**	**100**			

DF: Degree of freedom; F_{st}: Fixation index; Nm: Number of migrants

The percentage of polymorphic loci per population (PPL) ranged from 21.28% (Pop1) to 85.64% (Pop4), with an average of 54.23% (Table 4.3). The mean values for the expected heterozygosity (H_e), observed heterozygosity (H_o), unbiased expected heterozygosity (uH_e), and inbreeding coefficient (F_{is}) were 0.14, 0.08, 0.15 and 0.52, respectively. The gene diversity values, calculated as expected heterozygosity (H_e) in population varied from 0.08 (Pop 1) to 0.21 (Pop 4). The observed heterozygosity (Ho) value ranged from 0.02 (Pop5) and 0.13 (Pop4), whereas Inbreeding coefficient varied from 0.31 (Pop2) to 0.75 (Pop5).

Table 4.3: Genetic Diversity of Five Rice Sub-Populations Based on 8389 SNP Markers

	Sub-population size	PPL	H_e	H_o	uH_e	F_{is}
Pop1	4	21.28	0.08	0.05	0.09	0.52
Pop2	24	69.14	0.15	0.11	0.16	0.31
Pop3	21	65.36	0.18	0.07	0.19	0.62
Pop4	37	85.64	0.21	0.13	0.22	0.38
Pop5	8	29.73	0.09	0.02	0.10	0.75
Average		54.23	0.14	0.08	0.15	0.52

PPL: Percentage of Polymorphic Loci; H_o: Observed heterozygosity; H_e: Expected heterozygosity; uH_e: unbiased expected heterozygosity; F_{is}: inbreeding coefficient.

The population pairwise fixation indexes, presented in Table 4.4, estimates genetic differentiation among populations due to genetic structure. Pop5 showed greater genetic distance from Pop1, Pop2, Pop3, and Pop4 with F_{st} of 0.83, 0.78, 0.75 and 0.73, respectively. The minimum genetic distances were observed between Pop4 and

Pop3 (F_{st} = 0.07), between Pop3 and Pop2 (F_{st} = 0.11), and between Pop4 and Pop2 (F_{st} = 0.13).

Table 4.4: Population's Pairwise Genetic Differentiation Index (F_{st})

	Pop1	Pop2	Pop3	Pop4	Pop5
Pop1	0				
Pop2	0.43	0			
Pop3	0.33	0.11	0		
Pop4	0.26	0.13	0.07	0	
Pop5	0.83	0.78	0.75	0.73	0

The neighbor-joining tree (Figure 4.4) illustrates genetic relatedness among the five sub-populations. The analysis resulted in the formation of three distinct groups. The first group consisted of Pop2, Pop3, and Pop4. The second group comprised only Pop1, while the third group was composed of Pop5.

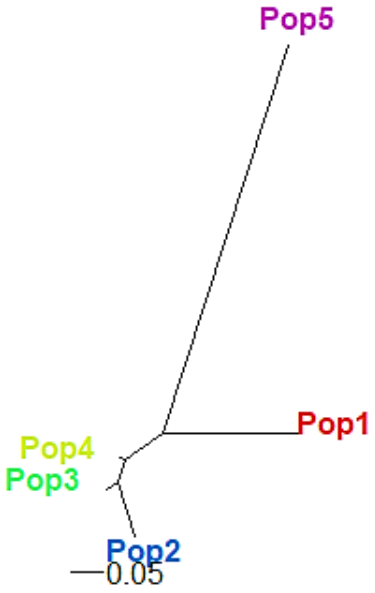

Figure 4.4: Neighbor-Joining Tree of Five Sub-Populations Based on the F_{st} Values

The Euclidean genetic distance among the 94 rice genotypes based on SNP markers ranged from 0.00 to 1.60, with an average of 0.87 (Appendix IV). The lowest genetic distances were observed in Pop5 between NERICA2 and NERICA10 (0.13), in Pop1 between Magoti and Runingu (0.15), Magoti and Jasmine (0.19), Jasmine and Runingu (0.2), in Pop2 between ARS755-3-3-1-B and ARS168-1-B-3-B (0.18). The highest genetic distance was exhibited between NERICA-L-19-sab-1 (from AfricaRice, Pop2) and NERICA12 (from IRRI-Kenya, Pop5), between NERICA-L-19-sab-1 (from AfricaRice, Pop2) and ARICA4 (from IRRI-Kenya, Pop5), between WAHX14N-926 (from AfricaRice, Pop3) and ARICA4 (from IRRI-Kenya, Pop5). The resulting genetic distance matrix was used to construct a Neighbor-Joining tree that classified the 94 rice genotypes into 2 distinct clusters (Figure 4.5). Cluster1 were made by Pop5, while Cluster2 grouped together genotypes from the rest of the sub-populations.

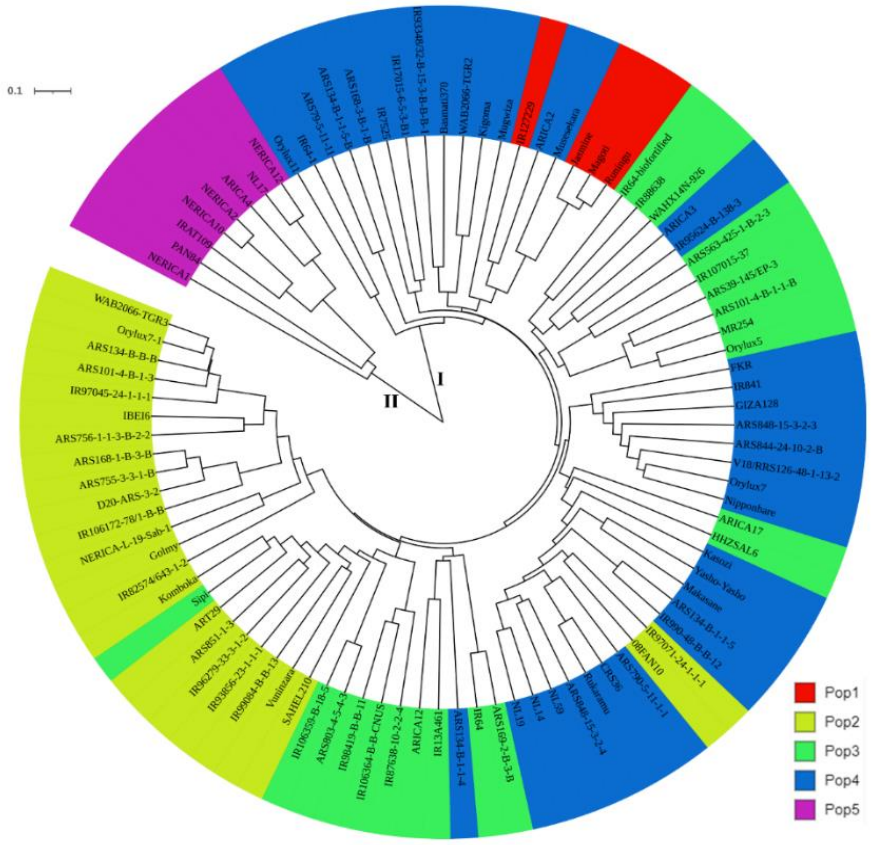

Figure 4.5: Phylogenetic Tree Showing Clustering Pattern among 94 Rice Genotypes Based on SNP Markers

The colors are based on the five sub-populations from ADMIXTURE results.

4.2. Identification of the Key Selection Criteria Associated with High Grain Yield among the Rice Germplasm Accessions

4.2.1 Analysis of Genetic Parameters

The results of the analysis of the genetic parameters presented in Table 4.5 revealed that the Phenotypic coefficient of variation (PCV) was higher than the corresponding Genotypic coefficient of variation (GCV) for all the traits.

Low PCV and GCV (values <10 %) were obtained with the days to flowering, days to maturity, panicle length, 1000grains weight, grain length, and grain yield although moderate PCV and GCV (10-20 %) were obtained with plant height, number of productive tillers per hill, number of spikelets per panicle, number of filled grains per panicle and panicle weight (Table 4.5). The number of primary branches/panicle showed low GCV (9.18 %) and moderate PCV (14.04 %) while a moderate GCV (19.25 %) with a high PCV (21.80 %) were obtained for grain width. High PCV and GCV (values > 20 %) were shown by grain length to grain width ratio. The GCV were greater than the corresponding environmental coefficient of variation (ECV) for days to flowering, days to maturity, plant height, panicle length, number of spikelets/panicle, number of filled grains/panicle, panicle weight, 1000 grains weight, grain width and grain length to grain width ratio. Number of primary branches/panicle, grain length and grain yield showed greater ECV than their corresponding GCV. Although equal GCV and ECV were observed on the number of productive tillers/hill.

The heritability in broad sense is categorized as low (< 30%), moderate (30% - 60%) and high (> 60%) (Robinson et al., 1949). For this study, it ranged from 13.35% to 89.84% (Table 4.5). A low heritability was observed for grain length (13.35%) while a moderate was obtained with panicle weight (59.34%), followed by the number of productive tillers/hill (50.04%), followed by grain yield (43.19%) and the number of primary branches/panicle (42.75%). High heritability was shown by days to maturity (89.84%), followed by days to flowering (88.58%), plant height (87.75%), grain length to grain width ratio (78.69%), grain width (77.98%), panicle length (71.31%), number of spikelet/panicle (68.70%), 1000 grains weight (66.64%), and number of filled grains/panicle (61.45%).

Table 4.5: Analysis of Genetic Parameters for Yield and Yield Components among 49 Rice Genotypes

Character(s)	MSg	MSe	σ^2g	σ^2p	ECV	GCV	PCV	h^2bs	GA	GAM
Days to flowering	106.83	4.40	34.14	38.55	1.90	5.31	5.64	88.58	11.33	10.29
Days to maturity	127.83	4.65	41.06	45.71	1.45	4.32	4.55	89.84	12.51	8.43
Plant height (cm)	275.86	12.27	87.86	100.13	3.72	9.94	10.61	87.75	18.09	19.18
Number of productive tillers/hill	18.44	4.61	4.61	9.22	13.46	13.47	19.05	50.04	3.13	19.63
Panicle length (cm)	8.38	0.99	2.46	3.45	4.02	6.33	7.50	71.31	2.73	11.02
Number of primary branches/panicle	4.07	1.25	0.94	2.19	10.62	9.18	14.04	42.75	1.30	12.36
Number of spikelet/panicle	2158.54	284.56	624.66	909.22	9.08	13.45	16.23	68.70	42.68	22.97
Number of filled grains/panicle	1909.78	330.28	526.50	856.78	10.80	13.64	17.40	61.45	37.05	22.03
Panicle weight (g)	0.68	0.13	0.18	0.31	10.35	12.50	16.23	59.34	0.68	19.84
1000 grains weight (g)	12.93	1.85	3.69	5.54	4.90	6.92	8.48	66.64	3.23	11.64
Grain length (mm)	0.26	0.17	0.03	0.20	4.72	1.85	5.07	13.35	0.12	1.39
Grain width (mm)	0.79	0.07	0.24	0.31	10.23	19.25	21.80	77.98	0.89	35.02
Grain length to grain width ratio	2.37	0.20	0.72	0.92	11.95	22.97	25.89	78.69	1.55	41.97
Grain yield (kg/m^2)	0.18	0.06	0.04	0.10	6.59	5.75	8.75	43.19	0.28	7.78

MSg: mean of squares for genotype; MSe: mean of squares for error; σ^2g: genotypic variance; σ^2p: phenotypic variance; ECV: environment coefficient of variance; GCV: genotypic coefficient of variance; PCV: phenotypic coefficient of variance; h^2bs: Broad sense heritability (%); GA: genetic advance (%); GAM: genetic advance as percentage of the mean

.

Genetic advance as a percent of the mean (GAM) classified as low (<10%), moderate (10%-20%), and high (>20%) (Johnson et al., 1955). In this study, it ranged from 1.39% to 41.97% (Table 4.5). A low GAM was shown by the grain length (1.39%), followed by grain yield (7.78%), and days to maturity (8.43%). Moderate GAM was obtained with days to flowering (10.29%), panicle length (11.02%), 1000 grains weight (11.64%), number of primary branches/panicle (12.36%), plant height (19.18%), number of productive tillers/hill (19.63%) and panicle weight (19.84%). High GAM was recorded with Grain length to grain width ratio (41.97%), followed by grain width (35.02%), then the number of spikelet/panicle (22.97%) and the number of filled grains/panicle (22.03%).

4.2.2 Principal Components Analysis

Data presented in Appendix III was used to perform principal components analysis. Components with eigen values greater than one were considered significant for discriminating the evaluated genotypes. The eigen values ranged from 4.88 (PC1) to 1.47 (PC4) (Table 4.6). The first four principal components accounted for 78.71% of total variation among 49 genotypes based on 14 morphological characters studied. It also indicates that the first principal component contributed 34.85% of the total variation. The second component accounted for 19.84% of the total variation. The third and the fourth principal components accounted for 13.53% and 10.48% of the total variance, respectively.

However, all measured traits have positively contributed to the variation in the first principal component except the 1000 grains weight (-0.004) and grain length to grain width ratio (-0.290) which negatively contributed to the variation. Although the major contributors to the variation in the first PC are panicle length (0.887), number of spikelets/panicle (0.881), number of filled grains/panicle (0.871), Number of primary branches/panicle (0.734), plant height (0.701), panicle weight (0.679), days to flowering (0.609), and Days to maturity (0.565). In the second principal component, the number of productive tillers/hill (0.777) and grain yield (0.764) were the most positively contributing characters.

Table 4.6: Contribution of 14 Traits to the Variation among 49 Rice Genotypes

Traits	PC1	PC2	PC3	PC4
Days to flowering	0.609	-0.474	-0.017	-0.364
Days to maturity	0.565	-0.500	-0.040	-0.387
Plant height (cm)	0.701	-0.312	0.408	-0.047
Number of productive tillers/hill	0.298	0.777	-0.187	0.095
Panicle length (cm)	0.887	0.032	0.289	0.015
Number of primary branches/panicle	0.734	0.353	-0.086	0.022
Number of spikelets/panicle	0.881	0.049	0.120	-0.083
Number of filled grains/panicle	0.871	0.098	0.076	-0.097
Panicle weight (g)	0.679	0.288	-0.069	0.502
1000 grains weight (g)	-0.004	-0.395	0.512	0.605
Grain length (mm)	0.030	-0.553	0.213	0.606
Grain width (mm)	0.341	-0.436	-0.749	0.314
Grain length to grain width ratio	-0.290	0.384	0.838	-0.147
Grain yield (kg/m^2)	0.292	0.764	-0.073	0.226
Eigenvalue	4.88	2.78	1.9	1.47
Percentage of variance	34.85	19.84	13.53	10.48
Cumulative percentage of variance	34.85	54.69	68.22	78.71

In the third principal component traits such as grain length to grain width ratio (0.838) and 1000 grains weight (0.512) explained significant and positive contribution to the variation. The important traits contributing positively and significantly to the variation in the fourth principal component were Grain length (0.606), 1000 grains weight (0.605), and panicle weight (0.502).

Two diverse major groups were formed based on the major yield components. The first includes representative genotypes such as 2 (IR64), 48 (D20-ARS-3-2) and 29 (IR13A461) exhibiting high values of grain yield, number of productive tillers/hill, number of primary branches/panicle, panicle weight, and number of filled grains/panicle. While the second was formed representatively by genotypes 13 (IR127229), 6 (Jasmine), and 44 (ARICA2) were commonly characterized by high values of plant height, days to flowering, days to maturity and grain width.

The PCA plot projections illustrate contribution and association among the variables (Figure 4.6), and Figure 4.7 illustrates the variation among the genotypes, showing how they are distributed along with PC1 and PC2.

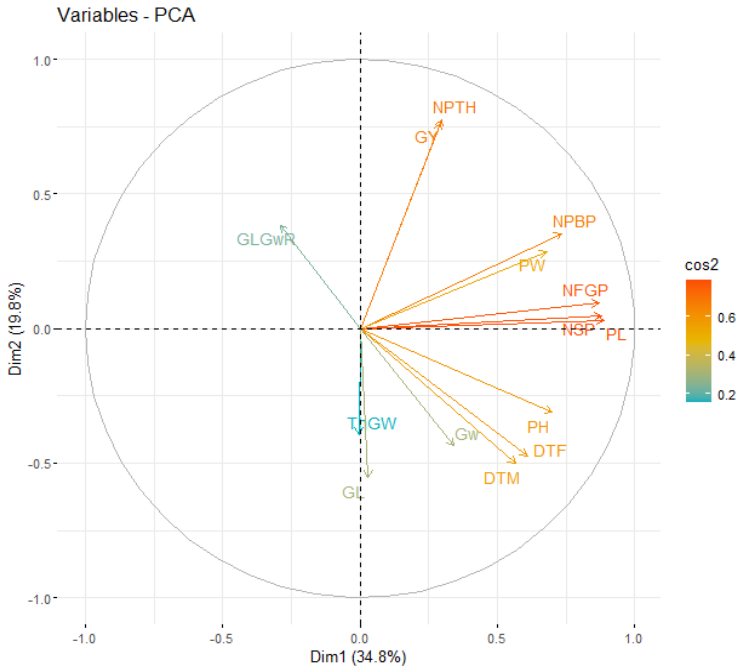

Figure 4.6: Scatter Plot Showing Contribution of the 14 Variables to the Variation Observed in PC1 and PC2.

The variable abbreviations are as presented in Table 3.2.

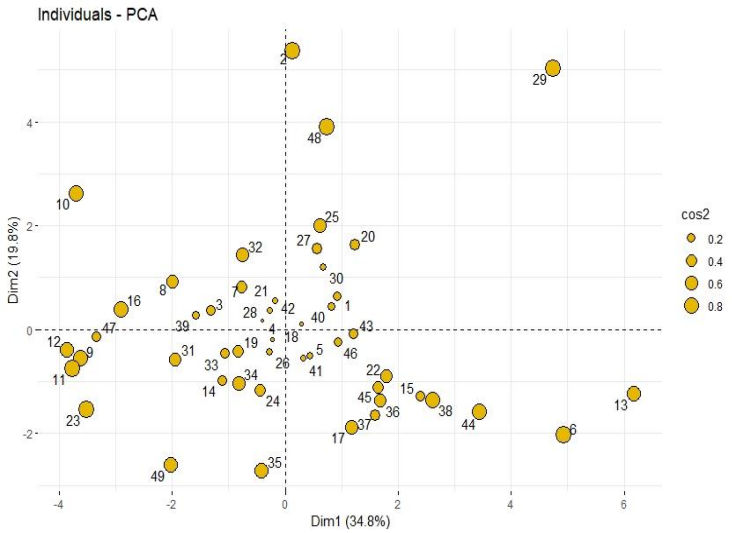

Figure 4.7: Scatter Plot Showing the Variation Observed among 49 Rice Genotypes Considering PC1 and PC2

The genotype names corresponding to the numbers are found in the Appendix I.

4.3.3 Phenotypic Correlation and Path Analysis

The phenotypic correlation (Figure 4.8) revealed a significant and positive correlation between grain yield and number of productive tillers per hill (r = 0.7), number of primary branches per panicle (r = 0.46), panicle weight (r = 0.45), and number of filled grains per panicle (r = 0.3). Days to flowering exhibited a highly and positive correlation with days to maturity (r = 0.79), plant height (r = 0.5), panicle length (r = 0.49), number of spikelets per panicle (r = 0.43), number of filled grains per panicle (r = 0.4), and number of primary branches per panicle (0.34), but significant and negatively correlated with grain length to grain width ratio (r = -0.3).

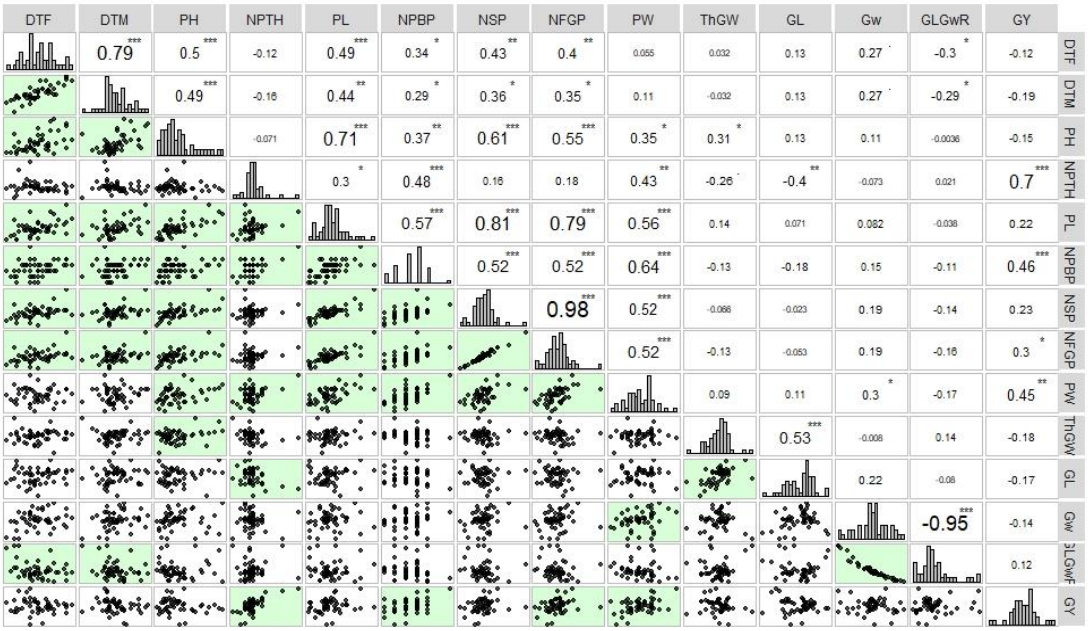

Figure 4.8: Correlation between Yield and Yield Related-Traits

DTF: Days to Flowering, DTM: Days to maturity, PH: Plant Height, NPTh: Number of productive Tillers/hill, PL: Panicle Length, NPBp: Number of primary branches/panicle, NSP: Number of Spikelet/panicle, NFGp: Number of filled grains/panicle, PW: Panicle weight, ThGW: Thousand Grains Weight, GY: Grain yield, GL: Grain Length, Gw: Grain Width and RGLGw: Ratio Grain Length/Grain Width; *, ** and ***: significance at 0.05 and 0.01 and 0.001 levels, respectively.

Days to maturity reflected a significant and positive relationship with plant number of filled grains per panicle (r = 0.35), and number of primary branches per panicle (r = 0.29). A significant and negative association was observed between days to maturity and grain length to grain width ratio (r = -0.29). Plant height showed a highly significant and positive relationship with panicle length (r = 0.71), number of spikelets

per panicle (r = 0.61), number of filled grains per panicle (r = 0.55), number of primary branches per panicle (r = 0.37) but significant and positive correlation with panicle weight (r = 0.35) and 1000grains weight (r = 0.31).

A positive and highly significant correlation was observed between the number productive of tillers per hill and the number of primary branches per panicle (r = 0.48), panicle weight (r = 0.43) but a significant and positive correlation was exhibited with the panicle length (r = 0.3) while significantly and negatively correlated with grain length (r = -0.4). Panicle length showed a highly significant and positive relationship with the number of spikelets per panicle (r = 0.81), the number of filled grains per panicle (r = 0.79), the number of primary branches per panicle (r = 0.57), and panicle weight (r = 0.56).

A positive and highly significant correlation was observed between the number of primary branches per panicle with panicle weight (r = 0.64), the number of spikelets per panicle (r = 0.52), and the number of filled grains per panicle (r = 0.52). The number of spikelets per panicle reflected a highly significant and positive correlation with the number of filled grains per panicle (r = 0.98) and panicle weight (r = 0.52). The number of filled grains per panicle exhibited a highly significant and positive association with panicle weight (r = 0.52). Panicle weight showed significant and positive relationship with grain width (r = 0.3). 1000grains weight reflected a highly significant and positive correlation with grain length (r = 0.53) while grain width showed a highly significant and negative correlation with grain length to grain width ratio (r = -0.95).

Phenotypic path analysis (Table 4.7) was performed to assess the direct and indirect effects of the yield components on grain yield. It revealed a strong and positive direct effect on grain yield by the number of productive tillers per hill (0.38), followed by the number of filled grains per panicle (0.38), panicle weight (0.34), weak and direct effect for days to flowering (0.08), grain length (0.08), 1000grains weight (0.07), and the number of primary branches per panicle (0.05). While negative direct effects on yield were observed on grain width (-0.52), grain length to grain width ratio (-0.33), plant height (-0.22), panicle length (-0.15), and days to maturity (-0.11).

Table 4.7: Phenotypic Path Analysis of Yield and Yield Related-Traits among 49 Rice Genotypes

	DTF	DTM	PH	NPTh	PL	NPBp	NSp	NFGp	PW	ThGW	GL	Gw	GLGwR
DTF	0.08	-0.08	-0.10	-0.03	-0.06	0.01	-0.01	0.11	0.02	0.00	0.00	-0.11	0.09
DTM	0.06	-0.11	-0.10	-0.06	-0.05	0.01	-0.01	0.10	0.03	0.00	0.01	-0.13	0.10
PH	0.04	-0.05	-0.22	-0.03	-0.09	0.01	-0.02	0.18	0.10	0.02	0.01	-0.05	0.00
NPTh	-0.01	0.02	0.02	0.38	-0.04	0.02	-0.01	0.08	0.14	-0.01	-0.02	0.02	0.00
PL	0.04	-0.04	-0.14	0.10	-0.15	0.03	-0.02	0.27	0.20	0.01	0.00	-0.05	0.02
NPBp	0.02	-0.02	-0.06	0.18	-0.08	0.05	-0.02	0.19	0.21	-0.01	-0.02	-0.09	0.06
NSp	0.03	-0.03	-0.12	0.07	-0.11	0.02	-0.03	0.37	0.19	0.00	0.00	-0.10	0.05
NFGp	0.02	-0.03	-0.11	0.08	-0.11	0.02	-0.03	0.38	0.19	-0.01	-0.01	-0.09	0.05
PW	0.00	-0.01	-0.06	0.15	-0.09	0.03	-0.02	0.21	0.34	0.00	0.00	-0.15	0.06
ThGW	0.00	0.00	-0.06	-0.08	-0.02	-0.01	0.00	-0.03	0.02	0.07	0.03	0.00	-0.05
GL	0.00	-0.01	-0.02	-0.12	0.00	-0.01	0.00	-0.03	0.02	0.03	0.08	-0.07	-0.03
Gw	0.02	-0.03	-0.02	-0.01	-0.01	0.01	-0.01	0.07	0.09	0.00	0.01	-0.53	0.32
GLGwR	-0.02	0.03	0.00	0.00	0.01	-0.01	0.00	-0.06	-0.06	0.01	0.01	0.48	-0.35

DTF: Days to Flowering, DTM: Days to maturity, PH: Plant Height, NPTh: Number of productive Tillers/hill, PL: Panicle Length, NPBp: Number of primary branches/panicle, NSP: Number of Spikelet/panicle, NFGp: Number of filled grains/panicle, PW: Panicle weight, ThGW: Thousand Grains Weight, GY: Grain yield, GL: Grain Length, Gw: Grain Width and RGLGw: Ratio Grain Length/Grain Width.

The indirect effect of all the traits on grain yield was not significant. However, considerable positive indirect effects on yield were observed in the number of spikelets per panicle (0.37) and panicle length (0.28) via the number of filled grains per panicle. Positive indirect effect for the number of spikelets via panicle weight (0.19), panicle length (0.2), and the number of filled grains per panicle (0.19) all via panicle weight. A considerable positive indirect effect on yield was also reflected by the number of primary branches per panicle via panicle weight (0.21) and via the number of filled grains per panicle (0.19).

4.4 Identification of aromatic genotypes and volatile organic compounds associated to the rice grain aroma trait

4.4.1 Classification of Rice Genotypes Based on Sensory Test and 2-Acetyl-1-Pyrroline Content

The sensory test on cooked brown flour revealed distinct difference in strength of aroma among the 86 rice genotypes evaluated; five genotypes were classified as aromatic, 32 genotypes as semi-aromatic, and 49 genotypes as non-aromatic (Table 4.8). A targeted metabolomics allowed the identification of the 2-AP in the rice samples. The level of 2-AP was below the detection limit in all the 50 non-aromatic rice genotypes, but detectable in all the five aromatic and in 14 semi-aromatic genotypes (Table 4.8). The 2-AP content in the aromatic class of rice ranged from 0.41 mg/kg in ARS169-2-B-3-B to 1.4 mg/kg in Basmati 370, with an average of 0.85mg/kg. In the semi-aromatic class, 2-AP concentration varied between 0.16 mg/kg (IR107015-37) and 0.74 mg/kg (ARS79-5-11-11).

Table 4.8: Sensory-Based Classification of Rice Genotypes and their 2-Acetyl-1-Pyrroline (2-AP) Contents

Entry No.	Genotype name	Class	2-AP content (mg/kg)	Entry No.	Genotype name	Class	2-AP content (mg/kg)
1	Komboka	Semi-aromatic	0.26±0.08	23	ARICA2	Semi-aromatic	nd
2	IR64	Semi-aromatic	0.18±0	24	NL19	Non aromatic	nd
3	IBEI6	Non aromatic	nd	25	NL14	Non aromatic	nd
4	GIZA128	Semi-aromatic	0.51±0.1	26	NL17	Semi-aromatic	nd
5	Nipponbare	Non aromatic	nd	27	D20-ARS-3-2	Semi-aromatic	nd
6	Jasmine	Aromatic	nd	28	IR96279-33-3-1-2	Non aromatic	nd
7	NL59	Non aromatic	nd	29	ARS134-B-1-1-5-B	Semi-aromatic	0.66±0.11
8	FKR	Non aromatic	nd	30	Orylux7-1	Non aromatic	nd
9	08FAN10	Non aromatic	nd	31	WAHX14N-926	Non aromatic	nd
10	WAB2066-TGR2	Aromatic	0.48±0.25	32	MR254	Non aromatic	nd
11	WAB2066-TGR3	Non aromatic	nd	33	Golmy	Non aromatic	nd
12	IR99084-B-B-13	Semi-aromatic	nd	34	ARS848-15-3-2-4	Non aromatic	nd
13	IR127229	Semi-aromatic	nd	35	IR93348:32-B-15-3-B-B-B-1	Non aromatic	nd
14	IR106172-78 :1-B-B	Non aromatic	nd	36	ARS168-3-B-1-B	Non aromatic	nd
15	ARS848-15-3-2-3	Semi-aromatic	0.33±0.2	37	IR88638	Semi-aromatic	nd
16	IR106364-B-B-CNUS	Non aromatic	nd	38	ARICA12	Semi-aromatic	0.39±0.01
17	ARS844-24-10-2-B	Non aromatic	nd	39	ARICA3	Semi-aromatic	nd
18	ARS168-1-B-3-B	Semi-aromatic	nd	40	IR64-sub-1	Non aromatic	nd
19	ARS851-1-3	Non aromatic	nd	41	HHZSAL6	Non aromatic	nd
20	IR87638-10-2-2-4	Semi-aromatic	nd	42	ARS755-3-3-1-B	Non aromatic	nd
21	IR98419-B-B-11	Non aromatic	nd	43	ARS134-B-1-1-5	Non aromatic	nd
22	IR97071-24-1-1-1	Non aromatic	nd	44	IR990-48-B-B-12	Non aromatic	nd
45	ARS803-4-5-4-3	Semi-aromatic	nd	66	IR64-biofortified	Semi-aromatic	nd
46	IR93856-23-1-1-1	Semi-aromatic	nd	67	IR107015-37	Semi-aromatic	0.16±0.04
47	ARS790-5-11-1-1	Non aromatic	nd	68	ARS79-5-11-11	Semi-aromatic	0.74±0.03
48	IR17015-6-5-3-B1	Semi-aromatic	nd	69	V18/RRS126-48-1-13-2	Semi-aromatic	0.27±0.07
49	IR106359-B-18-5	Semi-aromatic	nd	70	Orylux11	Non aromatic	nd
50	IR95624-B-138-3	Non aromatic	nd	71	ARS134-B-B-B	Non aromatic	nd
51	IR13A461	Non aromatic	nd	72	Magoti	Non aromatic	nd
52	Mugwiza	Non aromatic	nd	73	Runingu	Non aromatic	nd
53	Vuninzara	Non aromatic	nd	74	ARS169-2-B-3-B	Aromatic	0.41±0.09

Entry No.	Genotype name	Class	2-AP content (mg/kg)	Entry No.	Genotype name	Class	2-AP content (mg/kg)
54	IR97045-24-1-1-1	Non aromatic	nd	75	ARS134-B-1-1-4	Non aromatic	nd
55	Kigoma	Semi-aromatic	0.57±0.09	76	IR82574/643-1-2	Non aromatic	nd
56	Makasane	Non aromatic	nd	77	Orylux5	Non aromatic	nd
57	Rukaramu	Non aromatic	nd	78	SAHEL210	Non aromatic	nd
58	Musesekara	Non aromatic	nd	79	IR841	Non aromatic	nd
59	Yasho-Yasho	Semi-aromatic	nd	80	ARS39-145/EP-3	Non aromatic	nd
60	Kasozi	Semi-aromatic	nd	81	ARS101-4-B-1-1-B	Non aromatic	nd
61	IR7525	Semi-aromatic	nd	82	ARS101-4-B-1-3	Non aromatic	nd
62	Orylux7	Semi-aromatic	nd	83	NERICA-L-19-Sab-1	Non aromatic	nd
63	ART29	Semi-aromatic	0.21±0.03	84	ARS756-1-1-3-B-2-2	Semi-aromatic	0.24±0.07
64	Sipi	Semi-aromatic	0.44±0.04	85	ARS563-425-1-B-2-3	Aromatic	1.11±0.16
65	CRS36	Semi-aromatic	0.43±0.1	86	Basmati370	Aromatic	1.4±0.14

nd-: not detectable

4.4.2 Identification of Volatile Organic Compounds Associated with 2-AP

Genotypes from the aromatic and semi-aromatic classes were used to further investigate volatiles compounds using the untargerted metabolomic approach. The HS-SPME coupled with the GC-MS generated a total of 1834 chromatogram peaks which were detcted in 37 rice samples. The significant discriminating peaks among the two classes were identified using the Volcano plot analysis (Figure 4.9). Red points represent significantly different metabolites based on the results of fold change (FC) analysis and the t-test in positive and negative modes (FC > 2 and P < 0.05). Fold change is the mean value of peak area obtained from the aromatic class divide by the mean value of peak area obtained from the semi-aromatic class. When contrasting the aromatic and semi-aromatic classes, 216 peaks (blue and red) exhibited significant differences between the two groups while 1618 peaks were non significant. Among the 216 significant peaks, the intensity of 143 peaks (in red) were higher and 73 (in blue) were lower in the aromatic class of rice.

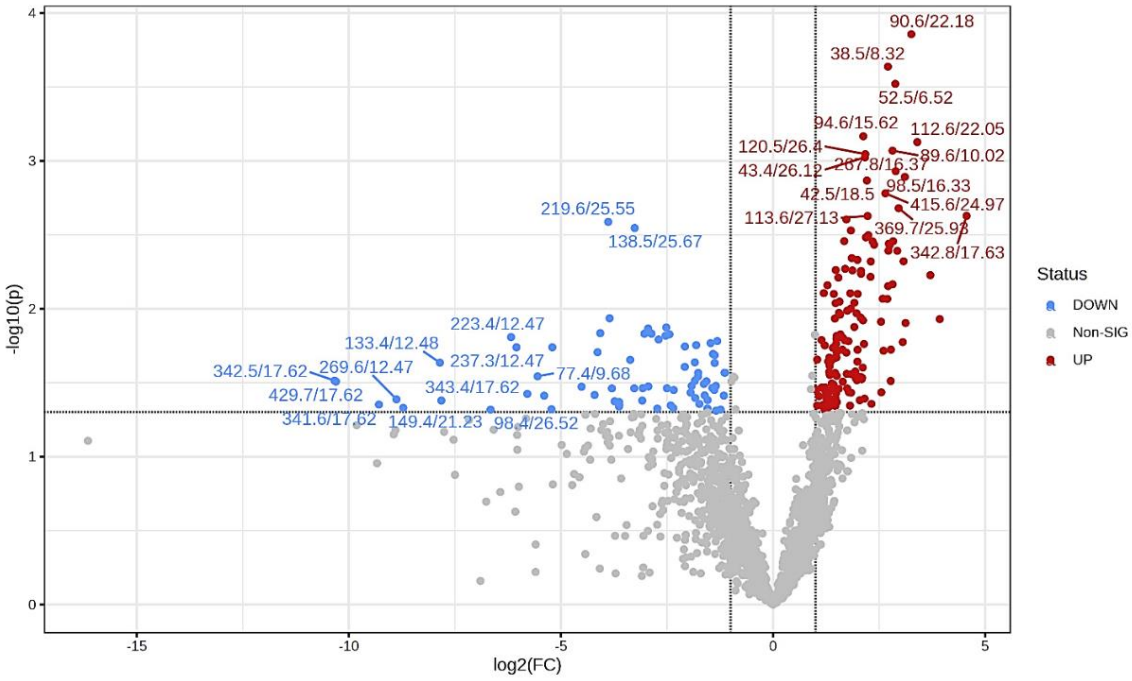

Figure 4.9: Volcano Plot Analysis of Aromatic vs Semi-Aromatic Class of Rice, Depicting Differential Chromatogram Peaks

Each data point on the plot corresponds to a specific peak. The x-axis represents the logarithm of the fold change in chromatogram peak abundance between two samples.

Further refinement was applied by selecting abundant peaks in the aromatic class, with a FC >2, VIP score > 0.8 and p-value < 0.05, resulting in a subset of 42 peaks including the 2AP-related peaks (111.5 m/z at 6.52 min) that exhibit substantial and statistically significant differences.

The correlation analysis was conducted to understand the association between the significant peaks. The heatmap of the correlation matrix was plotted to display the relationship between the 2-AP and other significant up-regulated peaks related to other compounds in rice samples ((Figure 4.10 (a)). Positive and highly significant correlation ($p < 0.005$) was found between 2AP-related peaks and 22 other peaks (Appendix V). No significant correlation was observed between 2-AP with 19 peaks. Details of the peaks positively associated with 2-AP are presented in Table 4.9, including the retention time, mass to charge ratio and peak intensities across the genotypes belonging to aromatic and semi-aromatic classes. The intensities (in Dalton) of significant peaks associated to 2AP Across the aromatic and semi-aromatic rice are presented in Appendix VI.

The heatmap showed distinct clustering of the genotypes with respect to the intensities of the significant peaks (Figure 4.10 (b)). The 37 samples were separated into three groups. Cluster I was constituted by 18 genotypes, including four aromatic genotypes (ARS563-425-1-B-2-3, ARS169-2-B-3-B, WAB2066-TGR2, and Jasmine) and 14 semi-aromatic genotypes (Kigoma, ART29, Sipi, ARS79-5-11-11, IR64, GIZA128, ARS848-15-3-2-3, V18RRS126-48-1-13-2, CRS36, ARICA12, ARS134-B-1-1-5-B, Komboka, IR107015-37, and ARS756-1-1-3-B-2-2) showing a moderate intensity of the peaks that were slightly correlated to 2-AP. Remarkably, 2-AP was low to moderate concentration in the samples of the cluster I. Thirty-two peaks, were had higher intensity in Basmati370 which formed the cluster II.

These included the peaks related to 2-AP and the most positively and highly correlated to it. Notably, Basmat370 showed also high 2-AP content what is believed to be a major volatile contributing to the rice flavor. Cluster III was composed by 18 genotypes, mostly characterized by features that are negatively correlated to 2-AP.

From the 22 signals correlated to 2-AP, 9 compounds were identified based on their retention indexes, and mass spectra fragmentation, while the other 13 were unknown (Figure 4.11). Details of the identified compounds are summarized in Table 4.9.

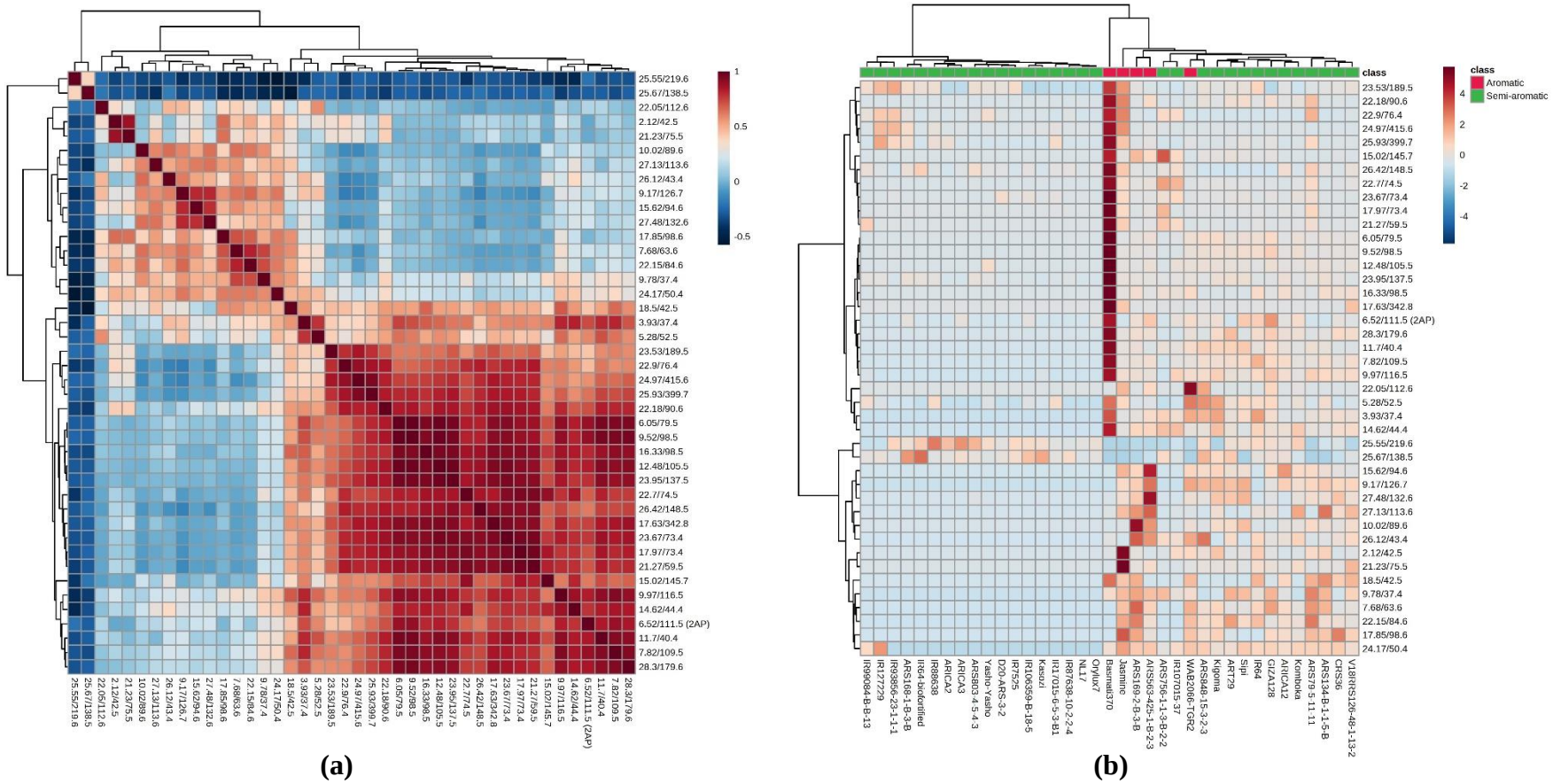

Figure 4.10: Heatmap Matrices

For (a) Correlation among Significant Up-Regulated Features; (b) between Rice Genotypes and Significant Up-Regulated Features. The red-colored and blue-colored squares indicate positive and negative associations, respectively.

These volatile compounds consisted of five aldehydes (n-Heptanal; Benzaldehyde; 1-Nonaldehyde; Decanal and Undecanal), three alcohols (Benzyl alcohol; 1-Decanol, 2-hexyl-; Isotridecanol-), one compounds with diverse functional groups (Mandelic acid, 2TBDMS derivative). All nine compounds identified showed large fold-changes when compared between the two classes of rice, were characterized with higher VIP scores (VIP > 1.0) and strongly correlated with higher and lower sensory ratings (p-value< 0.01, fold-change > 2). However, the results showed that the main volatiles, associated with 2-AP, found in the aromatic rice were Undecanal (FC = 13.06) and n-Heptanal (FC = 6.71).

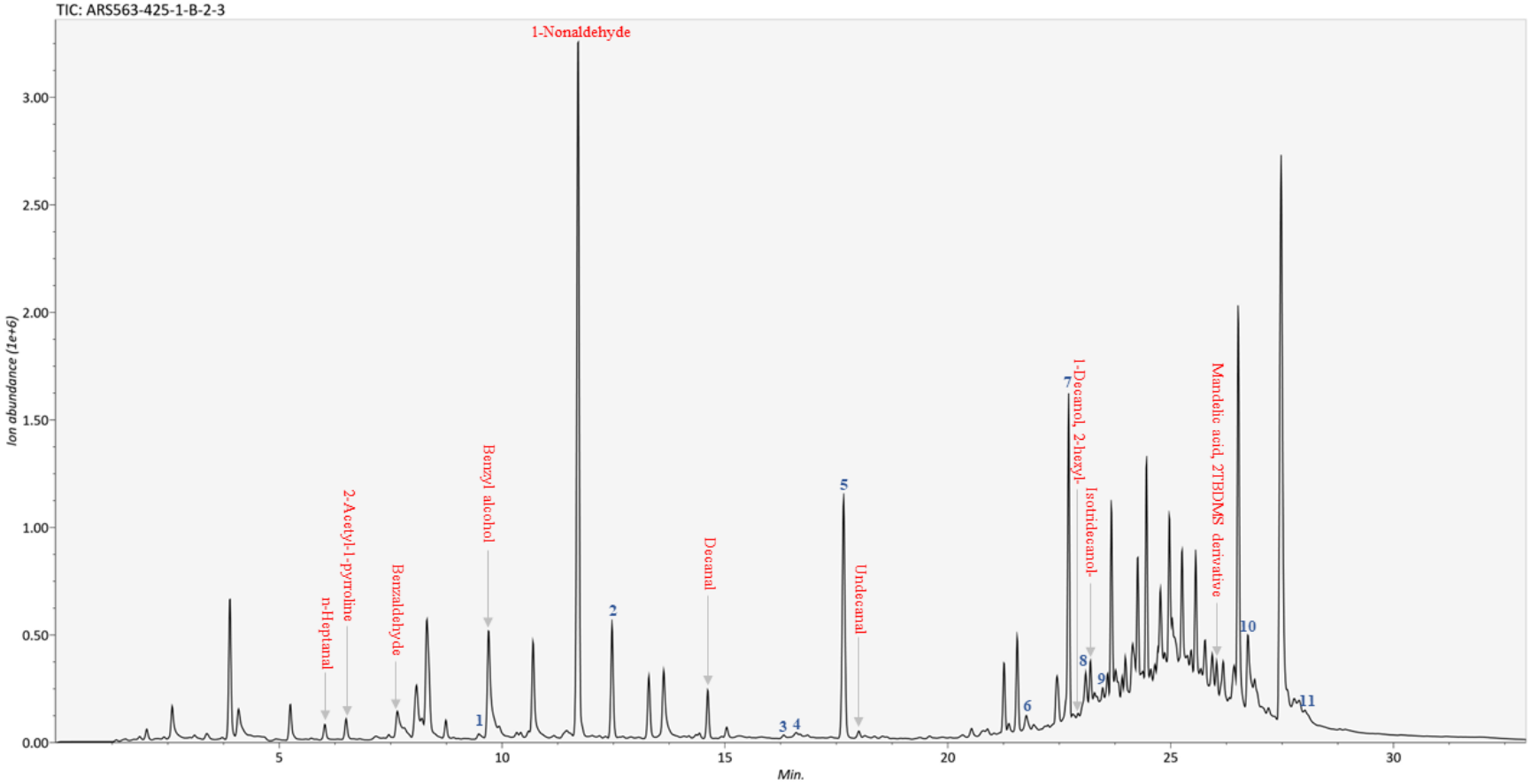

Figure 4.11: Chromatogram of an Aromatic Rice Depicting Significant Peaks Positively Associated to 2-AP

The identified compounds are labeled in red and the unknown are in blue.

Table 4.9: Characteristics of the Identified Volatile Organic Compounds Associated with 2-AP in Rice Samples

mz/rt	Metabolite name	Formula	Retention Index		VIP score	T-test (p-value)	FC
			Experiment	Reference			
79.5/6.05	n-Heptanal	$C_7H_{14}O$	900	897	2.77	0.0037	6.71
111.5/6.52	2-Acetyl-1-pyrroline	C_6H_9NO	918	918	2.41	0.0033	4.56
109.5/7.82	Benzaldehyde	C_7H_6O	961	963	2.76	0.0055	4.20
116.5/9.97	Benzyl alcohol	C_7H_8O	1039	1040	2.82	0.0054	3.25
40.4/11.7	1-Nonaldehyde	$C_9H_{18}O$	1095	1102	2.73	0.0058	4.21
44.4/14.62	Decanal	$C_{10}H_{20}O$	1184	1208	2.75	0.0035	3.20
73.4/17.97	Undecanal	$C_{11}H_{22}O$	1288	1279	2.71	0.0059	13.06
76.4/22.9	1-Decanol, 2-hexyl-	$C_{16}H_{34}O$	1481	1504	2.37	0.0061	4.91
137.5/23.95	Isotridecanol-	$C_{13}H_{28}O$	1566	1528	2.77	0.0037	5.21
399.7/25.93	Mandelic acid, 2TBDMS derivative	$C_{20}H_{36}O_3Si_2$	1828	1875	2.48	0.0048	4.93

mz/rt: mass to charge ratio/retention time; VIP: variable importance in the projection; FC: Fold change. The reference retention index values were obtained from https://webbook.nist.gov/ , accessed on the period between September 1st and 8th, 2023.

CHAPTER FIVE

DISCUSSION

5.1 Evaluation of Genetic Diversity and Population Structure of Rice Germplasm Accessions of the Eastern DRC

In this study, a total of 8,389 SNP markers passed quality control analyses designed to remove non-informative markers, as well as markers with a minor allele frequency of less than 5% and a call rate lower than 95%. This ensured that the selected markers were suitable for estimating the genetic diversity and structure of the 94 rice genotypes under investigation. It was observed that the number of SNPs generated using DArTseq technology in rice was higher compared to previous studies by Adeboye et al. (2020) and Thant et al. (2021), although lower than the findings reported by Ndjiondjop et al. (2018). Analysis of SNP marker distribution across the rice chromosomes revealed an average of 699 SNP markers per chromosome, indicating a wide distribution of markers. The abundance of polymorphic markers on chromosomes is commonly associated with the level of genetic diversity (Adeboye et al., 2020; Eltaher et al., 2018).

The PIC values for a set of genetic markers are a useful tool for evaluating the informativeness (usefulness) of these markers in population diversity studies (Botstein et al., 1980). In the current study, 42.4% of SNPs had a polymorphism information content (PIC) value between 0.5 and 0.25. According to Botstein et al. (1980), markers with PIC values greater than 0.5 are highly informative, those with PIC values between 0.5 and 0.25 are considered moderately informative, while markers with PIC values less than 0.25 remain to some extent informative. This implies that 42.4% of the SNP markers were moderately informative, which demonstrated the usefulness of these markers for genetic diversity analysis in the rice germplasm. However, the average SNP PIC value of 0.25 obtained in this study was slightly higher than what was found in previous studies in rice (Adeboye et al., 2020; Singh et al., 2013; Thant et al., 2021). According to Eltaher et al. (2018), due to their bi-allelic nature, SNP markers have limited informative value, resulting in low to moderate PIC values that are restricted to a maximum of 0.5. In this study a higher frequency of SNP transitions as compared

to that of SNP transversions. It is commonly observed that transitions are more frequent than transversions in true SNPs (Batley & Edwards, 2007; Boussaha et al., 2012; Lai et al., 2012). In fact, the ratio of transition to transversion frequencies is often used as a measure of evolutionary distance between species or individuals (Batley & Edwards, 2007). Moreover, it happens that in a set of three available SNPs, two of them are transitions while the third is a transversion (Batley & Edwards, 2007; Luo et al., 2019). Similar findings have also been reported on rice (Adeboye et al., 2020), maize (Kumar et al., 2022; Morton et al., 2006), *Camelina sativa* (Luo et al., 2019). According to Guo et al. (2017), transversions would have larger impacts in disrupting the transcription factors binding, leading to significant alterations in gene expression.

In the present study, 94 rice genotypes were assessed and classified into 5 sub-populations that displayed significant divergence among them, with varying degrees of diversity observed within each sub-population. It was observed that the composition of genotypes within each sub-population exhibited a dependence on their respective collection sources. The result highlights the source of the rice genotypes as a major factor influencing their genetic makeup, with Pop1 and Pop5 having a higher percentage of genotypes from specific sources (local landraces and IRRI-Kenya, respectively). However, Pop2, Pop3, and Pop4, predominantly made up of AfricaRice, IRRI-Burundi, and INERA-DRC genotypes, exhibited common genetic background. Xu et al. (2016) highlighted breeding efforts and Smith et al. (2021) pointed unrestricted movement across various institutions to contribute to the genetic similarities observed among the populations. Similarities between 22 diverse rice collections from different sources were also reported by Salem & Sallam, (2016).

The analysis of rice genotypes using the PCA plot revealed that the 94 genotypes could be grouped into 3 clusters. The distribution of the genotypes based on PCA was similar to that of the structure analysis using ADMIXTURE. These findings were consistent with previous reports (Mogga et al., 2018; Ndjiondjop et al., 2018). It has been noted that the first cluster comprised 3 genotypes, of which 2 were local landraces and 1 was from IRRI-Burundi. The second cluster was composed of genotypes from all 5 sources, and the third cluster was predominantly formed by IRRI-Kenya genotypes. This result

revealed the presence of common alleles among the genotypes within each cluster, which could be attributed to breeding activities such as selection for specific traits and hybridization (Xu et al., 2016).

The results of this study revealed that genetic diversity among the 94 rice genotypes were largely determined by the differentiation between sub-populations and within genotypes, which may be due to low genetic exchange and limited gene flow, respectively. According to Wright (1965), a number of migrants (Nm) value lower than one indicates limited gene flow among subpopulations. In this study, the observed Nm value was 0.23, which suggests limited genetic exchange and significant differentiation (F_{st} = 0.52) observed among sub-populations; consistent with the AMOVA. Since there are significant genetic differences among the sub-population, there is a potential for identifying and selecting diverse genotypes from different sub-populations for breeding programs (Salgotra & Chauhan, 2023). The genetic differences between genotypes within sub-populations contributed the least to the total genetic variation, indicating genetic similarity between the genotypes within sub-populations, which could be due to breeding practices, such as selection for specific traits or the use of parent lines with similar genetic backgrounds. This can be useful for establishing breeding populations with specific traits of interest (Li et al., 2014). The findings of this study corroborate those of previous research on rice genetic diversity and population structure, which had also reported a significant differentiation between sub-populations and within genotypes using SSR markers (Suvi et al., 2021), among genotypes from AfricaRice, IRRI and Tanzania which could be attributed to the different sources of materials. Nevertheless, the findings of this study disagree with those of Mogga et al. (2018), investigating diversity among landrace, upland and lowland rice collections using DArT-based SNPs who found low degree of differentiation among populations.

According to Eltaher et al. (2018) and Luo et al. (2019), genetic indices serve as indicators of genetic diversity. In this study, the mean H_o value (< 0.1) aligns with previous research on rice (Thant et al., 2021) (H_o = 0.03), (Mogga et al., 2018) (H_o = 0.0975), (Adeboye et al., 2020) (H_o < 1), but is slightly lower than findings by Suvi et al. (2021) (H_o = 0.17) using SSR markers. The lower H_o compared to H_e was expected

due to rice's self-pollinating nature (Sleper & Poehlman, 2006), resulting in a relatively higher degree of inbreeding within the population, as also supported by Ndjiondjop et al. (2018). Examining gene diversity based on PPL, H_o, H_e and F_{is} values, the genotypes within sub-populations (Pop2, Pop3, and Pop4) displayed diversity, suggesting possible occurrences of selection and hybridization among the accessions within these sub-populations. This genetic exchange contributes to increased diversity within the populations (Goulet et al., 2017). In contrast, Pop5 and Pop1 exhibited lower diversity, likely influenced by factors such as rice's inherent high inbreeding nature and specific selective breeding criteria employed by breeders on breeding lines forming Pop5, along with strict selection practices by farmers on the local landraces constituting Pop1.

Further, the F_{st} was used to quantify sub-population differentiation resulting from genetic structure. According to Wright (1968), F_{st} value of 0.25 or higher is considered significant in differentiating sub-populations, while values in the range of 0.15–0.25 indicate moderate differentiation. In contrast, differentiation is considered insignificant if the F_{st} value is 0.05 or less. Significant genetic differentiation was observed between all pairs of sub-populations, except for sub-populations 2 and 3, 2 and 4, and, 3 and 4, where insignificant genetic differentiation was observed. The lack of differentiation among sub-populations comprising genotypes from AfricaRice, IRRI-Burundi, and INERA-DRC may be due to the exchange of genetic materials, the effect of maintenance strategies, and selection by breeders across the involved institutes (Eltaher et al., 2018; Luo et al., 2019).

In this study, the average Euclidean genetic distance; a measure of genetic variation between pairs of genotypes, was found to be 0.87. This result is consistent with previous reports on elite rice genotypes from Chile (Becerra et al., 2015) and Ugandan rice genotypes (Mogga et al., 2018), which reported genetic distances of 0.87 and 0.86, respectively. The phylogenetic tree illustrates the distances between genotypes or groups, indicating their degree of relationship, with closely related groups positioned close to each other (Abaza, 2020). The NJ tree classified the 94 genotypes into two major groups, revealing a shared gene pool within each cluster. Cluster I comprised genotypes from AfricaRice, IRRI-Burundi, INERA-DRC, and the local landraces,

while cluster II mainly consisted of IRRI-Kenya genotypes. Low genetic distances observed in pairs of genotypes such as Magoti and Runingu, Magoti and Jasmine, Jasmine and Runingu, ARS755-3-3-1-B and ARS168-1-B-3-B, NERICA2 and NERICA10, suggest that these pairs were potentially collected separately but share a close genetic background. The high genetic distance observed between genotypes of different sub-populations may be attributed to the distinct genetic makeup of the IRRI-Kenya rice and local landraces collections, indicating differences from other sources, which aligns with the AMOVA results.

5.2 Selection Criteria for Grain Yield in Rice

To determine selection criteria for grain yield in rice, a set of 49 genotypes were randomly selected from the germplasm and assessed in two locations in Eastern DR Congo. The genotypes were significantly different ($P<0.001$) for all studied traits, which could be attributed to genetic diversity of the panel used, except for grain length. Similar results were reported by Abebe et al. (2017), indicating the existence of a wide range of genetic variation within the germplasm and revealing potential for genetic improvement through selection and hybridization for the evaluated traits. Rashid et al. (2017) also found similar results among 34 rice genotypes for all the traits they studied.

The high PCV compared to the corresponding GCV in this study explained the presence of environmental effects on the phenotypic expression of all the studied characters. Similar results were reported by Rashmi et al. (2017), Hannan et al. (2020), and Htwe et al. (2019), who have found lower GCV than the respective PCV, indicating the influence of environmental factors on the expression of the phenotypes.

However, the results showed high GCV than ECV for all the traits except number of primary branches/panicle, grain length, and grain yield. It could be noted the traits in the current study were mainly explained by genetic components and less affected by the environment. Studies reported by Abebe et al. (2017) and Girma et al. (2018) showed a small environmental influence on most of the yield components. Thus, selection based on the traits like days to flowering, days to maturity, number of productive tillers per hill, number of spikelets per panicle, number of filled grains per panicle, panicle weight, and 1000 grains weight could be effective in further

improvement. These findings were in accord with the results obtained by Dhakal et al. (2020).

Heritability is an important parameter for estimating the proportion of the phenotypic variation in a population that is explained by the genetic components (Dhakal et al., 2020). Heritability informs breeders on the magnitude of transmissibility of a particular trait while genetic advance estimates the measure of genetic gain during selection (Hannan et al., 2020). In the present study high heritability and genetic advance of the mean observed for grain length to grain width ratio, grain width, number of spikelet/panicle and number of filled grains/panicle, indicated that these traits are under high genetic control and less influenced by environment in their expression. therefore, improvement can be achieved effectively by direct selection based on these traits (Iqbal et al., 2018). Lipi et al. (2021) have also reported high heritability (97.47% and 97.39%) and genetic advance (52.45% and 64.7%) for the number of spikelet/panicle and the number of filled grains/panicle, respectively. A similar result was reported by Islam et al. (2015).

High heritability and moderate GAM reflected by days to flowering, plant height, panicle length, and 1000 grains weight, revealing that the characters are less influenced by environment in its expression and governed by both additive and non-additive gene action. This indicated a possibility of direct selection for the improvement of these traits (Girma et al., 2018).

High heritability coupled with low GAM days to maturity, showing non-additive gene action for the expressions of these characters. Direct selection for this trait might not be effective. Moderate heritability coupled with moderate GAM showed by panicle weight, the number of productive tillers/hill, and the number of primary branches/panicle, implies that improvement can be made through simple selection (Abebe et al., 2017). Grain yield showed moderate heritability and low GAM. This informed that this trait is totally governed by non-additive gene action and highly affected by the environment. Thus, heterosis breeding could be used for such traits (Girma et al., 2018).

Low heritability and low GAM exhibited by grain length, indicated that direct selection for this trait is not effective. Therefore, methods of selection based on families and progeny testing are more effective and efficient (Abebe et al., 2017).

Principal component analysis was used to identify the contribution of the variables (traits) towards variation observed in a given population (Kashyap & Yadav, 2020). It is very important in the selection procedure of the breeding program because it helps to identify the traits which have a great impact on the phenotype of the rice germplasm accessions (Burman et al., 2021).

The results of the principal component analysis showed that the first four principal components with eigen values greater than one accounted for 78.7% of the total variation. The characters associated with these four components are more useful in differentiating the genotypes. Similar results were also reported by Tonegnikes et al. (2019). The major discriminatory characteristics among the genotypes are panicle length, number of spikelets/panicle, number of filled grains/panicle, plant height, and number of productive tillers/hill. This is similar to the findings of Sudeepthi et al. (2020) for panicle weight and plant height. Therefore, selection of improved lines can be based on these morphological traits.

Grain yield is a complex trait in cereals, controlled by several genes and influenced more by the environment. Rice grain yield is determined by a combination of direct and indirect actions of the yield components (Li et al., 2014; Suvi et al., 2021). In rice improvement, the direct selection of genotypes based on yield may mislead the breeding program. Determining the direct and indirect effects of various traits on grain yield is critical in determining an appropriate selection criterion for high grain yield (Oladosu et al., 2018). The present study revealed high and positive correlation between grain yield and the number of productive tillers per hill, number of primary branches per panicle, panicle weight, and number of filled grains per panicle. These traits are more reliable components of grain yield. Improvement of these traits through direct selection are more likely to lead to the overall improvement of grain yield. Similar results have been obtained by Akhi et al. (2016), and (Saleh et al., 2020) on the number of productive tillers per hill and the number of filled grains per panicle.

Rashmi et al. (2017) and Kafi et al. (2021) also reported a positive association between grain yield with panicle weight, filled grains per panicle, and grains per panicle. In a study by Tonegnikes et al. (2019), in Nigeria using Korean rice germplasm also confirmed positive association between yield and number of productive tillers per hill, panicle weight, number of filled grains per panicle, and number of spikelets per panicle. Andrew et al. (2014) have also reported a positive and significant association between grain yield and effective tiller number, grain weight per plant, and the number of grains per panicle. Evaluating Tanzania rice germplasm collections, Suvi et al. (2021) found similar results for the association between grain yield and number of productive tillers/hill (number of panicles per plant), number of filled grains/panicle (percentage-filled grains). Amegan et al. (2020) have also reported no relationship between grain yield with plant height and 1000grains weight. These results do not corroborate findings of Suvi et al. (2021) on 1000grains weight.

Path analysis revealed a strong and positive direct contribution of the number of productive tillers per hill, number of filled grains per panicle, and panicle weight to grain yield. Positive indirect effects on grain yield were shown by the number of spikelets per panicle, panicle length and the number of primary branches per panicle. (Rashmi et al., 2017) have also reported that panicle weight, number of effective tillers per plant, and filled grains per panicle as the major direct contributors to grain yield.

5.3 Identification of Aromatic Genotypes and Determination of Volatile Organic Compounds Associated with Rice Grain Aroma

With the hypothesis that volatile compounds other than the 2-AP could be contributing to the aroma in rice, this objective tended to identify aromatic genotypes and volatile compounds associated with aroma that could be utilized as parents and biochemical markers, respectively in breeding programmes aimed at improving aroma in elite lines. The sensory method has been applied to identify fragrant rice as described by David et al. (2019). Based on this technique, by comparing the aroma strength of three well-known varieties, the genotypes were classified into Non aromatic, semi-aromatic and aromatic composed by 49, 32 and five genotypes, respectively. This classification has also been used in other studies (Ndikuryayo et al., 2023).

It has been observed that 2-AP was not detectable in all 50 non-aromatic rice genotypes tested. This result underscores the significance of 2-AP as a key aromatic compound, primarily associated with the aroma and flavor of certain rice varieties. Non-aromatic rice varieties are typically devoid of 2-AP, which is a defining feature of aromatic and semi-aromatic rice.

Among the aromatic rice genotypes, the 2-AP content exhibited variation, with concentrations ranging from 0.41 mg/kg in ARS169-2-B-3-B to 1.4 mg/kg in Basmati370. The average 2-AP content across the aromatic genotypes was determined to be 0.85 mg/kg. Variations in the aromatic profile concentration of 2-AP among cultivars were reported by Pachauri et al., (2010). This finding indicates that while 2-AP is consistently present in aromatic rice, however, there is still considerable diversity in its concentration among different aromatic cultivars. The highest 2-AP concentration were observed in Basmati370 (1.4mg/kg), followed by ARS563-425-1-B-2-3 (1.1mg/kg), suggesting that these possess a particularly strong aromatic profile, which aligns with the sensory evaluation. The reputation of Basmati rice for its distinctive aroma has been underlined by previous studies (Chandi & Sogi, 2008; David et al., 2019; Kasote et al., 2021).

In the semi-aromatic class, 2-AP concentrations also displayed variability, with levels ranging from 0.16 mg/kg in IR107015-37 to 0.74 mg/kg in ARS79-5-11-11. This indicates that, like aromatic rice, semi-aromatic rice varieties also contain 2-AP, although the concentrations tend to be lower compared to fully aromatic varieties. The fact that some semi-aromatic genotypes have 2-AP concentrations comparable to or even exceeding those of certain aromatic genotypes indicates that there are other volatiles compounds or factors that contribute to the aroma beyond 2-AP content. Therefore, it was important to determine the other components that could be responsible for the aroma.

The correlation analysis conducted in this study aimed to shed light on the associations between 2-AP and other volatile components in rice. The results of this work revealed a strong and positive correlation between 2-AP and other 22 chromatogram peaks; which suggest that 2-AP is closely linked to a group of compounds, potentially

contributing to the overall aroma of rice. Among these, the aldehydes (n-Heptanal; Benzaldehyde; 1-Nonaldehyde; Decanal and Undecanal) and alcohols (Benzyl alcohol; 1-Decanol, 2-hexyl-; Isotridecanol-) were the major groups associated to 2-AP. Similar results were found by Mathure et al. (2011) indicating that hexanal, nonanal, decanal, benzyl alcohol, vanillin, guaiacol and indole have significantly contributed to the variation in 2-AP among Indian scented rice that included Basmati, and other cultivars (Ambemohar, Kolam, Indrayani and local). These results tend to confirm these volatiles as aroma compounds found in Basmati rice, since the intensities of the identified compounds were relatively higher. Daygon et al. (2017), have reported association between 2-AP and other amine heterocycles. The dissimilarities in the compound profiles may be due to differences among the genotypes used in the studies. Jie et al. (2021), found that volatile metabolites significantly differ between three aromatic rice and non-aromatic rice varieties. In our study, suggests that these compounds could be used as identity markers of aromatic rice and also key contributors to the sensory experience, influencing consumer preferences and acceptance.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

Using DArTseq-derived SNP markers to examine the population structure and genetic diversity among 94 rice germplasm accessions in the Eastern DRC, this work established that these were five diverse sub-populations with limited gene flow. It revealed significant genetic differentiation between them and within the genotypes. Through the field experiment, this study has identified significance traits such as productive tillers, panicle weight, primary branches per panicle, filled grains per panicle, and spikelets per panicle, as selection criteria associated with high grain yield in rice. The study also identified potential aromatic rice genotypes by combining the sensory test and volatile organic compounds profiling. Basmati 370 and ARS563-425-1-B-2-3 were classified as aromatic rice and exhibited higher 2-acetyl-1-pyrroline (2-AP) content. Furthermore, 22 volatile compounds positively and significantly associated to 2-AP content in rice. These include 9 identified compounds, n-Heptanal, Benzaldehyde, 1-Nonaldehyde, Decanal, Undecanal, Benzyl alcohol, 1-Decanol, 2-hexyl-, Isotridecanol-, and Mandelic acid, 2TBDMS derivative and 13 unknows.

6.2 Recommendations

The study revealed significant genetic diversity among the evaluated genotypes and could be used for rice improvement. Plant breeders can further use these resources for genome-wide association studies to understand genetic factors controlling traits of interest. The yield-associated traits can serve as morphological markers when selecting for enhanced yield and productivity. The biochemical compounds associated with aroma could be used for selecting the genotypes. Future studies should be explored to identify the 13 associated unknown volatiles and elucidate the genetic basis and roles of both identified and unknown compounds in shaping the aroma characteristics of rice varieties.

REFERENCES

Abaza, S. (2020). What is and why do we have to know the phylogenetic tree? *Parasitologists United Journal*, *13*(2), 68–71. https://doi.org/10.21608/puj.2020.35843.1082

Abebe, T., Alamerew, S., & Tulu, L. (2017). Genetic Variability, Heritability and Genetic Advance for Yield and its Related Traits in Rainfed Lowland Rice (*Oryza sativa* L.) Genotypes at Fogera and Pawe, Ethiopia. *Advances in Crop Science and Technology*, *05*(02), 272. https://doi.org/10.4172/2329-8863.1000272

Adeboye, K. A., Oyedeji, O. E., Alqudah, A. M., Börner, A., Oduwaye, O., Adebambo, O., & Daniel, I. O. (2020). Genetic structure and diversity of upland rice germplasm using diversity array technology (DArT)-based single nucleotide polymorphism (SNP) markers. *Plant Genetic Resources: Characterization and Utilization*, *18*(5), 343–350. https://doi.org/10.1017/S1479262120000362

Adu, B. G., Akromah, R., Amoah, S., Nyadanu, D., Yeboah, A., Aboagye, L. M., Amoah, R. A., & Owusu, E. G. (2021). High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). *PLOS ONE*, *16*(7), e0255290. https://doi.org/10.1371/journal.pone.0255290

Afolayan, G., Deshpande, S. P., Aladele, S. E., Kolawole, A. O., Angarawai, I., Nwosu, D. J., Michael, C., Blay, E. T., & Danquah, E. Y. (2019). Genetic diversity assessment of sorghum (*Sorghum bicolor* (L.) Moench) accessions using single nucleotide polymorphism markers. *Plant Genetic Resources: Characterization and Utilization*, *17*(5), 412–420. https://doi.org/10.1017/S1479262119000212

Ahmadi, N., Audebert, A., Bennett, M. J., Bishopp, A., de Oliveira, A. C., Courtois, B., Diedhiou, A., Diévart, A., Gantet, P., Ghesquière, A., Guiderdoni, E., Henry, A., Inukai, Y., Kochian, L., Laplaze, L., Lucas, M., Luu, D. T., Manneh,

B., Mo, X., … Xu, J. (2014). The roots of future rice harvests. *Rice*, *7*(1), 29. https://doi.org/10.1186/s12284-014-0029-y

Akhi, A. H., Miah, M. A. k., Ivy, N. A., Islam, A., & Islam, M. Z. (2016). Association of yield and yield related traits in aromatic rice (*Oryza sativa* L.). *Bangladesh Journal Agril. Research*, *3*(41), 387–396.

Akhi, A., Miah, M., Ivy, N., Islam, A., & Islam, M. (2016). Association of yield and yield related traits in aromatic rice (*Oryza sativa* L.). *Bangladesh Journal of Agricultural Research*, *41*(3), 387–396. https://doi.org/10.3329/bjar.v41i3.29710

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Amarawathi, Y., Singh, R., Singh, A. K., Singh, V. P., Mohapatra, T., Sharma, T. R., & Singh, N. K. (2007). Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). *Molecular Breeding*, *21*(1), 49–65. https://doi.org/10.1007/s11032-007-9108-8

Amegan, E., Efisue, A., Akoroda, M., Shittu, A., & Tonegnikes, F. (2020). Genetic Diversity of Korean Rice (*Oryza Sativa* L.) Germplasm for Yield and Yield Related Traits for Adoption in Rice Farming System in Nigeria. *International Journal of Genetics and Genomics*, *8*(1), 19. https://doi.org/10.11648/j.ijgg.20200801.13

Amos, W., & Harwood, J. (1998). Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1366), 177–186. https://doi.org/10.1098/rstb.1998.0200

Andrew, A. E., Bianca, C. U., & Joseph, A. O. (2014). Effects of yield components on yield potential of some lowland rice (*Oryza sativa* L.) in coastal region of

Southern Nigeria. *Journal of Plant Breeding and Crop Science*, *6*(9), 119–127. https://doi.org/10.5897/JPBCS2014.0449

Arouna, A., Fatognon, I. A., Saito, K., & Futakuchi, K. (2021). Moving toward rice self-sufficiency in sub-Saharan Africa by 2030: Lessons learned from 10 years of the Coalition for African Rice Development. World Development Perspectives, 21, 100291. https://doi.org/10.1016/j.wdp.2021.100291

Bailey, P.-W., Nalley, L. L., Mulimbi, W., & Durand, A. (2020). *Assessing Rice Market Efficiencies in the Democratic Republic of the Congo*. 1–12.

Balasubramanian, V., Sie, M., Hijmans, R. J., & Otsuka, K. (2007). Increasing Rice Production in Sub-Saharan Africa: Challenges and Opportunities. *Advances in Agronomy*, 94(06). https://doi.org/10.1016/S0065-2113(06)94002-4

Batley, J., & Edwards, D. (2007). SNP Applications in Plants. In N. C. Oraguzie, E. H. A. Rikkerink, S. E. Gardiner, & H. N. De Silva (Eds.), *Association Mapping in Plants* (pp. 95–102). Springer New York. https://doi.org/10.1007/978-0-387-36011-9_6

Becerra, V., Paredes, M., Gutiérrez, E., & Rojo, C. (2015). Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. *Chilean Journal of Agricultural Research*, *75*(3), 267–274. https://doi.org/10.4067/S0718-58392015000400001

Behera, P. K., & Panda, D. (2023). Germplasm Resources, Genes and Perspective for Aromatic Rice. *Rice Science*, *30*(4), 294–305. https://doi.org/10.1016/j.rsci.2023.03.011

Bhandari, H., Bhanu, N., Srivastava, K., Singh, M., Shreya, & Hemantaranjan, A. (2017). Assessment of Genetic Diversity in Crop Plants—An Overview. *Advances in Plants & Agriculture Research*, *7*(3), 279–286. https://doi.org/10.15406/apar.2017.07.00255

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314–331.

Boussaha, M., Guyomard, R., Cabau, C., Esquerré, D., & Quillet, E. (2012). Development and characterisation of an expressed sequence tags (EST)-derived single nucleotide polymorphisms (SNPs) resource in rainbow trout. *BMC Genomics*, *13*(1), 238. https://doi.org/10.1186/1471-2164-13-238

Bradbury, L. M. T., Henry, R. J., Jin, Q., Reinke, R. F., & Waters, D. L. E. (2005). A Perfect Marker for Fragrance Genotyping in Rice. *Molecular Breeding*, *16*(4), 279–283. https://doi.org/10.1007/s11032-005-0776-y

Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., & Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. *Nature Climate Change*. https://doi.org/10.1038/s41558-022-01585-1

Brown, W. L. (1983). Genetic diversity and genetic vulnerability—An appraisal. *Economic Botany*, *37*(1), 4–12. https://doi.org/10.1007/BF02859301

Burman, M., Nair, S. K., & Sarawgi, A. K. (2021). Principal Component Analysis for Yield and its Attributing Traits in Aromatic Landraces of Rice (*Oryza sativa* L.). *International Journal of Bio-Resource and Stress Management*, *12*(5), 303–308. https://doi.org/10.23910/1.2021.2348a

Calingacion, M., Laborte, A., Nelson, A., Resurreccion, A., Concepcion, J. C., Daygon, V. D., Mumm, R., Reinke, R., Dipti, S., Bassinello, P. Z., Manful, J., Sophany, S., Lara, K. C., Bao, J., Xie, L., Loaiza, K., El-hissewy, A., Gayin, J., Sharma, N., … Fitzgerald, M. (2014). Diversity of Global Rice Markets and the Science Required for Consumer-Targeted Rice Breeding. *PLoS ONE*, *9*(1), e85106. https://doi.org/10.1371/journal.pone.0085106

Campos, H., & Caligari, P. D. S. (2017). Genetic improvement of tropical crops. In *Genetic Improvement of Tropical Crops*. https://doi.org/10.1007/978-3-319-59819-2

Chandi, G. K., & Sogi, D. S. (2008). Characterization of Traditional (Basmati 370) and Developed (Pusa Basmati 1) Basmati Rice. *International Journal of Food Properties*, *11*(4), 910–918. https://doi.org/10.1080/10942910701673501

Chen, S., Yang, Y., Shi, W., Ji, Q., He, F., Zhang, Z., Cheng, Z., Liu, X., & Xu, M. (2008). *Badh2*, Encoding Betaine Aldehyde Dehydrogenase, Inhibits the Biosynthesis of 2-Acetyl-1-Pyrroline, a Major Component in Rice Fragrance. *The Plant Cell*, *20*(7), 1850–1861. https://doi.org/10.1105/tpc.108.058917

Custodio, M. C., Cuevas, R. P., Ynion, J., Laborte, A. G., Velasco, M. L., & Demont, M. (2019). Rice quality: How is it defined by consumers, industry, food scientists, and geneticists? *Trends in Food Science & Technology*, *92*, 122–137. https://doi.org/10.1016/j.tifs.2019.07.039

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

David, O., Rongrong, Z., Martin, O., Michael, K., Angele, P. I., Agnes, A., Bill, W. K., Ephraim, N., Jimmy, L., Melissa, A. F., Venea, D. D., & Patrick, R. R. (2019). Relationship between 2-acetyl-1-pyrroline and aroma in Uganda rice populations with *Oryza* (*barthi*, *glaberrima* and *sativa*) backgrounds. *African Journal of Biotechnology*, *18*(31), 1016–1024. https://doi.org/10.5897/AJB2019.16960

Daygon, V. D., Calingacion, M., Forster, L. C., Voss, J. J. D., Schwartz, B. D., Ovenden, B., Alonso, D. E., McCouch, S. R., Garson, M. J., & Fitzgerald, M. A. (2017). Metabolomics and genomics combine to unravel the pathway for

the presence of fragrance in rice. *Scientific Reports*, *7*(1), 8767. https://doi.org/10.1038/s41598-017-07693-9

Deres, D., & Feyissa, T. (2022). Concepts and applications of diversity array technology (DArT) markers for crop improvement. *Journal of Crop Improvement*, 1–21. https://doi.org/10.1080/15427528.2022.2159908

Dhakal, A., Sharma, S., Pokhrel, A., & Poudel, A. (2020). Variability and heritability estimate of 30 rice landraces of lamjung and tanahun districts, Nepal. *Indonesian Journal of Agricultural Science*, *21*(1), 1. https://doi.org/10.21082/ijas.v21n1.2020.p1-10

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., Poland, J., & Baenziger, P. S. (2018). Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Frontiers in Genetics*, *9*, 76. https://doi.org/10.3389/fgene.2018.00076

Esquinas-Alcázar, J. (2005). Science and society: Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature Reviews. Genetics*, *6*(12), 946–953. https://doi.org/10.1038/nrg1729

Fitzpatrick, B. M. (2009). Power and sample size for nested analysis of molecular variance. *Molecular Ecology*, *18*(19), 3961–3966. https://doi.org/10.1111/j.1365-294X.2009.04314.x

Food and Agriculture Organization. (2022). *The future of food and agriculture: Drivers and triggers for transformation*. FAO. https://doi.org/10.4060/cc0959en

International Fund for Agricultural Development (IFAD). (2019). *Food loss analysis: Causes and solutions—Case studies on maize and rice in the Democratic Republic of Congo* (p. 188).

Franco, J., Crossa, J., Ribaut, J. M., Bertran, J., Warburton, M. L., & Khairallah, M. (2001). A method for combining molecular markers and phenotypic attributes for classifying plant genotypes: *Theoretical and Applied Genetics*, *103*(6–7), 944–952. https://doi.org/10.1007/s001220100641

Fufa, T. W., Abtew, W. G., Amadi, C. O., & Oselebe, H. O. (2022). DArTSeq SNP-based genetic diversity and population structure studies among taro [(*Colocasia esculenta* (L.) Schott] accessions sourced from Nigeria and Vanuatu. *PLOS ONE*, *17*(11), e0269302. https://doi.org/10.1371/journal.pone.0269302

Mirindi, F. G. (2017). *Comparative analysis of rice value chains in the Ruzizi Plain of the Economic Community of the Great Lakes Countries (CEPGL)*. University of Liege - Gembloux.

Girma, T. B., Amanuel Kitil, M., Gebre Banje, D., Mengistu Biru, H., & Bayisa Serbessa, T. (2018). Genetic Variability Study of Yield and Yield Related Traits in Rice (*Oryza sativa* L.) Genotypes. *Advances in Crop Science and Technology*, *06*(04). https://doi.org/10.4172/2329-8863.1000381

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics*, *74*(1), 21–30. https://doi.org/10.1007/BF00290078

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiology*, *173*(1), 65–78. https://doi.org/10.1104/pp.16.01340

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its

analytical perspectives. *Genetics Research International*, *2015*. https://doi.org/10.1155/2015/431487

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Gu, D., Andreev, K., E. Dupre, M., United Nations Population Division, New York, USA, & Department of Population Health Sciences & Department of Sociology, Duke University, North Carolina, USA. (2021). Major Trends in Population Growth Around the World. *China CDC Weekly*, *3*(28), 604–613. https://doi.org/10.46234/ccdcw2021.160

Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., & Reddy, T. E. (2017). Transversions have larger regulatory effects than transitions. *BMC Genomics*, *18*(1), 394. https://doi.org/10.1186/s12864-017-3785-4

Hannan, A., Rana, M., Hoque, M., & Sagor, G. (2020). Genetic Variability, Character Association and Divergence Analysis for Agro-morphological Traits of Local Rice (*Oryza sativa* L.) Germplasms in Bangladesh. *Journal of Bangladesh Agricultural University*, *0*, 1. https://doi.org/10.5455/JBAU.84906

Hashemi, F. S. G., Rafii, M. Y., Ismail, M. R., Mahmud, T. M. M., Rahim, H. A., Asfaliza, R., Malek, M. A., & Latif, M. A. (2013). Biochemical, Genetic and Molecular Advances of Fragrance Characteristics in Rice. *Critical Reviews in Plant Sciences*, *32*(6), 445–457. https://doi.org/10.1080/07352689.2013.807716

Honma, I., Ito, A., Ueda, R., Hayakawa, F., & Maruyama, K. (2019). Sensory Characterization of Cooked Brown Rice Using Quantitative Descriptive Analysis (QDA®). *Nippon Shokuhin Kagaku Kogaku Kaishi*, *66*(2), 33–40. https://doi.org/10.3136/nskkk.66.33

Htwe, N. M., Phyu, S. L., & Thu, C. N. (2019). Assessment of Genetic Variability and Character Association of Myanmar Local Rice (*Oryza sativa* L.) Germplasm. *Journal of Experimental Agriculture International*, 1–10. https://doi.org/10.9734/jeai/2019/v40i330369

Hu, S., Ren, H., Song, Y., Liu, F., Qian, L., Zuo, F., & Meng, L. (2023). Analysis of volatile compounds by GCMS reveals their rice cultivars. *Scientific Reports*, *13*(1), 7973. https://doi.org/10.1038/s41598-023-34797-2

Hu, X., Lu, L., Guo, Z., & Zhu, Z. (2020). Volatile compounds, affecting factors and evaluation methods for rice aroma: A review. *Trends in Food Science & Technology*, *97*, 136–146. https://doi.org/10.1016/j.tifs.2020.01.003

Imran, M., Shafiq, S., Ashraf, U., Qi, J., Mo, Z., & Tang, X. (2023). Biosynthesis of 2-Acetyl-1-pyrroline in Fragrant Rice: Recent Insights into Agro-management, Environmental Factors, and Functional Genomics. *Journal of Agricultural and Food Chemistry*, *71*(10), 4201–4215. https://doi.org/10.1021/acs.jafc.2c07934

Iqbal, T., Hussain, I., Ahmad, N., Muhammad, D., Ali., M., Saeed, S., Zia, M. & Ali, F. (2018). Genetic Variability, Correlation and Cluster Analysis in Elite Lines of Rice. *Journal of Scientific Agriculture*, *2*, 85. https://doi.org/10.25081/jsa.2018.v2.900

IRRI, (1980). *Descriptors for Rice, Oryza Sativa L.* Int. Rice Res. Inst. Manila, 15p

Islam, M., Raffi, S., Hossain, M., & Hasan, A. (2015). Analysis of genetic variability, heritability and genetic advance for yield and yield associated traits in some promising advanced lines of rice. *Progressive Agriculture*, *26*(1), 26–31. https://doi.org/10.3329/pa.v26i1.24511

Jie Y, Shi T, Zhang Z, Yan Q. (2021). Identification of Key Volatiles Differentiating Aromatic Rice Cultivars Using an Untargeted Metabolomics Approach. *Metabolites*, 11(8), 528. https://doi.org/10.3390/metabo11080528

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955). Estimates of Genetic and Environmental Variability in Soybeans 1. *Agronomy Journal*, *47*(7), 314–318. https://doi.org/10.2134/agronj1955.00021962004700070009x

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics (Oxford, England)*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Joshi, B. K., Ghimire, K. H., Neupane, S. P., Gauchan, D., & Mengistu, D. K. (2023). Approaches and Advantages of Increased Crop Genetic Diversity in the Fields. *Diversity*, *15*(5), 603. https://doi.org/10.3390/d15050603

Kafi, H. S., Andrew Abiodun, E., Bunmi, O., & Kyung-Ho, K. (2021). Correlation Coefficient and Path Analyses of Yield and Yield Related Traits of Korean Double Haploid Rice for Germplasm Improvement in Nigeria. *American Journal of Agriculture and Forestry*, *9*(3), 114. https://doi.org/10.11648/j.ajaf.20210903.13

Kamvar, Z. N., & Grunwald, N. J. (2021). *Algorithms and equations utilized in poppr version 2.9.3*.

Kashyap, A., & Yadav, V. K. (2020). Principal Component Analysis and Character Association for Yield Components in Rice (*Oryza sativa* L.) Genotypes of Salt Tolerance under Alkaline Condition. *International Journal of Current Microbiology and Applied Sciences*, *9*(10), 481–495. https://doi.org/10.20546/ijcmas.2020.910.059

Kasote, D., Singh, V. K., Bollinedi, H., Singh, A. K., Sreenivasulu, N., & Regina, A. (2021). Profiling of 2-Acetyl-1-Pyrroline and Other Volatile Compounds in Raw and Cooked Rice of Traditional and Improved Varieties of India. *Foods*, *10*(8), 1917. https://doi.org/10.3390/foods10081917

Kawuki, R. S., Herselman, L., Labuschagne, M. T., Nzuki, I., Ralimanana, I., Bidiaka, M., Kanyange, M. C., Gashaka, G., Masumba, E., Mkamilo, G., Gethi, J., Wanjala, B., Zacarias, A., Madabula, F., & Ferguson, M. E. (2013). Genetic

diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa. *Plant Genetic Resources*, *11*(2), 170–181. https://doi.org/10.1017/S1479262113000014

Khan, M. H., Dar, Z. A., & Dar, S. A. (2015). Breeding Strategies for Improving Rice Yield—A Review. *Agricultural Sciences*, *06*(05), 467–478. https://doi.org/10.4236/as.2015.65046

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In F. Pompanon & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics* (Vol. 888, pp. 67–89). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_5

Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A., & McCouch, S. R. (2009). The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences*, *106*(34), 14444–14449. https://doi.org/10.1073/pnas.0904077106

Kshirod, K. J., & Eero, A. J. N. (2017). Genetic Improvement of Rice (*Oryza sativa* L.). In H. Campos & P. D. S. Caligari, *Genetic Improvement of Tropical Crops* (pp. 111–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-59819-2

Kumar, B., Rakshit, S., Kumar, S., Singh, B. K., Lahkar, C., Jha, A. K., Kumar, K., Kumar, P., Choudhary, M., Singh, S. B., Amalraj, J. J., Prakash, B., Khulbe, R., Kamboj, M. C., Chirravuri, N. N., & Hossain, F. (2022). Genetic Diversity, Population Structure and Linkage Disequilibrium Analyses in Tropical Maize Using Genotyping by Sequencing. *Plants*, *11*(6), 799. https://doi.org/10.3390/plants11060799

Lai, K., Duran, C., Berkman, P. J., Lorenc, M. T., Stiller, J., Manoli, S., Hayden, M. J., Forrest, K. L., Fleury, D., Baumann, U., Zander, M., Mason, A. S., Batley, J., & Edwards, D. (2012). Single nucleotide polymorphism discovery from wheat next-generation sequence data: Wheat autoSNPdb. *Plant Biotechnology Journal*, *10*(6), 743–749. https://doi.org/10.1111/j.1467-7652.2012.00718.x

Lapchareonsuk, R., & Sirisomboon, P. (2015). Sensory Quality Evaluation of Rice Using Visible and Shortwave Near-Infrared Spectroscopy. *International Journal of Food Properties*, *18*(5), 1128–1138. https://doi.org/10.1080/10942912.2013.870572

Lee, Y.-S., Oh, Y., Kim, T.-H., & Cho, Y.-H. (2019). Quantitation of 2-acetyl-1-pyrroline in aseptic-packaged cooked fragrant rice by HS-SPME/GC-MS. *Food Science & Nutrition*, *7*(1), 266–272. https://doi.org/10.1002/fsn3.879

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, G., Zhang, J., Yang, C., Song, Y., Zheng, C., Wang, S., Liu, Z., & Ding, Y. (2014). Optimal yield-related attributes of irrigated rice for high yield potential based on path analysis and stability analysis. *Crop Journal*, *2*(4), 235–243. https://doi.org/10.1016/j.cj.2014.03.006

Li, R., Li, M., Ashraf, U., Liu, S., & Zhang, J. (2019). Exploring the relationships between yield and yield-related traits for rice varieties released in china from 1978 to 2017. *Frontiers in Plant Science*, *10*(May). https://doi.org/10.3389/fpls.2019.00543

Li, W. (2012). Volcano plots in analyzing differential expressions with mRNA microarrays. *Journal of Bioinformatics and Computational Biology*, *10*(06), 1231003. https://doi.org/10.1142/S0219720012310038

Lipi, L., Hasan, M., Akter, A., Quddus, M., Biswas, P., Ansari, A., & Akter, S. (2021). Genetic Variation, Heritability and Genetic Advance in Some Promising Rice

Hybrids. *SAARC Journal of Agriculture*, *18*(2), 39–49. https://doi.org/10.3329/sja.v18i2.51107

Liu, C.-C., Shringarpure, S., Lange, K., & Novembre, J. (2020). Exploring Population Structure with Admixture Models and Principal Component Analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *2090*, 67–86. https://doi.org/10.1007/978-1-0716-0199-0_4

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., Ge, Y., Fahlgren, N., & Abdel-Haleem, H. (2019). Genetic Diversity and Population Structure of a Camelina sativa Spring Panel. *Frontiers in Plant Science*, *10*, 184. https://doi.org/10.3389/fpls.2019.00184

Ma, J., & Amos, C. I. (2012). Principal Components Analysis of Population Admixture. *PLoS ONE*, *7*(7), e40115. https://doi.org/10.1371/journal.pone.0040115

Matessi, C., & Jayakar, S. D. (1973). Estimation of the amount of inbreeding due to subdivision of a population. *Theoretical Population Biology*, *4*(4), 395–417. https://doi.org/10.1016/0040-5809(73)90017-8

Mascellani, A., Hoca, G., Babisz, M., Krska, P., Kloucek, P., & Havlik, J. (2021). 1H NMR chemometric models for classification of Czech wine type and variety. *Food Chemistry*, *339*, 127852. https://doi.org/10.1016/j.foodchem.2020.127852

Mathure, S. V., Wakte, K. V., Jawali, N., & Nadaf, A. B. (2011). Quantification of 2-Acetyl-1-pyrroline and Other Rice Aroma Volatiles Among Indian Scented Rice Cultivars by HS-SPME/GC-FID. *Food Analytical Methods*, *4*(3), 326–333. https://doi.org/10.1007/s12161-010-9171-3

McCouch, S. (2004). Diversifying selection in plant breeding. *PLoS Biology*, *2*(10), e347. https://doi.org/10.1371/journal.pbio.0020347

McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., Singh, N., DeClerck, G., Agosto-Perez, F., Korniliev, P., Greenberg, A. J., Naredo, Ma. E. B., Mercado, S. M. Q., Harrington, S. E., Shi, Y., Branchini, D. A., Kuser-Falcão, P. R., Leung, H., Ebana, K., … Mezey, J. (2016). Open access resources for genome-wide association mapping in rice. *Nature Communications*, *7*(1), 10532. https://doi.org/10.1038/ncomms10532

Ministere de l'agriculture et du developpement rural/RDC. (2013). *Strategie Nationale De Developpement De La Riziculture (SNDR)* (Secrétariat Général de l'Agriculture Pêche et Elevage, Ed.).

Mogga, M., Sibiya, J., Shimelis, H., Lamo, J., & Yao, N. (2018). Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa* L.) using DArT markers. *PLOS ONE*, *13*(6), e0198012. https://doi.org/10.1371/journal.pone.0198012

Mohidem, N. A., Hashim, N., Shamsudin, R., & Che Man, H. (2022). Rice for Food Security: Revisiting Its Production, Diversity, Rice Milling Process and Nutrient Content. Agriculture, 12(6), 741. https://doi.org/10.3390/agriculture12060741

Morton, B. R., Bi, I. V., McMullen, M. D., & Gaut, B. S. (2006). Variation in Mutation Dynamics Across the Maize Genome as a Function of Regional and Flanking Base Composition. *Genetics*, *172*(1), 569–577. https://doi.org/10.1534/genetics.105.049916

Mudaki, P., Wamalwa, L. N., Muui, C. W., Nzuve, F., Muasya, R. M., Nguluu, S., & Kimani, W. (2023). Genetic Diversity and Population Structure of Sorghum (*Sorghum bicolor* (L.) Moench) Landraces Using DArTseq-Derived Single-Nucleotide Polymorphism (SNP) Markers. *Journal of Molecular Evolution*. https://doi.org/10.1007/s00239-023-10108-1

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption: Global

rice production, consumption, and trade. *Annals of the New York Academy of Sciences*, *1324*(1), 7–14. https://doi.org/10.1111/nyas.12540

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, *32*(2), 261–285. https://doi.org/10.1080/13102818.2017.1400401

Ndikuryayo, C., Ndayiragije, A., Kilasi, N., & Kusolwa, P. (2023). Screening for aroma in rice genotypes with African and Asian backgrounds. *Urban Agriculture & Regional Food Systems*, *8*(1), e20045. https://doi.org/10.1002/uar2.20045

Ndjiondjop, M. N., Semagn, K., Zhang, J., Gouda, A. C., Kpeki, S. B., Goungoulou, A., Wambugu, P., Dramé, K. N., Bimpong, I. K., & Zhao, D. (2018). Development of species diagnostic SNP markers for quality control genotyping in four rice (*Oryza* L.) species. *Molecular Breeding*, *38*(11), 131. https://doi.org/10.1007/s11032-018-0885-z

Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist*, *106*(949), 283–292. https://doi.org/10.1086/282771

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. https://doi.org/10.7312/nei-92038

Ng'endo, M., Kinyua, M., Chebet, L., Mutiga, S., Ndung'u, J., Nyongesa, O., Njau, S., Panchbhai, A., Musila, R., & Murori, R. (2022). The importance of market signals in crop varietal development: Lessons from Komboka rice variety. *CABI Agriculture and Bioscience*, 3(1), 57. https://doi.org/10.1186/s43170-022-00122-6

Niedziela, A., & Bednarek, P. T. (2023). Population structure and genetic diversity of a germplasm for hybrid breeding in rye (*Secale cereale* L.) using high-density

DArTseq-based silicoDArT and SNP markers. *Journal of Applied Genetics*. https://doi.org/10.1007/s13353-022-00740-w

Norman, J. C., & Kebe, B. (2004). African smallholder farmers: Rice production. *Plant Production*, 33–64.

Novoselović, D., Bentley, A. R., Šimek, R., Dvojković, K., Sorrells, M. E., Gosman, N., Horsnell, R., Drezner, G., & Šatović, Z. (2016). Characterizing Croatian Wheat Germplasm Diversity and Structure in a European Context by DArT Markers. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00184

Oladosu, Y., Rafii, M. Y., Magaji, U., Abdullah, N., Miah, G., Chukwu, S. C., Hussin, G., Ramli, A., & Kareem, I. (2018). Genotypic and Phenotypic Relationship among Yield Components in Rice under Tropical Conditions. *BioMed Research International*, *2018*, 1–10. https://doi.org/10.1155/2018/8936767

Pachauri, V., Mishra, V., Mishra, P., Singh, A. K., Singh, S., Singh, R., & Singh, N. K. (2014). Identification of candidate genes for rice grain aroma by combining QTL mapping and transcriptome profiling approaches. *Cereal Research Communications*, *42*(3), 376–388. https://doi.org/10.1556/CRC.42.2014.3.2

Pachauri, V., Singh, M. K., Singh, A. K., Singh, S., Shakeel, N. A., Singh, V. P., & Singh, N. K. (2010). Origin and Genetic Diversity of Aromatic Rice Varieties, Molecular Breeding and Chemical and Genetic Basis of Rice Aroma. *Journal of Plant Biochemistry and Biotechnology*, *19*(2), 127–143. https://doi.org/10.1007/BF03263333

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Prodhan, Z. H., & Qingyao, S. (2020). Rice Aroma: A Natural Gift Comes with Price and the Way Forward. *Rice Science*, *27*(2), 86–100. https://doi.org/10.1016/j.rsci.2020.01.001

R Core Team. (2023). *R: A language and environment for statistical computing* (4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rashid, M. M., Nuruzzaman, M., Hassan, L., & Begum, S. N. (2017). *Genetic variability analysis for various yield attributing traits in rice genotypes*. *15*(1), 15–19.

Rashmi, D., Saha, S., Loitongbam, B., Singh, S., & Singh, P. K. (2017). Genetic variability study for yield and yield components in rice (*Oryza sativa* L.). *International Journal of Agriculture, Environment and Biotechnology*, *10*(2), 171. https://doi.org/10.5958/2230-732X.2017.00020.1

Robinson, H. F., Comstock, R. E., & Harvey, P. H. (1949). Estimates of Heritability and the Degree of Dominance in Corn [1]. *Agronomy Journal*, *41*(8), 353–359. https://doi.org/10.2134/agronj1949.00021962004100080005x

Rundle, H. D., & Nosil, P. (2005). Ecological speciation: Ecological speciation. *Ecology Letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Saleh, M. M., Salem, K. F. M., & Elabd, A. B. (2020). Definition of selection criterion using correlation and path coefficient analysis in rice (*Oryza sativa* L.) genotypes. *Bulletin of the National Research Centre*, *44*(1), 143. https://doi.org/10.1186/s42269-020-00403-y

Salem, K. F. M., & Sallam, A. (2016). Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. *Comptes Rendus Biologies*, *339*(1), 1–9. https://doi.org/10.1016/j.crvi.2015.11.003

Salgotra, R. K., & Chauhan, B. S. (2023). Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes*, *14*(1), 174. https://doi.org/10.3390/genes14010174

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, *5*(S7), P54, 1753-6561-5-S7-P54. https://doi.org/10.1186/1753-6561-5-S7-P54

Schierenbeck, K. A. (2017). Population-level genetic variation and climate change in a biodiversity hotspot. *Annals of Botany*, *119*(2), 215–228. https://doi.org/10.1093/aob/mcw214

Seck, P. A., Touré, A. A., Coulibaly, J. Y., Diagne, A., & Wopereis, M. C. S. (2013). Africa's Rice Economy Before and After the 2008 Rice Crisis. In J. A. Wopereis MCS, Johnson DE, Ahmadi N, Tollens E (Ed.), *Realizing Africa's rice promise* (CABI, pp. 24–34).

Siddiq, E. A., & Vemireddy, L. R. (2021). Advances in Genetics and Breeding of Rice: An Overview. In J. Ali & S. H. Wani (Eds.), *Rice Improvement* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-66530-2_1

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., Singh, N. K., & Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PloS One*, *8*(12), e84136. https://doi.org/10.1371/journal.pone.0084136

Singh, S., & Pawar, I. S. (2005). Theory and Application of Biometrical Genetics. India: CBS Publishers

Sleper, D. A., & Poehlman, J. M. (2006). *Breeding field crops* (5th ed). Blackwell Pub.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Smith, S., Nickson, T. E., & Challender, M. (2021). Germplasm exchange is critical to conservation of biodiversity and global food security. *Agronomy Journal*, *113*(4), 2969–2979. https://doi.org/10.1002/agj2.20761

Sudeepthi, K., Srinivas, T., Ravi Kumar, B. N. V. S. R., Jyothula, D. P. B., & Nafeez Umar, Sk. (2020). Genetic Diversity Studies for Yield and Yield Component Traits using Principal Component Analysis in Rice (*Oryza sativa* L.). *Indian Journal of Pure & Applied Biosciences*, *8*(3), 228–235. http://dx.doi.org/10.18782/2582-2845.7925

Suvi, W. T., Shimelis, H., Laing, M., Mathew, I., & Shayanowako, A. I. T. (2021). Variation among Tanzania Rice Germplasm Collections Based on Agronomic Traits and Resistance to Rice Yellow Mottle Virus. *Agronomy*, *11*(2), 391. https://doi.org/10.3390/agronomy11020391

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science, 61*(2), 839–852. https://doi.org/10.1002/csc2.20377

Tabanao, D. A., Navarro, R. B., Millas, R. A., & Niño, M. C. (2021). Genetic Characterization of BADH2 in Philippine Aromatic Rice Cultivars. Plant Breeding and Biotechnology, 9(3), 227–238. https://doi.org/10.9787 /PBB.2021.9.3.227

Talukdar, P. R., Rathi, S., Pathak, K., Chetia, S. K., & Sarma, R. N. (2017). Population Structure and Marker-Trait Association in Indigenous Aromatic Rice. *Rice Science, 24*(3), 145–154. https://doi.org/10.1016/j.rsci.2016.08.009

Tanchotikul, U., & Hsieh, T. C. Y. (1991). An improved method for quantification of 2-acetyl-1-pyrroline, a “popcorn”-like aroma, in aromatic rice by high-resolution gas chromatography/mass spectrometry/selected ion monitoring. *Journal of Agricultural and Food Chemistry*, *39*(5), 944–947. https://doi.org/10.1021/jf00005a029

Thant, A. A., Zaw, H., Kalousova, M., Singh, R. K., & Lojka, B. (2021). Genetic Diversity and Population Structure of Myanmar Rice (*Oryza sativa* L.) Varieties Using DArTseq-Based SNP and SilicoDArT Markers. *Plants*, *10*(12), 2564. https://doi.org/10.3390/plants10122564

Tonegnikes, S., Efisue, A., Adetimirin, V., Shittu, A., & Amegan, E. (2019). Evaluation of Korea Rice Germplasm for Yield and Yield Components Adaptable to Nigeria Environmental Conditions. *Journal of Plant Sciences*, *7*(5), 106–116. https://doi.org/doi: 10.11648/j.jps.20190705.12

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., & Arita, M. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, *12*(6), 523–526. https://doi.org/10.1038/nmeth.3393

Twine, E. E., Ndindeng, S. A., Mujawamariya, G., Adur-Okello, S. E., & Kilongosi, C. (2023). Consumer preferences for rice in East Africa. *British Food Journal*, *125*(13), 316–329. https://doi.org/10.1108/BFJ-08-2022-0698

Ullah, F., Gao, Y., Sari, İ., Jiao, R.-F., Saqib, S., & Gao, X.-F. (2022). Macro-Morphological and Ecological Variation in Rosa sericea Complex. *Agronomy*, *12*(5), 1078. https://doi.org/10.3390/agronomy12051078

Verma, D. K., & Srivastav, P. P. (Eds.). (2018). *Science and Technology of Aroma, Flavor, and Fragrance in Rice* (0 ed.). Apple Academic Press. https://doi.org/10.1201/b22468

Verma, D. K., & Srivastav, P. P. (2020). A paradigm of volatile aroma compounds in rice and their product with extraction and identification methods: A comprehensive review. *Food Research International*, *130*, 108924. https://doi.org/10.1016/j.foodres.2019.108924

Verma, D. K., & Srivastav, P. P. (2022). Extraction, Identification and Quantification Methods of Rice Aroma Compounds with Emphasis on 2-Acetyl-1-Pyrroline (2-AP) and Its Relationship with Rice Quality: A Comprehensive Review.

Food Reviews International, *38*(2), 111–162. https://doi.org/10.1080/87559129.2020.1720231

Wang, L.-M., Wang, P., Teka, T., Zhang, Y.-C., Yang, W.-Z., Zhang, Y., Wang, T., Liu, L.-X., Han, L.-F., & Liu, C.-X. (2020). ^{1}H NMR and UHPLC/Q-Orbitrap-MS-Based Metabolomics Combined with 16S rRNA Gut Microbiota Analysis Revealed the Potential Regulation Mechanism of Nuciferine in Hyperuricemia Rats. *Journal of Agricultural and Food Chemistry*, *68*(47), 14059–14070. https://doi.org/10.1021/acs.jafc.0c04985

Wenig, P., & Odermatt, J. (2010). OpenChrom: A cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinformatics*, *11*, 405. https://doi.org/10.1186/1471-2105-11-405

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. *Proceedings of the National Academy of Sciences*, *101*(26), 9915–9920. https://doi.org/10.1073/pnas.0401076101

Wilkie, K., Wootton, M., & Paton, J. E. (2004). Sensory Testing of Australian Fragrant, Imported Fragrant, and Non-fragrant Rice Aroma. *International Journal of Food Properties*, *7*(1), 27–36. https://doi.org/10.1081/JFP-120022493

Wright, S. (1965a). The interpretation of population structure by f-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

Wright, S. (1965b). The interpretation of population structure by f-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

Wright, S. (1968). *Evolution and the genetics of populations: A treatise*. University of Chicago Press.

Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, *37*(Web Server), W652–W660. https://doi.org/10.1093/nar/gkp356

Xie, L., Tang, S., Wei, X., Jiao, G., Sheng, Z., & Hu, P. (2019). An optimized analysis of 2-acetyl-1-pyrroline content diversity in the core collection of aromatic rice germplasm (*Oryza sativa* L.). *Cereal Chemistry*, *96*(4), 698–707. https://doi.org/10.1002/cche.10166

Xu, Q., Yuan, X., Wang, S., Feng, Y., Yu, H., Wang, Y., Yang, Y., Wei, X., & Li, X. (2016). The genetic diversity and structure of indica rice in China as detected by single nucleotide polymorphism analysis. *BMC Genetics*, *17*(1), 53. https://doi.org/10.1186/s12863-016-0361-x

Yang, H.-C., Lin, H.-C., Huang, M.-C., Li, L.-H., Pan, W.-H., Wu, J.-Y., & Chen, Y.-T. (2010). A new analysis tool for individual-level allele frequency for genomic studies. *BMC Genomics*, *11*(1), 415. https://doi.org/10.1186/1471-2164-11-415

Yao, L., Sheflin, A. M., Broeckling, C. D., & Prenni, J. E. (2019). Data Processing for GC-MS- and LC-MS-Based Untargeted Metabolomics. In A. D'Alessandro (Ed.), *High-Throughput Metabolomics* (Vol. 1978, pp. 287–299). Springer New York. https://doi.org/10.1007/978-1-4939-9236-2_18

Zheng, Z., Zhang, C., Liu, K., & Liu, Q. (2022). Volatile Organic Compounds, Evaluation Methods and Processing Properties for Cooked Rice Flavor. *Rice*, *15*(1), 53. https://doi.org/10.1186/s12284-022-00602-3

APPENDICES

Appendix I: Rice Genotypes Used for Determining the Selection Criteria for Rice Grain Yield Selection

Entry No.	Genotype name	Source/program	Entry No.	Genotype name	Source/program
1	Komboka	IRRI-Burundi	26	IR17015-6-5-3-B1	INERA-DRC
2	IR64	IRRI-Burundi	27	IR106359-B-18-5	INERA-DRC
3	IBEI6	AfricaRice	28	IR95624-B-138-3	INERA-DRC
4	GIZA128	IRRI-Burundi	29	IR13A461	IRRI-Burundi
5	Nipponbare	IRRI-Burundi	30	Mugwiza	IRRI-Burundi
6	Jasmine	IRRI-Burundi	31	Vuninzara	IRRI-Burundi
7	NL59	AfricaRice	32	IR97045-24-1-1-1	IRRI-Burundi
8	FKR	AfricaRice	33	Kigoma	Local landrace
9	08FAN10	IRRI-Burundi	34	Makasane	IRRI-Burundi
10	WAB2066-TGR2	AfricaRice	35	Rukaramu	Local landrace
11	WAB2066-TGR3	AfricaRice	36	Musesekara	IRRI-Burundi
12	IR99084-B-B-13	INERA-DRC	37	Yasho-Yasho	Local landrace
13	IR127229	IRRI-Burundi	38	Kasozi	IRRI-Burundi
14	IR106172-78 :1-B-B	INERA-DRC	39	IR7525	IRRI-Burundi
15	ARS848-15-3-2-3	AfricaRice	40	Orylux7	AfricaRice
16	IR106364-B-B-CNUS	INERA-DRC	41	ART29	INERA-DRC
17	ARS844-24-10-2-B	AfricaRice	42	Sipi	INERA-DRC
18	ARS168-1-B-3-B	AfricaRice	43	CRS36	IRRI-Kenya
19	ARS851-1-3	AfricaRice	44	ARICA2	AfricaRice
20	IR87638-10-2-2-4	INERA-DRC	45	NL19	AfricaRice
21	IR98419-B-B-11	INERA-DRC	46	NL14	AfricaRice
22	IR97071-24-1-1-1	INERA-DRC	47	NL17	AfricaRice
23	ARS803-4-5-4-3	AfricaRice	48	D20-ARS-3-2	AfricaRice
24	IR93856-23-1-1-1	INERA-DRC	49	IR96279-33-3-1-2	IRRI-Burundi
25	ARS790-5-11-1-1	AfricaRice			

Appendix II: Rice Genotypes Used in Identifying Potential Volatile Organic Compounds Associated to the Rice Grain Aroma Trait

Entry No.	Genotype name	Entry No.	Genotype name	Entry No.	Genotype name
1	Komboka	30	Mugwiza	58	IR88638
2	IR64	31	Vuninzara	59	ARICA12
3	IBEI6	32	IR97045-24-1-1-1	60	ARICA3
4	GIZA128	33	Kigoma	61	IR64-sub-1
5	Nipponbare	34	Makasane	62	HHZSAL6
6	Jasmine	35	Rukaramu	63	ARS755-3-3-1-B
7	NL59	36	Musesekara	64	ARS134-B-1-1-5
8	FKR	37	Yasho-Yasho	65	IR990-48-B-B-12
9	08FAN10	38	Kasozi	66	IR64-biofortified
10	WAB2066-TGR2	39	IR7525	67	IR107015-37
11	WAB2066-TGR3	40	Orylux7	68	ARS79-5-11-11
12	IR99084-B-B-13	41	ART29	69	V18/RRS126-48-1-13-2
13	IR127229	42	Sipi	70	Orylux11
14	IR106172-78 :1-B-B	43	CRS36	71	ARS134-B-B-B
15	ARS848-15-3-2-3	44	ARICA2	72	Magoti
16	IR106364-B-B-CNUS	45	NL19	73	Runingu
17	ARS844-24-10-2-B	46	NL14	74	ARS169-2-B-3-B
18	ARS168-1-B-3-B	47	NL17	75	ARS134-B-1-1-4
19	ARS851-1-3	48	D20-ARS-3-2	76	IR82574/643-1-2
20	IR87638-10-2-2-4	49	IR96279-33-3-1-2	77	Orylux5
21	IR98419-B-B-11	50	ARS134-B-1-1-5-B	78	SAHEL210
22	IR97071-24-1-1-1	51	Orylux7-1	79	IR841
23	ARS803-4-5-4-3	52	WAHX14N-926	80	ARS39-145/EP-3
24	IR93856-23-1-1-1	53	MR254	81	ARS101-4-B-1-1-B
25	ARS790-5-11-1-1	54	Golmy	82	ARS101-4-B-1-3
26	IR17015-6-5-3-B1	55	ARS848-15-3-2-4	83	NERICA-L-19-Sab-1
27	IR106359-B-18-5	56	IR93348:32-B-15-3-B-B-B-1	84	ARS756-1-1-3-B-2-2
28	IR95624-B-138-3	57	ARS168-3-B-1-B	85	ARS563-425-1-B-2-3
29	IR13A461				

Appendix III: Mean Values for Yield and Yield Components of 49 Rice Genotypes Evaluated

Genotype	DTF	DTM	PH	NPTH	PL	NPBP	NSP	NFGP	PW	ThGW	GL	Gw	GLGwR	GY
Komboka	109	150	95.3	15	24.2	11	199	193	3.36	24.34	8.72	2.72	3.22	3.75
IR64	104	144	88.0	23	25.5	12	189	173	3.85	26.12	8.33	1.40	6.22	4.14
Hubei6	109	143	85.9	16	24.2	9	194	180	2.96	28.01	8.77	2.57	3.44	3.84
Giza182	111	149	94.0	17	24.3	9	182	170	3.48	28.16	9.14	2.55	3.88	3.65
Nipponbare	117	147	94.5	16	25.4	10	200	178	2.93	28.93	8.93	2.59	3.48	3.76
Jasmine	116	152	120.3	14	28.1	12	251	217	4.25	29.31	8.94	3.58	2.51	3.49
NL59	106	147	85.7	16	24.2	11	153	146	3.60	27.49	8.95	2.47	3.74	3.70
FKR28	109	143	86.8	17	24.2	10	162	148	2.35	28.45	8.57	2.20	4.04	3.70
08FAN10	107	145	89.3	16	23.2	9	153	131	2.90	26.96	8.66	2.26	3.79	3.21
WAB2066-TGR2	99	140	84.3	16	22.3	10	154	137	3.41	26.35	8.59	1.45	5.96	3.72
WAB2066-TGR3	110	145	82.3	14	22.4	8	127	124	2.99	28.72	9.07	2.46	3.71	3.56
IR99084-B-B-13	102	144	82.9	14	22.6	9	164	140	3.22	29.11	9.48	1.88	5.00	3.58
IR127229	121	162	112.4	14	30.1	11	290	267	3.73	28.65	9.08	2.04	4.36	3.52
IR106172-78 :1-B-B	108	145	88.4	17	23.4	9	163	145	3.69	28.56	8.87	3.36	2.65	3.34
ARS848-15-3-2-3	116	155	118.4	15	28.7	11	192	170	4.07	32.82	9.46	1.74	5.80	3.51
IR106364-B-B-CNUS	100	142	84.1	16	22.8	9	168	145	3.00	28.64	8.65	2.45	3.55	3.43
ARS844-24-10-2-B	112	152	102.8	15	25.3	10	209	183	3.27	28.12	8.77	3.34	2.65	3.39
ARS168-1-B-3-B	108	144	90.1	17	24.5	12	224	188	3.82	28.42	8.99	2.89	3.13	3.63
ARS 851-1-3	113	150	88.2	15	23.8	10	163	152	3.32	26.87	8.89	2.63	3.55	3.54
IR87638-10-2-2-4	101	144	90.2	16	26.6	10	216	206	4.13	25.56	9.20	2.50	3.97	3.80
IR98419-B-B-11	105	147	87.1	17	24.2	11	167	156	3.95	27.88	8.94	2.99	3.13	3.82
IR97071-24-1-1-1	110	149	91.4	16	24.2	11	196	182	3.66	26.78	9.28	3.19	2.91	3.83
ARS803-4-5-4-3	102	147	85.3	13	21.9	8	138	133	2.74	28.21	9.23	2.41	3.93	3.33
IR93856-23-1-1-1	114	151	91.6	14	24.4	11	166	142	3.36	29.48	9.09	2.58	3.59	3.55
ARS790-5-11-1-1	109	144	92.5	19	23.7	12	194	172	3.64	24.19	8.51	2.57	3.31	3.75
IR17015-6-5-3-B1	114	149	90.9	15	24.1	11	169	154	3.51	28.67	9.03	2.40	3.90	3.62
IR106359-B-18-5	107	139	89.4	18	24.9	12	206	182	3.81	29.58	8.86	2.79	3.24	3.92
IR95624-B-138-3	104	137	87.3	15	25.3	11	186	166	3.73	30.35	9.19	2.60	3.55	3.51
IR13A461	107	143	98.1	27	28.4	14	232	207	4.65	27.13	8.24	2.99	2.81	4.23
Mugwiza	111	149	111.0	16	25.9	11	187	179	3.04	24.21	8.08	1.75	4.68	3.31
Vuninzara	107	147	90.6	15	22.6	10	160	144	2.85	27.52	9.07	2.95	3.14	3.70
IR97045-24-1-1-1	105	147	87.1	17	24.7	11	169	160	3.29	25.76	8.72	2.61	3.35	3.75

Genotype	DTF	DTM	PH	NPTH	PL	NPBP	NSP	NFGP	PW	ThGW	GL	Gw	GLGwR	GY
Kigoma	103	144	96.0	15	23.9	10	175	148	3.44	27.51	8.91	3.04	2.95	3.48
Makasane	115	152	86.5	15	24.7	10	163	150	3.26	28.36	8.78	2.49	3.55	3.48
Rukaramu	112	152	98.3	14	24.9	11	187	161	2.99	28.19	8.94	3.20	2.83	2.99
Musesekara	123	159	97.6	16	26.8	11	183	167	3.46	28.75	9.21	2.70	3.47	3.65
Yasho Yasho	121	161	88.8	13	23.6	12	186	172	3.91	26.31	8.47	3.29	2.65	3.35
Kasozi	118	157	108.1	14	26.9	12	215	191	3.67	28.10	9.08	2.63	3.50	3.67
IR7525	116	147	82.9	14	23.9	10	177	159	2.59	21.74	8.58	2.09	4.37	3.32
Orylux6	113	159	92.0	19	25.2	11	181	170	3.49	29.47	8.61	2.53	3.44	3.85
ART29	111	160	97.8	14	24.2	11	197	170	3.47	25.26	8.61	2.58	3.33	3.51
SIPI	111	146	92.8	15	24.4	10	192	190	3.08	26.03	8.82	2.60	3.30	3.73
CRS36	120	155	97.2	17	25.5	11	195	170	3.00	26.62	8.74	2.22	3.87	3.65
ARICA2	122	163	115.7	16	26.3	12	221	194	3.86	29.05	9.02	2.85	3.28	3.70
NL19	110	152	102.7	15	25.7	10	183	174	4.39	28.20	8.93	3.37	2.86	3.45
NL14	116	149	107.5	18	25.5	10	180	170	3.82	28.85	8.96	2.45	3.73	3.60
NL17	101	132	104.4	12	23.0	9	160	144	2.92	34.29	8.83	1.42	6.25	3.35
D20-ARS-3-2	107	144	91.0	19	25.2	11	211	193	3.86	25.65	8.41	1.83	4.60	4.09
IR96279-33-3-1-2	107	153	95.0	13	24.3	9	176	148	2.65	29.47	8.65	2.57	3.29	2.96
Mean	110	148	94.3	16	24.8	11	186	168	3.44	27.78	8.85	2.55	3.7	3.6
CV (%)	1.9	1.5	3.7	13.5	4	10.6	9.1	10.8	10.3	4.9	4.7	10.2	4.8	6.6

DTF: Days to Flowering, DTM: Days to maturity, PH: Plant Height, NPTh: Number of productive Tillers/hill, PL: Panicle Length, NPBp: Number of primary branches/panicle, NSP: Number of Spikelet/panicle, NFGp: Number of filled grains/panicle, PW: Panicle weight, ThGW: Thousand Grains Weight, GY: Grain yield, GL: Grain Length, Gw: Grain Width and RGLGw: Ratio Grain Length/Grain Width.

Appendix IV: Euclidean Genetic Distances among the 94 Rice Genotypes based on 8389 DArTseq SNP Markers

The information can be downloaded from the link, https://www.mdpi.com/article/10.3390/agronomy13071906/s1

Appendix V: Correlation Coefficients between 2AP and the Significant Features

Nº	Chromatogram features	6.52/111.5 (2AP)	Nº	Chromatogram features	6.52/111.5 (2AP)
1	25.55/219.6	-0.21	22	24.97/415.6	0.57
2	25.67/138.5	-0.28	23	25.93/399.7	0.64
3	22.05/112.6	0.13	24	22.18/90.6	0.71
4	2.12/42.5	-0.03	25	6.05/79.5	0.89
5	21.23/75.5	-0.02	26	9.52/98.5	0.90
6	10.02/89.6	0.13	27	16.33/98.5	0.81
7	27.13/113.6	0.15	28	12.48/105.5	0.83
8	26.12/43.4	0.23	29	23.95/137.5	0.79
9	9.17/126.7	0.17	30	22.7/74.5	0.68
10	15.62/94.6	0.20	31	26.42/148.5	0.62
11	27.48/132.6	0.21	32	17.63/342.8	0.74
12	17.85/98.6	0.01	33	23.67/73.4	0.77
13	7.68/63.6	0.20	34	17.97/73.4	0.71
14	22.15/84.6	0.06	35	21.27/59.5	0.70
15	9.78/37.4	0.26	36	15.02/145.7	0.62
16	24.17/50.4	0.30	37	9.97/116.5	0.83
17	18.5/42.5	0.43	38	14.62/44.4	0.78
18	3.93/37.4	0.72	39	11.7/40.4	0.84
19	5.28/52.5	0.59	40	7.82/109.5	0.88
20	23.53/189.5	0.43	41	28.3/179.6	0.85
21	22.9/76.4	0.45			

Appendix VI: Intensities (Dalton) of Significant Peaks Associated to 2AP Across the Aromatic and Semi-Aromatic Rice

mz	rt	ARS169-2-B-3-B	ARS563-425-1-B-2-3	Basmati370	Jasmine	WAB2066-TGR2	ARICA12	ARICA2
		Aromatic	Aromatic	Aromatic	Aromatic	Aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	652.04	1034.81	11084.08	367.65	1198.55	419.86	0.00
111.5	6.52	2377.29	4651.10	15138.87	0.00	2922.89	1486.21	0.00
109.5	7.82	763.25	990.74	6344.25	689.09	689.21	1169.49	0.00
98.5	9.52	669.82	613.40	13878.01	732.11	1044.13	723.11	0.00
116.5	9.97	193.34	160.31	1139.16	215.82	221.04	215.70	0.00
40.4	11.7	4130.78	1527.05	32996.44	3388.50	5995.83	3755.12	0.00
105.5	12.48	236.87	484.60	5693.84	367.68	458.14	284.05	0.00
44.4	14.62	3940.27	4127.84	13905.33	2031.77	2218.78	2625.59	0.00
98.5	16.33	1416.45	426.16	11461.80	719.20	1810.07	1462.02	0.00
267.8	16.37	0.00	137.36	776.00	212.89	0.00	0.00	0.00
342.8	17.63	8337.39	16123.66	1061321.09	232815.34	0.00	8127.33	0.00
73.4	17.97	411.47	204282.73	6195137.71	959891.52	9962.42	130173.11	1093.03
59.5	21.27	1727.83	2505.73	82665.59	18136.47	473.28	1758.92	0.00
90.6	22.18	42.99	199.39	1177.28	650.90	25.58	58.67	0.00
74.5	22.7	9286.54	13816.81	173678.87	29557.96	4063.04	318.61	0.00
76.4	22.9	180.35	0.00	6040.61	3741.96	0.00	16.15	42.74
189.5	23.53	1180.61	863.05	5465.47	3533.81	947.43	1077.19	125.83
73.4	23.67	65128.89	84805.22	2149642.09	357848.97	53460.76	24341.46	0.00
137.5	23.95	3412.20	2856.53	40670.41	3717.43	3275.23	4169.27	0.00
415.6	24.97	1146.71	2621.60	44387.67	22574.96	989.61	724.62	1270.59
399.7	25.93	211.12	301.42	4217.12	1264.73	226.97	61.83	534.60
148.5	26.42	1243.73	865.29	15452.12	4375.81	778.35	2321.11	160.85
179.6	28.3	179.56	167.06	1298.55	112.78	131.34	146.76	0.00

mz	rt	ARICA3	ARS134-B-1-1-5-B	ARS168-1-B-3-B	ARS756-1-1-3-B-2-2	ARS79-5-11-11	Yasho-Yasho
		Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	0.00	843.34	0.00	0.00	553.24	0.00
111.5	6.52	1598.51	0.00	1396.33	164.26	0.00	28.23
109.5	7.82	0.00	1130.88	0.00	0.00	683.54	0.00
98.5	9.52	0.00	622.30	0.00	0.00	525.24	2.33
116.5	9.97	0.00	287.70	0.00	0.00	336.48	0.00
40.4	11.7	0.00	3462.84	0.00	3794.02	4449.78	0.00
105.5	12.48	0.00	436.73	0.00	0.00	493.03	781.11
44.4	14.62	0.00	1872.51	0.00	5438.98	2237.00	0.00
98.5	16.33	0.00	1432.56	0.00	1257.89	241.02	0.00
267.8	16.37	0.00	0.00	0.00	0.00	106.17	0.00
342.8	17.63	0.00	0.00	0.00	0.00	10771.24	0.00
73.4	17.97	0.00	19896.52	13406.50	1859728.88	1030305.37	1426.53
59.5	21.27	0.00	1699.71	0.00	20320.54	9409.31	0.00
90.6	22.18	0.00	25.29	0.00	0.00	370.77	0.00
74.5	22.7	0.00	712.72	0.00	75146.87	31436.78	0.00
76.4	22.9	111.43	0.00	489.98	1320.60	2532.28	1058.55
189.5	23.53	175.84	1047.45	1676.19	704.43	946.75	1294.18
73.4	23.67	0.00	46957.96	0.00	218746.16	239750.72	0.00
137.5	23.95	0.00	4661.15	3975.54	1957.35	3441.49	2148.89
415.6	24.97	1081.46	696.27	7396.77	1721.39	765.03	1195.63
399.7	25.93	22.79	90.14	625.76	9.54	882.75	0.00
148.5	26.42	860.09	1300.41	1863.13	1619.93	3866.76	1215.11
179.6	28.3	0.00	105.07	0.00	0.00	63.91	0.00

mz	rt	ARS803-4-5-4-3	ARS848-15-3-2-3	ART29	CRS36	D20-ARS-3-2	GIZA128
		Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	0.00	1548.61	763.75	512.61	0.00	1744.61
111.5	6.52	0.00	1579.63	0.00	1528.79	0.00	7867.91
109.5	7.82	0.00	1041.71	1091.23	636.18	0.00	1367.83
98.5	9.52	25.94	1293.61	1311.20	809.88	0.00	2909.52
116.5	9.97	0.00	247.35	173.70	236.38	0.00	383.49
40.4	11.7	0.00	8571.65	7758.22	1714.87	0.00	6339.72
105.5	12.48	0.00	526.00	468.98	285.45	0.00	591.28
44.4	14.62	0.00	2347.15	2481.58	3135.46	0.00	2686.13
98.5	16.33	0.00	0.00	0.00	1227.68	0.00	449.32
267.8	16.37	0.00	65.73	0.00	0.00	0.00	103.16
342.8	17.63	0.00	17958.37	10785.73	1366.47	0.00	9070.57
73.4	17.97	8612.31	59297.54	95203.33	477.22	6023.03	45782.87
59.5	21.27	0.00	620.51	0.00	1685.53	0.00	941.64
90.6	22.18	0.00	109.45	147.89	24.70	0.00	172.75
74.5	22.7	3496.24	7595.07	7353.74	3930.76	0.00	4796.06
76.4	22.9	284.93	0.00	0.00	10.29	270.31	0.00
189.5	23.53	1451.23	1352.64	1186.55	1150.64	1254.95	0.00
73.4	23.67	0.00	64304.30	93041.73	33124.60	251860.79	96461.11
137.5	23.95	3526.17	1932.21	4651.48	2456.82	749.66	1682.65
415.6	24.97	428.66	1054.86	1030.76	644.05	878.10	1159.76
399.7	25.93	24.24	145.56	212.76	150.58	37.22	315.22
148.5	26.42	2563.54	0.00	0.00	708.53	1132.10	0.00
179.6	28.3	0.00	53.82	405.48	253.20	0.00	318.64

mz	rt	IR106359-B-18-5	IR107015-37	IR127229	IR17015-6-5-3-B1	IR64-biofortified	IR64
		Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	0.00	398.04	0.00	0.00	0.00	1229.49
111.5	6.52	1282.41	1694.50	30.73	0.00	182.52	4289.56
109.5	7.82	0.00	579.64	0.00	0.00	0.00	2399.42
98.5	9.52	0.00	59.55	0.00	0.00	0.00	1775.42
116.5	9.97	0.00	330.14	0.00	0.00	0.00	276.36
40.4	11.7	0.00	3241.57	0.00	0.00	0.00	3501.01
105.5	12.48	0.00	495.89	0.00	0.00	0.00	220.15
44.4	14.62	0.00	5539.67	0.00	0.00	0.00	4197.49
98.5	16.33	0.00	869.74	0.00	0.00	0.00	1765.31
267.8	16.37	0.00	267.28	0.00	0.00	0.00	0.00
342.8	17.63	0.00	0.00	0.00	0.00	0.00	16473.41
73.4	17.97	8511.33	3881.46	7579.26	9961.27	5272.12	59623.23
59.5	21.27	0.00	12935.06	0.00	0.00	0.00	2700.35
90.6	22.18	0.00	0.00	0.00	0.00	0.00	60.93
74.5	22.7	0.00	58958.72	0.00	0.00	0.00	4439.85
76.4	22.9	107.65	1254.66	1869.24	0.00	40.05	31.50
189.5	23.53	42.52	580.21	2731.27	226.36	1407.34	2027.51
73.4	23.67	145526.48	366259.96	0.00	95367.13	0.00	10067.53
137.5	23.95	427.81	2509.85	2509.19	670.73	3239.46	1448.94
415.6	24.97	1889.43	1246.62	15290.46	3057.44	837.94	357.70
399.7	25.93	3.15	0.00	1753.80	761.50	2.59	157.17
148.5	26.42	1315.82	2609.36	825.32	1230.18	3879.46	0.00
179.6	28.3	0.00	0.00	0.00	0.00	0.00	197.87

mz	rt	IR7525	IR87638-10-2-2-4	IR88638	IR93856-23-1-1-1	IR99084-B-B-13	Kasozi
		Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	0.00	0.00	0.00	0.00	0.00	0.00
111.5	6.52	32.14	15.97	35.72	0.00	228.85	0.00
109.5	7.82	0.00	0.00	90.58	0.00	0.00	0.00
98.5	9.52	0.00	0.00	0.00	0.00	4.45	0.00
116.5	9.97	0.00	0.00	0.00	0.00	79.45	78.62
40.4	11.7	0.00	0.00	0.00	0.00	0.00	0.00
105.5	12.48	0.00	0.00	0.00	0.00	0.00	0.00
44.4	14.62	0.00	0.00	0.00	0.00	0.00	0.00
98.5	16.33	0.00	0.00	0.00	0.00	0.00	0.00
267.8	16.37	0.00	0.00	0.00	0.00	0.00	0.00
342.8	17.63	0.00	0.00	0.00	0.00	0.00	0.00
73.4	17.97	0.00	8267.51	4873.64	4671.69	895.32	609.08
59.5	21.27	0.00	0.00	0.00	1790.47	17770.55	0.00
90.6	22.18	0.00	0.00	0.00	0.00	0.00	0.00
74.5	22.7	0.00	0.00	0.00	0.00	0.00	0.00
76.4	22.9	89.41	249.59	205.87	1991.88	675.79	101.59
189.5	23.53	1673.14	317.34	1489.62	3183.85	1610.72	0.00
73.4	23.67	66708.09	0.00	0.00	69824.15	5829.06	47315.66
137.5	23.95	842.50	1376.35	2384.84	0.00	2866.80	312.24
415.6	24.97	607.44	1254.90	3966.94	15940.33	1079.71	1772.59
399.7	25.93	0.00	38.59	513.65	909.87	21.42	0.00
148.5	26.42	435.35	1298.93	888.31	161.66	1848.51	431.06
179.6	28.3	0.00	0.00	0.00	0.00	0.00	0.00

mz	rt	Kigoma	Komboka	NL17	Orylux7	Sipi	V18RRS126-48-1-13-2
		Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic	Semi-aromatic
79.5	6.05	1612.23	950.73	0.00	0.00	1012.14	691.23
111.5	6.52	2282.55	2850.61	34.76	1025.27	4597.63	992.72
109.5	7.82	949.61	605.48	0.00	0.00	1297.78	1055.46
98.5	9.52	1574.04	1348.26	0.00	4.24	1675.68	707.57
116.5	9.97	339.36	132.70	0.00	0.00	156.77	263.56
40.4	11.7	7832.88	2393.64	0.00	0.00	9051.34	1626.14
105.5	12.48	507.96	311.14	0.00	0.00	383.09	466.95
44.4	14.62	5403.03	4836.37	0.00	0.00	1377.44	3341.61
98.5	16.33	0.00	598.86	0.00	0.00	0.00	1407.40
267.8	16.37	0.00	0.00	0.00	0.00	0.00	97.81
342.8	17.63	1726.18	7373.88	0.00	0.00	8280.48	259874.85
73.4	17.97	45464.18	93643.79	4887.29	8108.75	73150.59	1916.01
59.5	21.27	1765.54	1944.82	0.00	0.00	687.19	2431.15
90.6	22.18	51.15	99.77	0.00	0.00	177.80	0.00
74.5	22.7	7488.67	6432.67	0.00	0.00	6194.43	8926.84
76.4	22.9	0.00	0.00	0.00	32.23	0.00	177.66
189.5	23.53	1051.22	529.65	247.73	339.57	1311.10	1292.01
73.4	23.67	72605.47	149756.54	9583.39	0.00	124850.75	37104.68
137.5	23.95	4152.18	1743.86	0.00	273.54	4533.78	1361.36
415.6	24.97	977.59	1059.65	1285.02	996.62	1052.32	613.15
399.7	25.93	144.92	127.61	14.82	24.38	364.46	116.50
148.5	26.42	0.00	442.57	722.27	1031.21	0.00	1742.79
179.6	28.3	233.71	168.09	0.00	0.00	46.57	201.09

Appendix VII: List of Publications and Conference Presentation

Peer-reviewed articles:

Kimwemwe, P.K.; Bukomarhe, C.B.; Mamati, E.; Githiri, S.M.; Rene M.C.; Mignouna, J.; Kimani, W., & Fofana, M. (2023). Population Structure and Genetic Diversity of Rice (*Oryza sativa* L.) Germplasm from the Democratic Republic of Congo (DRC) Using DArTseq-Derived Single Nucleotide Polymorphism (SNP). *Agronomy*, 13 (1906). https://doi.org/10.3390/agronomy13071906

Bukomarhe C.B.; **Kimwemwe P.K.**; Githiri, S.M.; Mamati, E.G.; Kimani W.; Mutai C.; Nganga F.; Nguezet PM. D.; Mignouna J.; Civava R.M.; and Fofana M. (2023). Association mapping of candidate genes associated with Iron and Zinc Content in Rice (*Oryza sativa* L.) Grains. *Genes*, 14 (1815). https://doi.org/10.3390/genes14091815

Conference presentation: Second Biennial Africa Climate-Smart Agriculture (CSA) Stakeholders Conference, Accra, Ghana. 14-16 September 2023.

Kimwemwe K.P., Bukomarhe B.C., Mudarhi B.L., Munkumba D.D., Tshiabukole P.K.J., Okonya S.J., Warinda E., Ndikumana I., Mamati G.E. and Fofana M. (2023) Phenotypic variation among rice (*Oryza sativa* L.) germplasm accessions for the Eastern Democratic Republic of Congo and traits association based on yield and yield components. *FARA Research Report*, Vol 7(49):622-641. https://library.faraafrica.org/storage/2023/04/FRR-Vol-749622-641.pdf